The Annals of Applied Statistics
2014, Vol. 8, No. 2, 674–702
DOI: 10.1214/14-AOAS729
© Institute of Mathematical Statistics, 2014

ESTIMATION IN THE PARTIALLY OBSERVED STOCHASTIC MORRIS–LECAR NEURONAL MODEL WITH PARTICLE FILTER AND STOCHASTIC APPROXIMATION METHODS

BY SUSANNE DITLEVSEN[*,1] AND ADELINE SAMSON[†,‡,2]

University of Copenhagen, Université Paris Descartes[†] and Université Grenoble-Alpes[‡]*

Parameter estimation in multidimensional 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, defined by a nonlinear two-dimensional stochastic differential equation. The coordinates are coupled, that is, the unobserved coordinate is nonautonomous, the model exhibits oscillations to mimic the spiking behavior, which means it is not of gradient-type, and the measurement noise from intracellular recordings is typically negligible. Therefore, the hidden Markov model framework is degenerate, and available methods break down. The main contributions of this paper are an approach to estimate in this ill-posed situation and nonasymptotic convergence results for the method. Specifically, 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. 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. An experimental data set of intracellular recordings of the membrane potential of a spinal motoneuron of a red-eared turtle is analyzed, and the performance is further evaluated in a simulation study.

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. Often these models are compared to experimental data by hand-tuning to reproduce the

Received March 2013; revised October 2013.

[1]Supported in part by grants from the Danish Council for Independent Research | Natural Sciences.

[2]Supported in part by grants from the University Paris Descartes PCI.

Key words and phrases. Sequential Monte Carlo, diffusions, pseudo likelihood, Stochastic Approximation Expectation Maximization, motoneurons, conductance-based neuron models, membrane potential.

qualitative behaviors observed in experimental data, but without any formal statistical analysis. It is of particular interest to estimate conductances, which reflect the synaptic input from the surrounding network. These can be estimated from intracellular recordings, where the neuronal membrane potential is recorded at high frequency, and are typically done using only subthreshold fluctuations, ignoring the dynamics during action potentials [Berg, Alaburda and Hounsgaard (2007), Berg and Ditlevsen (2013), Borg-Graham, Monier and Frégnac (1998), Monier, Fournier and Frégnac (2008), Pospischil et al. (2009), Rudolph et al. (2004)]. The aim of this article is to estimate such biophysical parameters during the dynamics of spiking from intracellular data.

The Morris–Lecar model [Morris and Lecar (1981)] is a simple biophysical model and a prototype for a wide variety of neurons. It is a conductance-based model [Gerstner and Kistler (2002)], introduced to explain the dynamics of the barnacle muscle fiber. It is given by two coupled first order differential equations, the first modeling the membrane potential evolution and the second 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. If one of these noise sources are zero, a hypoelliptic diffusion arises leading to singular transition densities and particular statistical challenges [Pokern, Stuart and Wiberg (2009), Samson and Thieullen (2012)]. Typically, the membrane potential will be measured discretely at high frequency, whereas the second variable cannot be observed. Our goal is to estimate model parameters from discrete observations of the first coordinate in the nonsingular case of nonnegligible noise on both coordinates. This includes estimation of a central rate parameter characterizing the channel kinetics of the unobserved component, which we believe has not been done before.

Estimation in these conductance-based models is not straightforward. Because of the coupling between the coordinates of the stochastic differential equation (SDE), the unobserved coordinate is nonautonomous, and the model does not fit into the (nondegenerate) Hidden Markov Model (HMM) framework, as explained in Section 2.3. Furthermore, the diffusion is not time reversible and the likelihood is generally not tractable. Thus, the problem of inference is complex. The literature contains various methodologies when all the coordinates are observed [Aït-Sahalia (2002), Beskos et al. (2006), Durham and Gallant (2002), Jensen et al. (2012), Pedersen (1995), Sørensen (2004, 2012)] or the hidden state is Markovian [Ionides et al. (2011)]. They strongly rely on the Markov property and are hard to generalize to the non-Markovian case we are studying. In the non-Markovian case, methods are mainly based on data augmentation. The idea is that the likelihood can be approximated given the entire path or a sufficient partition of it. Therefore, the unobserved coordinates are treated as missing data and are imputed. Most methods propose to approximate the transition density by the Euler–Maruyama scheme and consider a Bayesian point of view to estimate the posterior distribution of the parameters [Elerian, Chib and Shephard (2001), Eraker (2001), Golightly and

Wilkinson (2006, 2008)]. Golightly and Wilkinson (2006) study a model similar to us but with low frequency data. So they need to impute data between observations, which is computationally costly. Furthermore, there exists a strong dependence between the imputed sample paths and the diffusion coefficient, and it is not possible to estimate the diffusion parameter with this kind of approach. An alternative is reparametrization of the diffusion, but it is limited to scalar diffusions [Roberts and Stramer (2001)] or an autonomous hidden coordinate [Kalogeropoulos (2007)].

In this paper, we propose to estimate the parameters with a maximum likelihood approach. We approximate the SDE through an Euler–Maruyama scheme to obtain a tractable pseudo-likelihood. Then we consider the statistical model as an incomplete data model and maximize the pseudo-likelihood through a stochastic Expectation–Maximization (EM) algorithm, where the unobserved data are imputed at each iteration of the algorithm. We are in the setting of high frequency data so we do not need to impute data between observations, but our approach could be extended to that type of data as well. A similar but different method has been proposed by Huys, Ahrens and Paninski (2006), where up to 10^4 parameters are estimated in a detailed multicompartmental single neuron model. However, 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. Because of the measurement noise, the model fits into the HMM framework and they can use a standard particle filter. But again, only parameters entering linearly in the pseudo-likelihood are considered. In particular, all parameters of the hidden coordinate are assumed known.

Here, we also want to estimate parameters from the hidden coordinate and we do not consider measurement noise. We propose to impute the hidden non-Markovian path in the stochastic EM algorithm with a Sequential Monte Carlo (SMC) algorithm. Monte Carlo methods for nonlinear filtering are widely spread, with, among other algorithms, sequential importance sampling, sequential importance sampling with resampling (SISR), auxiliary SISR and stratified resampling [see Cappé, Moulines and Rydén (2005) for a general presentation]. All SISR algorithms are now called SMC. Most of them are designed for HMM. In the specific setting of multidimensional SDEs, Del Moral, Jacod and Protter (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, they propose to simulate it at each iteration of the filter. Fearnhead, Papaspiliopoulos and Roberts (2008) generalize this particle filter to a nonautonomous hidden path but with drift of gradient type. In the ergodic case this corresponds to a time-reversible

diffusion. In particular, models exhibiting oscillations are not covered, which is the case of any realistic neuronal model.

These algorithms cannot be directly applied because we are studying a multidimensional coupled SDE that is not of gradient type. Thus, we consider the SMC algorithm proposed by Doucet, de Freitas and Gordon (2001) for more general dynamic models than HMM. As we combine this SMC with the Stochastic Approximation Expectation–Maximization (SAEM) algorithm which maximizes the pseudo-likelihood based on an Euler–Maruyama approximation of the SDE defining the model, we need nonasymptotic convergence results for the SMC to obtain the convergence of the SAEM–SMC. Nonasymptotic results for SMC, such as deviation inequalities, have been proposed in the literature only in the HMM framework [Del Moral and Miclo (2000), Del Moral, Jacod and Protter (2001), Douc et al. (2011), Künsch (2005)], and the Markovian structure of the hidden path is a key element in the proofs. A major contribution here is that we are able to extend this result to a SMC for a non-Markovian hidden path. Then 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 noise structure is motivated, and the pseudo-likelihood arising from the Euler–Maruyama approximation is found. In Section 3 the filtering problem is presented, as well as the SMC algorithm and deviation inequalities. In Section 4 we present the estimation procedure and the assumptions needed for the convergence results to hold. In Section 5 we apply the method on an experimental data set of intracellular recordings of the membrane potential of a motoneuron of a turtle, and in Section 6 we conduct a simulation study to document the performance of the method. Proofs and technical results can be found in the Appendix.

2. Stochastic Morris–Lecar model.

2.1. *Exact diffusion model.* The stochastic Morris–Lecar model including both current and channel noise is defined as the solution to

$$\begin{cases} dV_t = f(V_t, U_t)\,dt + \gamma\,d\tilde{B}_t, \\ dU_t = b(V_t, U_t)\,dt + \sigma(V_t, U_t)\,dB_t, \end{cases} \tag{1}$$

where

$$f(V_t, U_t) = \frac{1}{C}\big(-g_{\mathrm{Ca}} m_\infty(V_t)(V_t - V_{\mathrm{Ca}}) - g_{\mathrm{K}} U_t (V_t - V_{\mathrm{K}}) - g_{\mathrm{L}}(V_t - V_{\mathrm{L}}) + I\big),$$

$$b(V_t, U_t) = \big(\alpha(V_t)(1 - U_t) - \beta(V_t)U_t\big),$$

$$m_\infty(v) = \frac{1}{2}\left(1 + \tanh\left(\frac{v - V_1}{V_2}\right)\right),$$

$$\alpha(v) = \frac{1}{2}\phi \cosh\left(\frac{v - V_3}{2V_4}\right)\left(1 + \tanh\left(\frac{v - V_3}{V_4}\right)\right),$$
$$\beta(v) = \frac{1}{2}\phi \cosh\left(\frac{v - V_3}{2V_4}\right)\left(1 - \tanh\left(\frac{v - V_3}{V_4}\right)\right),$$

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 time t, and U_t represents the normalized conductance of the K^+ current. It varies between 0 and 1, and can 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 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; V_{Ca}, V_K and V_L are reversal potentials for Ca^{2+}, K^+ and leak currents; 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.

Various noise sources are present in single neurons, and they act on many different spatial and temporal scales [Gerstner and Kistler (2002), Longtin (2013)]. A main component arises from the synaptic bombardment from other neurons in the network, and in the diffusion limit appears as an additive noise on the current equation. Parameter γ scales this current noise. Conductance fluctuations caused by random opening and closing of ion channels leads to multiplicative noise on the conductance equation. Function $\sigma(V_t, U_t)$ models this 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 (2013)]: $\sigma(V_t, U_t) = \sigma\sqrt{2\frac{\alpha(V_t)\beta(V_t)}{\alpha(V_t)+\beta(V_t)}U_t(1 - U_t)}$. A trajectory of the model is simulated in Figure 1. The peaks of (V_t) correspond to spikes of the neuron.

2.2. *Observations and approximate model.* Data are discrete measurements of (V_t), while (U_t) is not measured. We denote $t_0 \leq t_1 \leq \cdots \leq t_n$ 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. Let $\theta \in \Theta \subseteq \mathbb{R}^p$ be the vector of parameters to be estimated. We consider estimation of all identifiable parameters of the observed coordinate and the rate parameter of the unobserved channel dynamics $\theta = (g_{Ca}, g_K, g_L, V_{Ca}, V_K, I, \gamma, \phi)$. 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. We conjecture that the information about σ in the observed coordinate is close to zero and, thus, in practice, also σ is unidentifiable from observations of $V_{0:n}$ only, at least for any

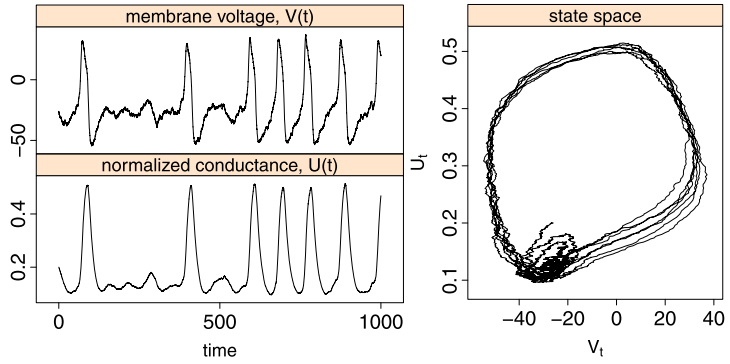

FIG. 1. *Simulated trajectory of the stochastic Morris–Lecar model*: (V_t) *as a function of time* (*left*, *top*), (U_t) *as a function of time* (*left*, *bottom*), *and* (U_t) *against* (V_t) (*right*). *Parameters are given in Section* 6. *Time is measured in ms*, *voltage in mV*, *and the conductance is normalized between* 0 *and* 1.

finite sample size. This happens because σ is mainly shaping the dynamics of U_t between spikes, while the dynamics during spikes resemble deterministic behavior, and the influence of U_t on V_t is only strong during spikes. This is confirmed in Sections 5 and 6 where misspecification of σ is shown not to deteriorate the estimation of θ. Finally, we assume the scaling parameters V_1–V_4 known because otherwise the model does not belong to an exponential family, as required by assumption (M1) below. This could be solved by introducing an extra optimization step in the EM algorithm at the cost of precision and computer time. It is not pursued further in this work.

The aim is to estimate θ by maximum likelihood. However, this likelihood is intractable, as the transition density of model (1) is not explicit. Let Δ denote the step size between two observation times, which for simplicity we assume do 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:

$$
\begin{aligned}
V_{i+1} &= V_i + \Delta f(V_i, U_i) + \sqrt{\Delta}\gamma\tilde{\eta}_i, \\
U_{i+1} &= U_i + \Delta b(V_i, U_i) + \sqrt{\Delta}\sigma(V_i, U_i)\eta_i,
\end{aligned} \tag{2}
$$

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.

2.3. *Property of the observation model.* The observation model is a degenerate HMM. Let us recall the definition proposed by Cappé, Moulines and Rydén (2005). A HMM with not countable state space is defined as a bivariate Markov

chain (X_i, Y_i) with only partial observations Y_i, whose transition kernel has a special structure: both the joint process (X_i, Y_i) *and* the marginal hidden chain (X_i) are Markovian.

In our model, (U_i) is not Markovian, only (V_i, U_i) is Markovian. So set $X_i = (V_i, U_i)$, with Markov kernel $R(X_{i-1}, dX_i) = p_\Delta(dV_i, dU_i|V_{i-1}, U_{i-1})$, the transition density of model (2), and $Y_i = X_i^{(1)}$, the first coordinate of X_i with transition kernel $F(X, dY) = \mathbb{1}_{\{Y=X^{(1)}\}}$. Here, $\mathbb{1}_x$ is the Dirac measure in x. Thus, the kernel F is zero almost everywhere and the HMM is degenerate. This leads to an intrinsic degeneracy of the particle filter used in the standard HMM toolbox, as explained below.

Therefore, we consider the observation model as a bivariate Markov chain (V_i, U_i) with only partial observations V_i whose hidden coordinate U_i is *not* Markovian. It is not a HMM but a general dynamic model as considered by Andrieu, Doucet and Punskaya (2001). The hidden process U_i is distributed as

$$U_0 \sim \mu(dU_0), \qquad U_i|(U_{0:i-1}, V_{0:i-1}) \sim K(dU_i|U_{0:i-1}, V_{0:i-1})$$

for some conditional distribution function K and the observed process V_i is distributed as

$$V_i|(U_{0:i}, V_{0:i-1}) \sim G(dV_i|U_{0:i}, V_{0:i-1})$$

for some distribution function G. Given the Markovian structure of the pair (V_i, U_i), we have $K(dU_i|U_{0:i-1}, V_{0:i-1}) = K(dU_i|U_{i-1}, V_{i-1})$ and $G(dV_i|U_{0:i}, V_{0:i-1}) = G(dV_i|U_{i-1:i}, V_{i-1})$. To simplify, we use the same notation for random variables and their realizations and assume that $G(dV_i|U_{0:i}, V_{0:i-1}) = G(V_i|U_{0:i}, V_{0:i-1})\,dV_i$.

2.4. *Likelihood function.* We want to estimate the parameter θ by maximum likelihood of the approximate model, with likelihood

$$p_\Delta(V_{0:n}; \theta) = \int p(V_0, U_0; \theta) \prod_{i=1}^{n} p_\Delta(V_i, U_i|V_{i-1}, U_{i-1}; \theta)\, dU_{0:n}. \tag{3}$$

It corresponds to a pseudo-likelihood for the exact diffusion. The multiple integrals of equation (3) are difficult to handle and it is not possible to maximize the pseudo-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. To maximize the likelihood of the complete data vector $(V_{0:n}, U_{0:n})$, we propose to use a stochastic version of the EM algorithm, namely, the SAEM algorithm [Delyon, Lavielle and Moulines (1999)]. Simulation under the smoothing distribution $p_\Delta(U_{0:n}|V_{0:n}; \theta)$ is likely to be difficult, and direct simulation of the nonobserved data $(U_{0:n})$ is not possible. A SMC algorithm, also known as Particle Filtering, provides a way to approximate this distribution

[Doucet, de Freitas and Gordon (2001)]. We have adapted this algorithm to handle a coupled two-dimensional SDE, that is, the unobserved coordinate is nonautonomous and non-Markovian. Then, 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 (2). Details on the filtering are given in Section 3, and the SAEM algorithm is presented in Section 4.1. To prove the convergence of this new SAEM–SMC algorithm, a nonasymptotic deviation inequality is required for the SMC algorithm. Then we derive the convergence of the SAEM–SMC algorithm to a maximum of the likelihood.

3. Filtering.

3.1. *The filtering problem and the SMC algorithm.* For any bounded Borel function $f:\mathbb{R}\mapsto\mathbb{R}$, we denote $\pi_{n,\theta}f=\mathbb{E}_\Delta(f(U_n)|V_{0:n};\theta)$, the conditional expectation under the exact smoothing distribution $p_\Delta(U_{0:n}|V_{0:n};\theta)$ of the approximate model. The aim is to approximate this distribution for a fixed value of θ. When included in the stochastic EM algorithm, this value will be the current value $\widehat{\theta}_m$ at the given iteration. For notational simplicity, θ is omitted in the rest of this section.

We now argue why the HMM point of view is ill-posed for the filtering problem. Considering the model as a HMM, $X_i=(V_i,U_i)$ is the hidden Markov chain and $Y_i=X_i^{(1)}$. But then the filtering problem $\pi_n f$ is the ratio of $\int\mu(dU_0)R(X_0,dX_1)F(X_0;Y_1)\cdots R(X_{n-1},dX_n)F(X_{n-1};Y_n)f(X_n)$ and $\int\mu(dU_0)R(X_0,dX_1)F(X_0;Y_1)\cdots R(X_{n-1},dX_n)F(X_{n-1};Y_n)$. Since $F(X_{n-1};Y_n)=\mathbb{1}_{\{Y_n=X_{n-1}^{(1)}\}}$ and the state space is continuous, the denominator is zero almost surely and the filtering problem is ill-posed.

Now consider the model in a more general framework with the hidden state U_i not Markovian, and introduce for $i=1,\dots,n$ the kernels H_i from $\mathbb{R}$ into itself by

$$\begin{aligned} H_i f(u) &= \int K(dz|u,V_{i-1})G(V_i|u,V_{i-1},z)f(z) \\ &= \int p_\Delta(V_i,z|V_{i-1},u)f(z)\,dz. \end{aligned} \tag{4}$$

Then π_n can be expressed recursively by

$$\pi_n f=\frac{\pi_{n-1}H_n f}{\pi_{n-1}H_n 1}=\frac{\int\mu(U_0)\prod_{i=1}^n p_\Delta(V_i,U_i|V_{i-1},U_{i-1})f(U_n)\,dU_{0:n}}{\int\mu(U_0)\prod_{i=1}^n p_\Delta(V_i,U_i|V_{i-1},U_{i-1})\,dU_{0:n}}. \tag{5}$$

Note that the denominator of (5) is $\mu H_1\cdots H_n 1=p_\Delta(V_{0:n})$, which is different from 0 since its support is the real line. Thus, the filtering problem is well-posed.

The kernels H_i are extensions of the kernels considered by Del Moral, Jacod and Protter (2001) in the context of two-dimensional SDEs with hidden coordinate U_t autonomous (and thus Markovian). We do not extend their particle

filter since it is based on simulation of both V_i and U_i with transition kernel $p_\Delta(V_i, U_i|V_{i-1}, U_{i-1})$. They avoid the degeneracy of the weights by introducing an instrumental function ψ and the weights are computed as $\psi(V_i^{(k)} - V_i)$. The choice of this instrumental function may influence the numerical properties of the filter. Therefore, we adopt the general filter proposed by Andrieu, Doucet and Punskaya (2001) for a more general dynamic system, that we recall here.

The SMC algorithm provides a set of K particles $(U_{0:n}^{(k)})_{k=1,\dots,K}$ and weights $(W_{0:n}^{(k)})_{k=1,\dots,K}$ approximating the conditional smoothing distribution $p_\Delta(U_{0:n}|V_{0:n})$ [see Doucet, de Freitas and Gordon (2001)]. The SMC method relies on proposal distributions $q(U_i|V_i, V_{i-1}, U_{i-1})$ to sample what we call particles from these distributions. We write $V_{0:i} = (V_0, \dots, V_i)$ and likewise for $U_{0:i}$.

ALGORITHM 1 (SMC algorithm).

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

$$w_0(U_0^{(k)}) = p(V_0, U_0^{(k)}), \qquad W_0(U_0^{(k)}) = \frac{w_0(U_0^{(k)})}{\sum_{k=1}^K w_0(U_0^{(k)})}.$$

- *At time* $i = 1, \dots, n$*:* $\forall k = 1, \dots, K$:
  1. sample indices $A_{i-1}^{(k)} \sim r(\cdot|W_{i-1}(U_{0:i-1}^{(1)}), \dots, W_{i-1}(U_{0:i-1}^{(K)}))$. Set $U_{0:i-1}^{\prime(k)} = U_{0:i-1}^{(A_{i-1}^{(k)})}$,
  2. sample $U_i^{(k)} \sim q(\cdot|V_{i-1:i}, U_{i-1}^{\prime(k)})$ and set $U_{0:i}^{(k)} = (U_{0:i-1}^{\prime(k)}, U_i^{(k)})$,
  3. compute and normalize the weights

$$W_i(U_{0:i}^{(k)}) = \frac{w_i(U_{0:i}^{(k)})}{\sum_{k=1}^K w_i(U_{0:i}^{(k)})} \qquad \text{with}$$

$$w_i(U_{0:i}^{(k)}) = \frac{p_\Delta(V_{0:i}, U_{0:i}^{(k)})}{p_\Delta(V_{0:i-1}, U_{0:i-1}^{\prime(k)}) q(U_i^{(k)}|V_{i-1:i}, U_{0:i-1}^{\prime(k)})}.$$

The SMC algorithm provides an empirical measure $\Psi_n^K = \sum_{k=1}^K W_n(U_{0:n}^{(k)}) \mathbb{1}_{U_{0:n}^{(k)}}$ which is an approximation to the smoothing distribution $p_\Delta(U_{0:n}|V_{0:n})$. A draw from this distribution can be obtained by sampling an index k from a multinomial distribution with probabilities $W_n(U_{0:n}^{(1)}), \dots, W_n(U_{0:n}^{(K)})$ and setting the draw $U_{0:n}$ equal to $U_{0:n} = U_{0:n}^{(k)}$.

The variable $A_{i-1}^{(k)}$ plays an important role to discard the samples with small weights and multiply those with large weights [Gordon, Salmond and Smith

(1993)]. It generates a number of offspring $N_{i-1}^{(\ell)}$, $\ell = 1, \dots, K$, such that $\sum_{\ell=1}^{K} N_{i-1}^{(\ell)} = K$ and $\mathbb{E}(N_{i-1}^{(\ell)}) = K W_{i-1}(U_{0:i-1}^{(l)})$. Many schemes for r have been presented in the literature, including multinomial sampling [Gordon, Salmond and Smith (1993)], residual sampling [Liu and Chen (1998)] or stratified resampling [Doucet, Godsill and Andrieu (2000)]. They differ in terms of $\operatorname{var}(N_{i-1}^{(\ell)})$ [see Doucet, Godsill and Andrieu (2000)]. The key property that we need in order to prove the deviation inequality is that $\mathbb{E}(\mathbb{1}_{\{A_{i-1}^{(k)}=\ell\}}) = W_{i-1}(U_{0:i-1}^{(l)})$.

Since our model is not a HMM, the weights $w_i(U_{0:i}^{(k)})$ cannot be written in terms of a Markov transition kernel of the hidden path as is usually done. It follows that the proposal q, which is crucial to ensure good convergence properties, has to depend on V_i. The first classical choice of q is $q(U_i|V_{i-1:i}, U_{i-1}) = p_\Delta(U_i|V_{i-1}, U_{i-1})$, that is, the transition density. In this case, the weight reduces to $w_i(U_{0:i}^{(k)}) = p_\Delta(V_i|V_{i-1}, U_{0:i}^{(k)})$. A second choice for the proposal is $q(U_i|V_{i-1:i}, U_{i-1}) = p_\Delta(U_i|V_{i-1:i}, U_{i-1})$, that is, the conditional distribution. In this case, the weight reduces to $w_i(U_{0:i}^{(k)}) = p_\Delta(V_i|V_{i-1}, U_{0:i-1}^{(k)})$. Transition densities and conditional distributions are detailed in Appendix A. When the two Brownian motions are independent, as we assume, the two choices are equivalent.

This SMC algorithm is plugged into the EM algorithm to estimate the parameters. We thus need nonasymptotic convergence results on the SMC algorithm to ensure the convergence of the EM algorithm. This is discussed in the next section.

3.2. *Deviation inequality.* In the literature, deviation inequalities for SMC algorithms only appear for HMM. To our knowledge, this is the first nonasymptotic result proposed for a SMC applied to a non-Markovian hidden path. The only result of this type with SDEs has been obtained by Del Moral, Jacod and Protter (2001), with autonomous second coordinate. Here, we generalize their deviation inequality to a nonautonomous hidden path.

For a bounded Borel function f, denote $\Psi_n^K f = \sum_{k=1}^{K} f(U_n^{(k)}) W_{n,\theta}(U_{0:n}^{(k)})$, the conditional expectation of f under the empirical measure $\Psi_{n,\theta}^K$ obtained by the SMC algorithm for a given value of θ. We have the following:

PROPOSITION 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*

$$\mathbb{P}\big(|\Psi_{n,\theta}^K f - \pi_{n,\theta} f| \geq \varepsilon\big) \leq C_1 \exp\left(-K \frac{\varepsilon^2}{C_2 \|f\|^2}\right), \tag{6}$$

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

The proof is provided in Appendix D. A similar result can be obtained with respect to the exact smoothing distribution of the exact diffusion model, under assumptions on the number of particles and the step size of the Euler approximation.

4. Estimation method.

4.1. *SAEM algorithm.* The EM algorithm [Dempster, Laird and Rubin (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 $Q(\theta|\theta') = \mathbb{E}_\Delta[\log p_\Delta(V_{0:n}, U_{0:n};\theta)|V_{0:n};\theta']$, 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 (2) 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 $\widehat{\theta}_{m-1}$, 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 the following:

(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, that is, $\log p_\Delta(V_{0:n}, U_{0:n};\theta) = -\psi(\theta) + \langle S(V_{0:n}, U_{0:n}), \nu(\theta)\rangle$, 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 (2) satisfies this assumption when the scaling parameters V_1, V_2, V_3 and V_4 are known. Details of the sufficient statistic S are given in Appendix B.

Under assumption (M1), the E-step reduces to the computation of $\mathbb{E}_\Delta[S(V_{0:n}, U_{0:n})|V_{0:n};\widehat{\theta}_{m-1}]$. When this expectation has no closed form, Delyon, Lavielle and Moulines (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 nonobserved 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 $\mathbb{E}_\Delta[S(V_{0:n}, U_{0:n})|V_{0:n};\widehat{\theta}_{m-1}]$ with a sequence of positive numbers $(a_m)_{m\in\mathbb{N}}$ decreasing to zero. We write s_m for the approximation of this expectation. At the S-step, the simulation under the smoothing distribution is done by SMC, as explained in Section 3. We call this algorithm the SAEM–SMC algorithm. Iterations of the SAEM–SMC algorithm are written as follows:

ALGORITHM 2 (SAEM–SMC algorithm).

- *Iteration* 0*:* initialization of $\widehat{\theta}_0$ and set $s_0 = 0$.
- *Iteration* $m \geq 1$*:*
  - *S-step:* simulation of the nonobserved data $(U_{0:n}^{(m)})$ with SMC targeting the distribution $p_\Delta(U_{0:n}|V_{0:n};\widehat{\theta}_{m-1})$.

SA-step: update s_{m-1} using the stochastic approximation:

$$s_m = s_{m-1} + a_{m-1}\bigl[S\bigl(V_{0:n}, U_{0:n}^{(m)}\bigr) - s_{m-1}\bigr]. \tag{7}$$

M-step: update of $\widehat{\theta}_m$ by $\widehat{\theta}_m = \arg\max_{\theta\in\Theta}(-\psi(\theta) + \langle s_m, \nu(\theta)\rangle)$.

Standard errors of the 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 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.

4.2. *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 toward the maximum of the likelihood of the approximated diffusion (2). 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, Lavielle and Moulines (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 $L:\mathcal{S}\times\Theta\to\mathbb{R}$ by $L(s,\theta) = -\psi(\theta) + \langle s, \nu(\theta)\rangle$. There exists a function $\hat{\theta}:\mathcal{S}\to\Theta$ such that $\forall\theta\in\Theta, \forall s\in\mathcal{S}, L(s,\hat{\theta}(s)) \geq L(s,\theta)$.
- (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 $\Gamma(\theta) = \mathrm{Cov}_\theta(S(\cdot, U_{0:n}))$ is continuous, where the covariance is under the conditional distribution $p_\Delta(U_{0:n}|V_{0:n};\theta)$.
- (SAEM4) Let $\{\mathcal{F}_m\}$ be the increasing family of σ-algebras generated by the random variables $s_0, U_{0:n}^{(1)}, U_{0:n}^{(2)}, \ldots, U_{0:n}^{(m)}$. For any positive Borel function f, $\mathbb{E}_\Delta(f(U_{0:n}^{(m+1)})|\mathcal{F}_m) = \int f(u)p_\Delta(u|v,\widehat{\theta}_m)\,du$.

Assumptions (M1)–(M5) ensure the convergence of the EM algorithm when the E-step is exact [Delyon, Lavielle and Moulines (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, Lavielle and Moulines (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, Lavielle and Moulines (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* $p_\Delta(V_{0:n};\theta)$.

The classical assumptions (M1)–(M5) are usually 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.

In practice, the SAEM algorithm is implemented with an increasing number equal to the iteration number, which satisfies Assumption (SMC1). 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, Jacod and Protter (2001), which resimulates the variables V at each iteration. Assumption (SMC3) is satisfied if we require that γ is strictly bounded away from zero; $\gamma \geq \varepsilon > 0$.

4.3. *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 (2), with the true likelihood $p(V_{0:n}; \theta)$. The result is based on the bound of the Euler approximation proved by Gobet and Labart (2008). 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:

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

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

Proof is given in Appendix D. Assumption (H1) is a strong assumption, which is sufficient and not necessary. It does not hold for the Morris–Lecar model. Different sets of weaker assumptions have been proposed to prove the convergence of the Euler scheme in the strong sense (expectation of the absolute error between the exact and approximated process). The proofs are mainly based on localization arguments; see Kloeden and Neuenkirch (2012) for a review paper. The convergence of the densities has been less studied, and it is beyond the scope of this paper.

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, Alaburda and Hounsgaard (2007), Berg, Ditlevsen and Hounsgaard (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). However, in Jahn et al. (2011), only one-dimensional diffusions are considered, and spikes are modeled as single points in time by adding a jump term with state-dependent intensity function to the SDE, ignoring the detailed dynamics during spikes. In this

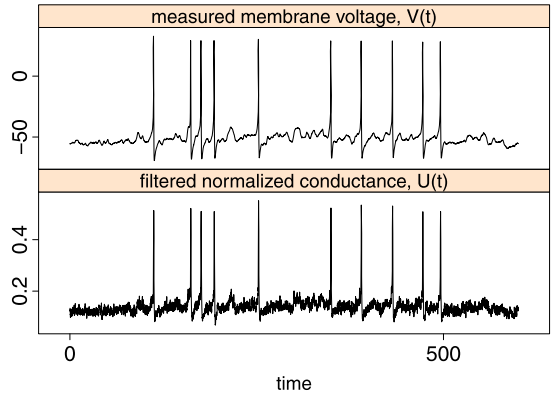

FIG. 2. *Observations of the membrane potential in a spinal motoneuron of an adult red-eared turtle during* 600 *ms* (*upper panel*) *and the filtered hidden process of the normalized conductance associated with* K^+ *current* (*lower panel*) *for the estimated parameters with the scaling parameters fixed at* $V_1 = -2.4$ *mV*, $V_2 = 36$ *mV*, $V_3 = 4$ *mV and* $V_4 = 60$ *mV*.

paper we aim at estimating parameters during spiking activity by explicit modeling of time-varying conductances. Therefore, we only analyze four traces during on-cycles [following Jahn et al. (2011)] where spikes occur. Furthermore, in these time windows, the input is approximately constant, which is required for the Morris–Lecar model with constant parameters. An example of the analyzed data is plotted in Figure 2, 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 given in Rinzel and Ermentrout (1989) and used in Section 6 below; see Table 1 for one of the traces. 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 mV (esti-

TABLE 1

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. With* $\sigma = 0.05$ *fixed. First trace*

	Parameter							
	g_{L}	g_{Ca}	g_{K}	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	−0.296	11.274	6.553	2.801	−124.481	1.989	35.769	−5.024
SE	0.001	0.028	0.049	0.001	14.563	0.000	0.122	0.052
With $V_1 = -2.4$ mV, $V_2 = 36$ mV, $V_3 = 4$ mV, $V_4 = 60$ mV								
Estimate	1.046	12.906	20.878	2.466	−67.097	2.153	98.698	−65.403
SE	0.009	0.008	0.021	0.001	0.227	0.001	0.198	1.204

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 three different values of σ. With $V_1 = -2.4$ mV, $V_2 = 36$ mV, $V_3 = 4$ mV, $V_4 = 60$ mV fixed. First trace

	Parameter							
	g_L	g_{Ca}	g_K	γ	V_K	ϕ	V_{Ca}	I
σ fixed to 0.02								
Estimate	1.302	12.460	16.550	2.280	−74.301	2.881	99.398	−52.742
SE	0.003	0.028	0.131	0.000	0.703	0.001	2.384	2.727
σ fixed to 0.05								
Estimate	1.046	12.906	20.878	2.466	−67.097	2.153	98.698	−65.403
SE	0.009	0.008	0.021	0.001	0.227	0.001	0.198	1.204
σ fixed to 0.15								
Estimate	1.308	12.442	16.419	2.301	−74.576	2.911	99.417	−52.341
SE	0.002	0.007	0.120	0.000	0.021	0.000	4.529	1.558

mated to 44.7 mV), and the leak conductance, which is estimated to be negative. Conductances are always nonnegative. This is probably due to wrong choices of the scaling constants V_1–V_4. For the parameters given in Rinzel and Ermentrout (1989), 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 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 when V_t is at its equilibrium value. In this case all parameters are reasonable and in agreement with the expected order of magnitudes.

To check the robustness to misspecifications in the diffusion parameter σ of the unobserved coordinate, we fitted the model for three different values of σ; see Table 2. Results are stable and suggest that σ is primarily affecting the subthreshold fluctuations of the channel dynamics, and mainly the spiking dynamics of the unobserved coordinate influences the first coordinate.

Final results for all four traces are presented in Table 3. It is reassuring that the parameter estimates seem so reproducible over different traces; the largest variation was below 10%. This is not due to starting values, for example, the starting value for g_L was 0.1, and all four estimates ended up between 1.3 and 1.4, and the starting value for V_K was −55, and all four estimates ended up between −75.4 and −74.4.

6. Simulation study. Parameter values of the Morris–Lecar model used in the simulations are the same as those of Rinzel and Ermentrout (1989), Tateno

TABLE 3

Parameter estimates obtained from four different traces of the membrane potential of a spinal motoneuron of an adult red-eared turtle. Each trace has 6000 *observations points with a sampling step of* 0.1 *ms. With* $V_1 = -2.4$ *mV*, $V_2 = 36$ *mV*, $V_3 = 4$ *mV*, $V_4 = 60$ *mV and* $\sigma = 0.05$ *fixed*

	Parameter							
	g_L	g_{Ca}	g_K	γ	V_K	ϕ	V_{Ca}	I
First trace								
Estimate	1.046	12.906	20.878	2.466	−67.097	2.153	98.698	−65.403
SE	0.009	0.008	0.021	0.001	0.227	0.001	0.198	1.204
Second trace								
Estimate	1.430	11.705	15.791	2.253	−73.029	3.138	103.709	−53.183
SE	0.003	0.008	0.029	0.000	0.641	0.001	1.269	0.651
Third trace								
Estimate	1.371	11.878	15.379	2.210	−75.024	3.004	99.887	−49.499
SE	0.002	0.013	0.017	0.000	0.195	0.000	0.464	0.614
Fourth trace								
Estimate	1.197	11.452	12.521	2.012	−85.982	3.776	99.615	−37.017
SE	0.002	0.055	0.017	0.000	0.089	0.000	1.466	0.861

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 as follows: $V_K = -84$ 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, $\phi = 0.04\ \text{ms}^{-1}$. 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 $\gamma = 1\ \text{mV ms}^{-1/2}$, $\sigma = 0.03\ \text{ms}^{-1/2}$. Trajectories are simulated with time step $\delta = 0.01$ ms and $n = 2000$ points are subsampled with observations time step $\Delta = 10\delta$. Then θ is estimated on each simulated trajectory. A hundred repetitions are used to evaluate the performance of the estimators. An example of a simulated trajectory (for $n = 10{,}000$) is given in Figure 1.

6.1. *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 θ fixed at its true value. The SMC Particle filter algorithm is implemented with $K = 100$ particles and the transition density as proposal; see Figure 3. 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).

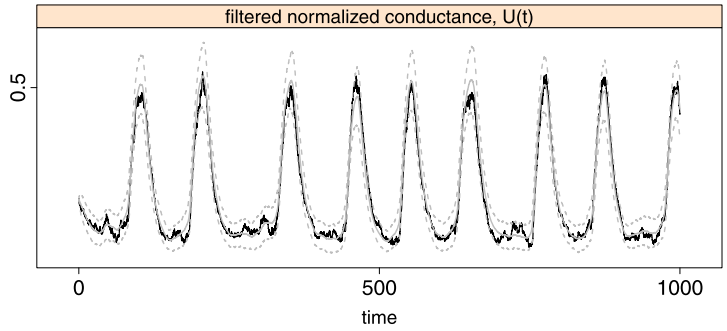

FIG. 3. *Filtering of* (U_t) *with the particle filter algorithm* (100 *particles*): *hidden simulated trajectory of the Morris–Lecar model* (U_t) (*black*), *mean filtered signal* (*grey full drawn line*), 95% *confidence interval of filtered signal* (*grey dashed lines*).

6.2. *Estimation results*. The performance of the SAEM–SMC algorithm is illustrated on 100 simulated trajectories. The SAEM algorithm is implemented with $m = 200$ iterations and a sequence (a_m) equal to 1 during the 100 first iterations and equal to $a_m = 1/(m - 100)^{0.8}$ for $m > 100$. The SMC algorithm is implemented with $K(m) = \min(m, 100)$ particles at each iteration of the SAEM algorithm. The SAEM algorithm is initialized by a random draw of $\widehat{\theta}_0$ not centered around the true value: $\widehat{\theta}_0 = \theta_{\text{true}} + 0.1 + \theta_{\text{true}}/3\mathcal{N}(0, 1)$.

An example of the convergence of the SAEM algorithm for one of the iterations is presented in Figure 4. It is seen that the algorithm converges for most of the parameters in a few iterations to a neighborhood of the true value, even if the initial values are far from the true ones. Only for ϕ more iterations are needed, which is expected since this parameter appears in the second, nonobserved coordinate.

The SAEM estimator is compared with the pseudo maximum likelihood estimator obtained if both V_t and U_t were observed. Results are given in Table 4. The parameters are well estimated in this ideal case. 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 Figure 4 we show boxplots of the estimates of the eight parameters for the three estimation settings; both coordinates observed or only one observed with σ fixed at either the true or a wrong value. All parameters appear well estimated. As expected, the variance of the estimator of ϕ hugely increases 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.

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 data sets. As an example, the SE for one data set estimated by SAEM are reported in the last line of Table 4. The estimated SE are satisfactory for most of the parameters, but tend to underestimate.

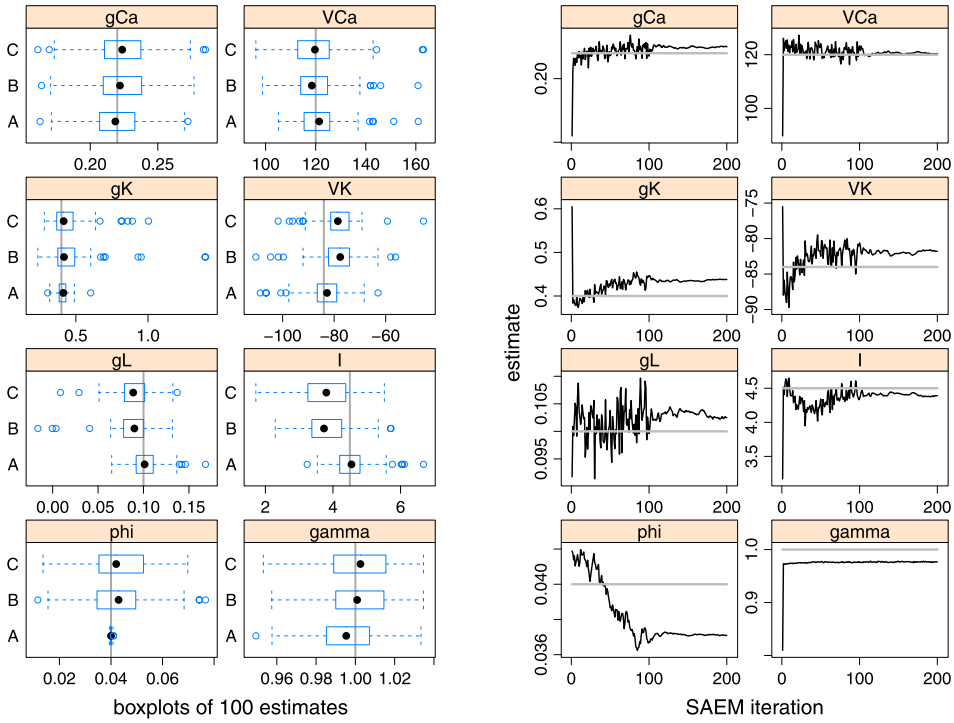

FIG. 4. *Estimation results for simulated data. Left panels*: *Boxplots of* 100 *estimates from simulated data sets for the* 8 *parameters. True values used in the simulations are given by the gray lines. A*: *Both* V_t *and* U_t *are observed. B*: *Only* V_t *is observed,* σ *is fixed at the true value* 0.03. *C*: *Only* V_t *is observed,* σ *is fixed at a wrong value* 0.04. *Right panels*: *Convergence of the SAEM algorithm for the* 8 *estimated parameters on a simulated data set. True values used in the simulation are given by the gray lines.*

7. Discussion. The main contributions of this paper are an algorithm to handle a more general model than a HMM and to show nonasymptotic convergence results for the method. It turns out that some of the common problems encountered with particle filters are not present in our case, namely, the filter does not degenerate, and we run the algorithm on large data sets of 6000 observations points in reasonable time (35 minutes on a standard portable computer for one of the simulated data sets).

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 does 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 a

TABLE 4

Simulation results obtained from 100 *simulated Morris–Lecar trajectories* ($n = 2000$, $\Delta = 0.1$ *ms*). *Two estimators are compared*: *The pseudo maximum likelihood estimator in the ideal case where both V_t and U_t are observed, and the 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 data set is also given*

	Parameter							
Estimator **True values:**	g_L **0.100**	g_{Ca} **0.220**	g_K **0.400**	γ **1.00**	V_K **−84.00**	ϕ **0.040**	V_{Ca} **120.00**	I **4.400**
With both V_t and U_t observed (pseudo maximum likelihood estimator)								
Mean	0.101	0.219	0.411	0.996	−83.20	0.040	121.97	4.539
RMSE	0.017	0.019	0.041	0.019	7.61	0.001	8.50	0.560
With only V_t observed (SAEM estimator)								
Mean	0.090	0.225	0.464	1.003	−78.622	0.041	119.677	4.060
RMSE	0.021	0.024	0.144	0.017	9.459	0.013	10.218	1.028
Estimated SE	0.016	0.019	0.042	0.016	4.96	0.001	7.31	0.561

small sampling step, that is, 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. Second, 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 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), and the number of SAEM iterations, needs further investigation. Finally, an automated procedure to find starting values for the procedure is warranted.

APPENDIX A: DISTRIBUTIONS OF APPROXIMATE MODEL

Consider the general approximate model [see (2)]

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

with ρ the correlation coefficient between the two Brownian motions or perturbations. The distribution of (V_{i+1}, U_{i+1}) conditionally on (V_i, U_i) is

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

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

$$\begin{aligned} V_{i+1}|V_i, U_i &\sim \mathcal{N}\big(V_i + \Delta f(V_i, U_i), \Delta(\gamma^2 + \rho^2)\big), \\ U_{i+1}|V_i, U_i &\sim \mathcal{N}\big(U_i + \Delta b(V_i, U_i), \Delta(\sigma^2(V_i, U_i) + \rho^2)\big). \end{aligned} \tag{8}$$

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

$$\begin{aligned} V_{i+1}|U_{i+1}, V_i, U_i &\sim \mathcal{N}(m_V, Var_V), \\ U_{i+1}|V_{i+1}, V_i, U_i &\sim \mathcal{N}(m_U, Var_U), \end{aligned} \tag{9}$$

where

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

The distributions in (8) and (9) are equal when the Brownian motions are independent, that is, when $\rho = 0$.

APPENDIX B: SUFFICIENT STATISTICS

We here provide the sufficient statistics of the approximate model (2). Consider the $n \times 6$-matrix

$$\begin{aligned} X = \big(&-V_{0:(n-1)}, -m_\infty(V_{0:(n-1)})V_{0:(n-1)}, \\ &- U_{0:(n-1)}V_{0:(n-1)}, U_{0:(n-1)}, \mathbf{1}, m_\infty(V_{0:(n-1)})\big), \end{aligned}$$

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

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

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

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

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

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

APPENDIX C: FISHER INFORMATION MATRIX

The standard errors (SE) of the parameter estimators can be evaluated from 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, Lavielle and Moulines (1999) and based on the Louis' missing information principle.

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

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

The derivatives $\partial_\theta L(S(V_{0:n}, U_{0:n}), \theta)$ and $\partial_\theta^2 L(S(V_{0:n}, U_{0:n}), \theta)$ 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:

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

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.

APPENDIX D: PROOF OF THE CONVERGENCE RESULTS

D.1. Convergence results of Proposition 1. We omit θ in the proof for clarity. The conditional expectation $\pi_n f$ is given by (5) and the kernels H_i from $\mathbb{R}$ into itself are defined by (4). We write $\nu_n = \mu H_1 \cdots H_n 1$ for the constant conditioned on the observed values $V_{0:n}$. Also, (4) is bounded, that is, $H_i 1(u) \le C$ for all $u \in \mathbb{R}$ and $i = 1, \dots, n$, for some constant C. It directly follows that $\mu H_1 \cdots H_{i-1} 1 \le C^{i-1}$. Furthermore, we obtain the bound

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

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

$$\frac{\nu_n}{C^{n-1}} \le \pi_{i-1} H_i 1 \le C. \tag{10}$$

Define the two empirical measures obtained at time i: $\Psi_i'^K = \frac{1}{K}\sum_{k=1}^K \mathbb{1}_{U_{0:i}'^{(k)}}$ and $\Psi_i^K = \sum_{k=1}^K W_i(U_{0:i}^{(k)}) \mathbb{1}_{U_{0:i}^{(k)}}$. 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, Jacod and Protter (2001)] for $\xi_1, \dots, \xi_K$ random variables, which conditioned on a σ-field $\mathcal{G}$ are independent, centered and bounded $|\xi_k| \le a$. Then for any $\varepsilon > 0$ we have

$$\mathbb{P}\left(\left|\frac{1}{K}\sum_{k=1}^K \xi_k\right| \ge \varepsilon\right) \le 2\exp\left(-K\frac{\varepsilon^2}{2a^2}\right). \tag{11}$$

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

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

fulfills the conditions for (11) 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 $\varepsilon > 0$,

$$\mathbb{P}\bigl(|\Psi_i'^K f - \Psi_i^K f| \ge \varepsilon\bigr) \le 2\exp\left(-K\frac{\varepsilon^2}{8\|f\|^2}\right). \tag{12}$$

Define $Q_i(f)(u) = \int q(u'|V_i, V_{i-1}, u) f(u')\,du'$. By definition of the unnormalized weights in step 3 of the SMC algorithm, $w_i(u,u') = p_\Delta(V_i, V_{i-1}, u, u')/p_\Delta(V_{i-1}, u) q(u'|V_i, V_{i-1}, u)$, so that $Q_i(f w_i)(u) = \int p_\Delta(V_i, u'|V_{i-1}, u) \times f(u')\,du' = H_i f(u)$. We therefore have

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

which fulfills the conditions for (11) to hold, now with $a = 2C\|f\|$ and $\mathcal{G}$ is the σ-algebra generated by $U_{0:i-1}^{\prime(k)}$, since $U_i^{(k)}$ is drawn from $q(\cdot|V_{i-1:i}, U_{i-1}^{\prime(k)})$; see step 2 of the SMC algorithm. Hence, for any $\varepsilon > 0$ we obtain

$$\mathbb{P}\big(\big|\Upsilon_i^K f - \Psi_{i-1}^{\prime K} H_i f\big| \geq \varepsilon\big) \leq 2\exp\bigg(-K\frac{\varepsilon^2}{8C^2\|f\|^2}\bigg). \tag{13}$$

We want to show the following two bounds:

$$\mathbb{P}\big(\big|\Psi_i^K f - \pi_i f\big| \geq \varepsilon\big) \leq 2I_i \exp\bigg(-K\frac{\varepsilon^2}{8J_i\|f\|^2}\bigg), \qquad i = 1, \ldots, n, \tag{14}$$

$$\mathbb{P}\big(\big|\Psi_i^{\prime K} f - \pi_i f\big| \geq \varepsilon\big) \leq 2I_i' \exp\bigg(-K\frac{\varepsilon^2}{8J_i'\|f\|^2}\bigg), \qquad i = 0, 1, \ldots, n, \tag{15}$$

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 $U_0^{\prime(k)}$ are i.i.d. with law μ, then (11) with $\xi_k = f(U_0^{\prime(k)}) - \mu(f)$ yields (15) for $i = 0$ with $I_i' = J_i' = 1$. Let $i \geq 1$ and assume (15) holds for $i - 1$. We can write

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

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 (10) to see that

$$\big|\Psi_i^K f - \pi_i f\big| \leq \frac{C^{n-1}}{\nu_n}\big(\|f\||L_i 1| + |L_i f|\big)$$

and

$$|L_i f| \leq \big|\Upsilon_i^K f - \Psi_{i-1}^{\prime K} H_i f\big| + \big|\Psi_{i-1}^{\prime K} H_i f - \pi_{i-1}H_i f\big|.$$

Assuming that (15) holds for $i - 1$ and using (13) and that $\|H_i f\| \leq C\|f\|$ yield

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

We obtain

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

Hence, (14) 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 (12) and (14) we then conclude that (15) 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, (6) 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. To prove the convergence of the SAEM–SMC algorithm, we study the stochastic approximation scheme used during the SA step. The scheme (7) can be decomposed into

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

with

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

where we denote by $\pi_{n,\theta} S = \mathbb{E}_\Delta(S(V_{0:n}, U_{0:n})|V_{0:n}; \theta)$ the expectation of the sufficient statistic S under the exact distribution $p_\Delta(U_{0:n}|V_{0:n}; \theta)$, and by $\Psi_{n,\hat{\theta}(s_m)}^{K(m)} S$ 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, Lavielle and Moulines (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 (7). Assertion 2 is proved by Lemma 2 of Delyon, Lavielle and Moulines (1999) under assumptions (M1)–(M5) and (SAEM2). Assertion 3 is proved similarly as Theorem 5 of Delyon, Lavielle and Moulines (1999). By construction of the SMC algorithm, the equivalent of assumption (SAEM3) is checked for the expectation taken under the approximate empirical measure $\Psi_{n;\widehat{\theta}_m}^{K(m)}$. Indeed, the assumption of independence of the nonobserved 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, $\mathbb{E}_\Delta^{K(m)}(f(U_{0:n}^{(m+1)})|\mathcal{F}_m) = \Psi_{n;\widehat{\theta}_m}^{K(m)} f$. 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 Proposition 1. Under assumptions (SMC2)–(SMC3) and assertion 1, Proposition 1 yields that for any $\varepsilon > 0$, there exist two constants C_1, C_2, independent of θ, such that

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

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

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

which is finite when $M \to \infty$, proving the a.s. convergence of r_m to 0.

D.3. Proof of Theorem 2. The Markov property yields

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

Gobet and Labart (2008) 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

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

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

$$\int |p(V_i, U_i|V_{i-1}, U_{i-1}; \theta) - p_\delta(V_i, U_i|V_{i-1}, U_{i-1}; \theta)| \times \prod_{j=1}^{i-1} p(V_j, U_j|V_{j-1}, U_{j-1}; \theta) \times \prod_{j=i+1}^{n} p_\delta(V_j, U_j|V_{j-1}, U_{j-1}; \theta)\, dU_{0:n} \le C\delta.$$

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

Acknowledgments. The authors are grateful to Rune W. Berg for making his experimental data available. We thank E. Gobet for helpful discussions about convergence of the Euler scheme. The work is part of the Dynamical Systems Interdisciplinary Network, University of Copenhagen.

REFERENCES

AÏT-SAHALIA, Y. (2002). Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation approach. *Econometrica* **70** 223–262. MR1926260

ANDRIEU, C., DOUCET, A. and PUNSKAYA, E. (2001). Sequential Monte Carlo methods for optimal filtering. In *Sequential Monte Carlo Methods in Practice* 79–95. Springer, New York. MR1847787

BERG, R. W., ALABURDA, A. and HOUNSGAARD, J. (2007). Balanced inhibition and excitation drive spike activity in spinal half-centers. *Science* **315** 390–393.

BERG, R. W. and DITLEVSEN, S. (2013). Synaptic inhibition and excitation estimated via the time constant of membrane potential fluctuations. *J. Neurophys.* **110** 1021–1034.

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

BESKOS, A., PAPASPILIOPOULOS, O., ROBERTS, G. O. and FEARNHEAD, P. (2006). Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **68** 333–382. MR2278331

BORG-GRAHAM, L. J., MONIER, C. and FRÉGNAC, Y. (1998). Visual input evokes transient and strong shunting inhibition in visual cortical neurons. *Nature* **393** 369–373.

CAPPÉ, O., MOULINES, E. and RYDÉN, T. (2005). *Inference in Hidden Markov Models*. Springer, New York. MR2159833

DELYON, B., LAVIELLE, M. and MOULINES, E. (1999). Convergence of a stochastic approximation version of the EM algorithm. *Ann. Statist.* **27** 94–128. MR1701103

DEL MORAL, P., JACOD, J. and PROTTER, P. (2001). The Monte-Carlo method for filtering with discrete-time observations. *Probab. Theory Related Fields* **120** 346–368. MR1843179

DEL MORAL, P. and MICLO, L. (2000). Branching and interacting particle systems approximations of Feynman–Kac formulae with applications to non-linear filtering. In *Séminaire de Probabilités, XXXIV. Lecture Notes in Math.* **1729** 1–145. Springer, Berlin. MR1768060

DEMPSTER, A. P., LAIRD, N. M. and RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B* **39** 1–38. MR0501537

DITLEVSEN, S. and GREENWOOD, P. (2013). The Morris–Lecar neuron model embeds a leaky integrate-and-fire model. *J. Math. Biol.* **67** 239–259. MR3071157

DOUC, R., GARIVIER, A., MOULINES, E. and OLSSON, J. (2011). Sequential Monte Carlo smoothing for general state space hidden Markov models. *Ann. Appl. Probab.* **21** 2109–2145. MR2895411

DOUCET, A., DE FREITAS, N. and GORDON, N. (2001). An introduction to sequential Monte Carlo methods. In *Sequential Monte Carlo Methods in Practice* 3–14. Springer, New York. MR1847784

DOUCET, A., GODSILL, S. and ANDRIEU, C. (2000). On sequential Monte Carlo sampling methods for Bayesian filterin. *Stat. Comput.* **10** 197–208.

DURHAM, G. B. and GALLANT, A. R. (2002). Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. *J. Bus. Econom. Statist.* **20** 297–338. With comments and a reply by the authors. MR1939904

ELERIAN, O., CHIB, S. and SHEPHARD, N. (2001). Likelihood inference for discretely observed nonlinear diffusions. *Econometrica* **69** 959–993. MR1839375

ERAKER, B. (2001). MCMC analysis of diffusion models with application to finance. *J. Bus. Econom. Statist.* **19** 177–191. MR1939708

FEARNHEAD, P., PAPASPILIOPOULOS, O. and ROBERTS, G. O. (2008). Particle filters for partially observed diffusions. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **70** 755–777. MR2523903

GERSTNER, W. and KISTLER, W. M. (2002). *Spiking Neuron Models*: *Single Neurons*, *Populations*, *Plasticity*. Cambridge Univ. Press, Cambridge. MR1923120

GOBET, E. and LABART, C. (2008). Sharp estimates for the convergence of the density of the Euler scheme in small time. *Electron. Commun. Probab.* **13** 352–363. MR2415143

GOLIGHTLY, A. and WILKINSON, D. J. (2006). Bayesian sequential inference for nonlinear multivariate diffusions. *Stat. Comput.* **16** 323–338. MR2297534

GOLIGHTLY, A. and WILKINSON, D. J. (2008). Bayesian inference for nonlinear multivariate diffusion models observed with error. *Comput. Statist. Data Anal.* **52** 1674–1693. MR2422763

GORDON, N. J., SALMOND, D. J. and SMITH, A. F. M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings-F* **140** 107–113.

HUYS, Q. J. M., AHRENS, M. and PANINSKI, L. (2006). Efficient estimation of detailed single-neuron models. *J. Neurophysiol.* **96** 872–890.

HUYS, Q. J. M. and PANINSKI, L. (2009). Smoothing of, and parameter estimation from, noisy biophysical recordings. *PLoS Comput. Biol.* **5** e1000379, 16. MR2516090

IONIDES, E. L., BHADRA, A., ATCHADÉ, Y. and KING, A. (2011). Iterated filtering. *Ann. Statist.* **39** 1776–1802. MR2850220

JAHN, P., BERG, R. W., HOUNSGAARD, J. and DITLEVSEN, S. (2011). Motoneuron membrane potentials follow a time inhomogeneous jump diffusion process. *J. Comput. Neurosci.* **31** 563–579. MR2864731

JENSEN, A. C., DITLEVSEN, S., KESSLER, M. and PAPASPILIOPOULOS, O. (2012). Markov chain Monte Carlo approach to parameter estimation in the FitzHugh–Nagumo model. *Physical Review E* **86** 041114.

KALOGEROPOULOS, K. (2007). Likelihood-based inference for a class of multivariate diffusions with unobserved paths. *J. Statist. Plann. Inference* **137** 3092–3102. MR2364153

KLOEDEN, P. and NEUENKIRCH, A. (2012). Convergence of numerical methods for stochastic differential equations in mathematical finance. Available at arXiv:1204.6620v1 [math.NA].

KÜNSCH, H. R. (2005). Recursive Monte Carlo filters: Algorithms and theoretical analysis. *Ann. Statist.* **33** 1983–2021. MR2211077

LIU, J. S. and CHEN, R. (1998). Sequential Monte Carlo methods for dynamic systems. *J. Amer. Statist. Assoc.* **93** 1032–1044. MR1649198

LONGTIN, A. (2013). Neuronal noise. *Scholarpedia* **8** 1618.

MONIER, C., FOURNIER, J. and FRÉGNAC, Y. (2008). In vitro and in vivo measures of evoked excitatory and inhibitory conductance dynamics in sensory cortices. *J. Neurosci. Methods* **169** 323–365.

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

PEDERSEN, A. R. (1995). A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. *Scand. J. Stat.* **22** 55–71. MR1334067

POKERN, Y., STUART, A. M. and WIBERG, P. (2009). Parameter estimation for partially observed hypoelliptic diffusions. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **71** 49–73. MR2655523

POSPISCHIL, M., PIWKOWSKA, Z., BAL, T. and DESTEXHE, A. (2009). Extracting synaptic conductances from single membrane potential traces. *Neurosci.* **158** 545–552.

RINZEL, J. and ERMENTROUT, G. B. (1989). *Methods in Neural Modeling*, *Chapter Analysis of Neural Excitability and Oscillations*. MIT Press, Cambridge, MA.

ROBERTS, G. O. and STRAMER, O. (2001). On inference for partially observed nonlinear diffusion models using the Metropolis–Hastings algorithm. *Biometrika* **88** 603–621. MR1859397

RUDOLPH, M., PIWKOWSKA, Z., BADOUAL, M., BAL, T. and DESTEXHE, A. (2004). A method to estimate synaptic conductances from membrane potential fluctuations. *J. Neurophysiol.* **91** 2884–2896.

SAMSON, A. and THIEULLEN, M. (2012). A contrast estimator for completely or partially observed hypoelliptic diffusion. *Stochastic Process. Appl.* **122** 2521–2552. MR2926166

SØRENSEN, H. (2004). Parametric inference for diffusion processes observed at discrete points in time: A survey. *Int. Stat. Rev.* **72** 337–354.

SØRENSEN, M. (2012). Estimating functions for diffusion-type processes. In *Statistical Methods for Stochastic Differential Equations*. *Monogr. Statist. Appl. Probab.* **124** 1–107. CRC Press, Boca Raton, FL. MR2976982

TATENO, T. and PAKDAMAN, K. (2004). Random dynamics of the Morris–Lecar neural model. *Chaos* **14** 511–530. MR2089477

DEPARTMENT OF MATHEMATICAL SCIENCES
UNIVERSITY OF COPENHAGEN
UNIVERSITETSPARKEN 5
DK-2100 COPENHAGEN
DENMARK
E-MAIL: susanne@math.ku.dk
URL: http://www.math.ku.dk/~susanne
DYNAMICAL SYSTEMS INTERDISCIPLINARY NETWORK
URL: http://dsin.ku.dk

CNRS UMR8145
LABORATOIRE MAP5
UNIVERSITÉ PARIS DESCARTES
45 RUE DES SAINTS PÈRES
75006 PARIS
FRANCE
AND
LABORATOIRE JEAN KUNTZMANN UMR
CNRS 5224
UNIVERSITÉ GRENOBLE-ALPES
51 RUE DES MATHÉMATIQUES
38041 GRENOBLE CEDEX 9
FRANCE
E-MAIL: Adeline.leclercq-Samson@imag.fr
URL: http://adeline.e-samson.org/en/