Markov chain Monte Carlo approach to parameter estimation in the FitzHugh-Nagumo model

Anders Chr. Jensen* and Susanne Ditlevsen
Department of Mathematical Sciences, University of Copenhagen, Copenhagen, Denmark

Mathieu Kessler
Departamento de Matemática Aplicada y Estadística, Universidad Politécnica de Cartagena, Cartagena, Spain

Omiros Papaspiliopoulos
ICREA and Department of Economics, Universitat Pompeu Fabra, Barcelona, Spain

(Received 23 May 2012; revised manuscript received 6 September 2012; published 10 October 2012)

Excitability is observed in a variety of natural systems, such as neuronal dynamics, cardiovascular tissues, or climate dynamics. The stochastic FitzHugh-Nagumo model is a prominent example representing an excitable system. To validate the practical use of a model, the first step is to estimate model parameters from experimental data. This is not an easy task because of the inherent nonlinearity necessary to produce the excitable dynamics, and because the two coordinates of the model are moving on different time scales. Here we propose a Bayesian framework for parameter estimation, which can handle multidimensional nonlinear diffusions with large time scale separation. The estimation method is illustrated on simulated data.

DOI: 10.1103/PhysRevE.86.041114 PACS number(s): 02.50.−r, 05.10.−a, 87.19.L−, 87.10.−e

I. INTRODUCTION

An excitable system is characterized by a resting state from which it only escapes if perturbed by a sufficiently large stimulus. Weak stimuli only result in a small amplitude linear response, whereas strong stimuli cause a highly nonlinear response, where the system variables make a large excursion through state space called firing, whereafter it returns to its resting state after a refractory period. Under a continuous stimulus, the system can enter into an oscillatory mode. Thus, an excitable system operates close to a bifurcation point, and is sensitive to small perturbations, e.g., caused by noise. It is observed in many natural systems, such as neuronal dynamics, ion channels, chemical reactions, climate dynamics, or wildfires [1–3]. Noise can have a dramatic effect on excitable systems, inducing stochastic limit cycles on otherwise stable dynamics. A prototype of an excitable system is the FitzHugh-Nagumo (FHN) model, a minimal representation of more realistic excitable systems, such as the Hodgkin-Huxley model, modeling the firing mechanisms in a neuron [4–6]. It is a generalization of the van der Pol equations. It allows for coordinates to evolve on different time scales, and the time scale separation parameter is essential for the understanding of the dynamical behavior of the system. The larger the time scale separation, the more an all-or-nothing response is observed to a perturbation, mimicking the response of the simpler threshold models, such as leaky integrate-and-fire models [7,8].

The FHN model is defined by two coupled differential equations, representing the neuronal membrane potential and a recovery variable, respectively, where the recovery variable models the channel kinetics. Extending the model by adding a noise term governed by Brownian motion results in a diffusion process. The noise term in the FHN model accounts for various sources of noise affecting the neuronal behavior, such as random opening and closing of ion channels or noisy presynaptic currents [7].

Diffusions are defined through a stochastic differential equation, and for all but a few models an explicit expression for the transition density is unattainable. This problem complicates parameter estimation. Though many methods deal with this problem (see Ref. [9] and also Refs. [10,11]), they tend to be highly complicated to implement and apply in practice, especially when the dimension of the diffusion is larger than one. The Euler-Maruyama, the Milstein, and other schemes offer easy-to-implement approximations to the transition density. However, if the time step between observations is too large, the approximation will be inaccurate.

Within the past decade, novel Bayesian methods have been developed which can be used for statistical inference (see Refs. [12–16]). We describe a Markov chain Monte Carlo method and adapt it to the two-dimensional stochastic FHN model for parameter inference. Our approach involves imputation of data from the distribution of the underlying diffusion process, and application of a Gibbs sampler to iteratively update parameters and imputed data. We apply an independent Metropolis-Hastings (MH) step to update the imputed data conditional on parameters, and sample the parameters conditional on the imputed data directly. Parameter sampling relies on a Gaussian prior for the parameters, and when this assumption is not met, a Gaussian random walk MH step may be applied (see Appendix B for details).

Typically, in experimental settings only the slow variable of the membrane potential is observed through intracellular recordings in single neurons, whereas the channel kinetics are unobserved. This largely complicates the statistical inference, e.g., because the observed process is no longer Markovian. One approach to this problem is to assume the channel kinetics known [17,18]. We will not assume the channel kinetics known, but instead assume that the recovery variable is observed.

*anders@math.ku.dk

Our methodology may be extended to the partially observed case and this is work in progress. However, the first goal, which is achieved in this paper, is to make the statistical procedure work in a computationally efficient manner in the two-dimensional nonlinear model with time scale separation.

The effects of noise on the FHN model have been extensively studied (see, e.g., Refs. [1,3,19–21]), whereas papers devoted to its comparison with experimental data are rare. In this paper we use simulated data to estimate parameters of the stochastic FHN model from discrete observations of the state variables.

In Sec. II we introduce the deterministic FHN model and a stochastic extension, and Sec. III describes the estimation procedure in the case where the diffusion coefficient is assumed to be known. This approach simplifies the exposition and speeds up the practical implementation considerably. Section IV deals with the procedure when also the diffusion coefficient is estimated, and Sec. V includes a small simulation study.

II. THE FITZHUGH-NAGUMO MODEL

Consider the FHN model [19] (see also Refs. [4,5,7,22]),

$$\frac{d}{dt}x_t = \frac{1}{\varepsilon}\left(x_t - x_t^3 - y_t + s\right), \tag{1}$$

$$\frac{d}{dt}y_t = \gamma x_t - y_t + \beta, \tag{2}$$

with $\varepsilon > 0$. In the modeling of neuronal spike generation in axons, x describes the membrane potential and y is a recovery variable. The parameter ε is a time scale separator, typically smaller than one, such that x is the fast and y is the slow variable. Furthermore, s denotes the input current and γ and β determine the locations of the fixed point(s).

For the model (1) and (2), the nullclines are cubic and linear, respectively,

$$y = x - x^3 + s, \quad y = \gamma x + \beta,$$

and in general there exist at most three fixed points. For $\gamma > 1$ there is exactly one fixed point which is stable or unstable, depending on the parameter values. Figure 1 shows phase and time plots for two sets of parameter values, leading to excitatory and oscillatory behavior, respectively. In both cases there is only one fixed point. The parameters in the upper panels are chosen such that the fixed point is stable and the model spikes one time and then relaxes to the resting state and stays there. In the lower panels the fixed point is unstable and a limit cycle with spikes appears. A detailed exposition of the dynamics of the model can be found in Refs. [7,22].

A. Stochastic extension

We include additive noise in both coordinates and obtain the following stochastic model:

$$dX_t = \frac{1}{\varepsilon}\left(X_t - X_t^3 - Y_t + s\right)dt + \sigma_1\, dB_t^{(1)}, \tag{3}$$

$$dY_t = (\gamma X_t - Y_t + \beta)\,dt + \sigma_2\, dB_t^{(2)}, \tag{4}$$

where $(B_t^{(1)}, B_t^{(2)})^T$ is a two-dimensional standard Brownian motion and $t \in [0,T]$,$(X_0, Y_0) = (x_0, y_0)$. The parameters of the model are $(\theta,\sigma)^T$ with the drift parameter $\theta = (\varepsilon,s,\gamma,\beta)^T \in \mathbb{R}_+ \times \mathbb{R}^3$ and diffusion parameter $\sigma = (\sigma_1,\sigma_2)^T \in \mathbb{R}_+^2$. The qualitative behavior of the stochastic model is different from the deterministic model because the random perturbations can affect the system and lead to emergence of repeated spiking activity, also when the fixed point is stable (see Fig. 2).

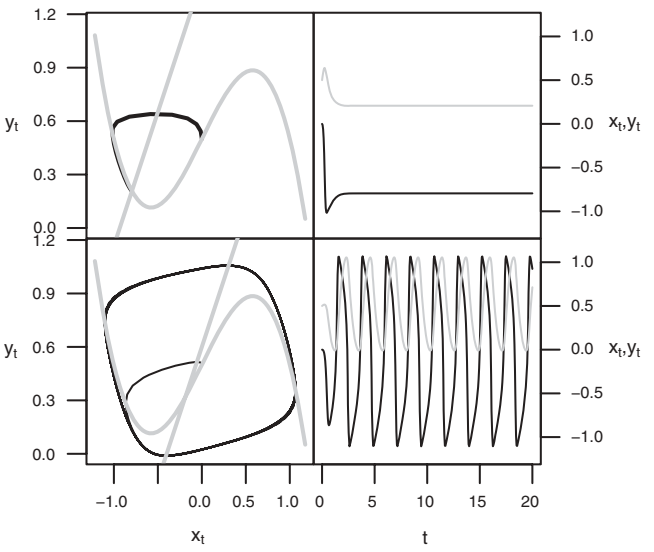

FIG. 1. Deterministic FHN model. Left: Phase portraits with x and y nullclines (gray) and simulated trajectory (black). Right: Time plots of x (black) and y (gray). For all plots $\varepsilon = 0.1$, $s = 0.5$, $\gamma = 1.5$. Top: $\beta = 1.4$, excitatory behavior, the fixed point is stable. Bottom: $\beta = 0.6$, oscillatory behavior, the fixed point is unstable.

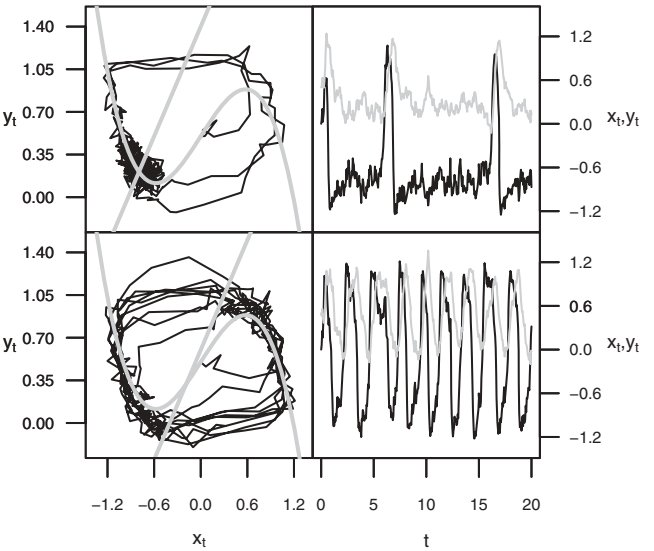

FIG. 2. Stochastic FHN model. Left: Phase portraits with x and y nullclines (gray) and simulated trajectory (black). Right: Time plots of x (black) and y (gray). For all plots $\varepsilon = 0.1$, $s = 0.5$, $\gamma = 1.5$, $\sigma_1 = 0.5, \sigma_2 = 0.3$. Top: $\beta = 1.4$, excitatory behavior, the fixed point is stable. Bottom: $\beta = 0.6$, oscillatory behavior, the fixed point is unstable.

B. Statistical model

In the following we shall state the model in more general terms, as

$$dV_t = b(V_t,\theta)dt + \Sigma(\sigma)dB_t, \quad V_0 = v_0, \tag{5}$$

with $V_t = (X_t,Y_t)^T$ a two-dimensional stochastic process, B_t a two-dimensional standard Brownian motion, and functions b and Σ taking values in $\mathbb{R}^2$ and the set of two-by-two matrices, respectively. We will assume that $\Gamma := \Sigma\Sigma^T$ is invertible and that b is linear in the drift parameters θ:

$$b(V_t,\theta) = f_0(V_t) + \sum_{i=1}^{k} \theta_i f_i(V_t), \tag{6}$$

with f_i a 2×1 vector.

We assume equidistant observations

$$D_n = \{V_{t_0}, V_{t_1}, \ldots, V_{t_n}\},$$

with $t_i - t_{i-1} = \Delta$, and $t_0 = 0, t_n = T$, and aim to make inference for the parameters θ and σ governing the diffusion (5). The observation times are assumed equidistant only for notational simplicity and because it is consistent with the type of experimental design for which the analysis in this paper is relevant for; it will be clear that our methodology does not require this assumption.

Note that for the FHN model (3) and (4), the reparametrization

$$(\tilde{\varepsilon},\tilde{s},\gamma,\beta) = (1/\varepsilon, s/\varepsilon, \gamma, \beta) \tag{7}$$

makes the model linear in the drift parameters. For models where the drift is not linear in θ, the estimation procedure we describe will still work, at the cost of an additional MH step, and this approach is described in Appendix B.

III. ESTIMATION OF DRIFT PARAMETERS WITH KNOWN DIFFUSION

The aim of this section is to estimate the parameter vector θ governing the drift, while σ is assumed known. This assumption simplifies the exposition, while at the same time it provides the foundation of the methodology used in the setting where also the diffusion is unknown. For notational simplicity σ is neglected in this section.

Since the model is Markovian, the distribution of θ conditional on the observed data D_n is

$$p(\theta \mid D_n) \propto p(\theta)p(D_n \mid \theta) = p(\theta)\prod_{i=1}^{n} p\left(V_{t_i} \mid V_{t_{i-1}},\theta\right), \tag{8}$$

where $p(\theta)$ is the prior distribution of θ, $p(D_n \mid \theta)$ is the distribution of the observed data given θ, and $p(V_{t_i} \mid V_{t_{i-1}},\theta)$ is the transition density of the process V, from $V_{t_{i-1}}$ to V_{t_i}, conditional on θ. Following standard convention in Bayesian statistics, proportionality is understood with respect to the argument of the density on the left hand side of the equation, i.e., θ in the above equation. For all but a few models, the transition density is not explicitly known and this is indeed a problem in the FHN model. To overcome this difficulty, consider a theoretical data augmentation step that imputes a latent data path $\bar{V}_i$ on each interval (t_i,t_{i+1}), distributed according to the underlying model (5). Denote the collection of latent paths $\bar{V} := \cup_{i=0}^{n-1} \bar{V}_i$, and change focus to the posterior of θ and the imputed data conditional on the observed data:

$$p(\theta,\bar{V} \mid D_n). \tag{9}$$

A Gibbs sampler can be applied to obtain a sample from (9) from which marginal inference about the posterior of θ can be drawn. The algorithm alternates between updating the parameter and the latent data while keeping the other fixed. After a suitable burn-in period, the result is a sample $(\theta^{(k)},\bar{V}^{(k)})_k$ from the distribution $p(\theta,\bar{V} \mid D_n)$, from which a summarized inference about the posterior of θ can be drawn, e.g., mean, variance, and tail probabilities. The algorithm is given in Table I.

TABLE I. Gibbs sampler for $p(\theta,\bar{V} \mid D_n)$.

Initialize
(1) Initialize $\theta^{(0)}$ and imputed data $\bar{V}^{(0)}$.
Iterate
At iteration k:
(2a) Sample $\bar{V}^{(k)}$ from $p(V \mid \theta^{(k-1)},D_n)$.
(2b) Sample $\theta^{(k)}$ from $p(\theta \mid \bar{V}^{(k)},D_n)$.

A few remarks are in order: On the theoretical level the imputed paths $\bar{V}$ are infinite dimensional. Therefore, each path is projected onto a discrete subset of the continuous path, and the accuracy of the algorithm depends to some extent on the accuracy of this approximation. The paths sampled from step (2a) in Table I are conditioned on the observed data D_n, and therefore a method for simulation of processes conditional on both start and endpoint (diffusion bridges) is needed. Simulating a diffusion bridge is in general not an easy task though much progress in this area has been achieved in the past decade; see Refs. [13,14].

A. Sampling from $p(\bar{V} \mid \theta, D_n)$

Due to the Markovian property, the following relation holds,

$$p(\bar{V} \mid \theta,D_n) = \prod_{i=1}^{n} p\left(\bar{V}_i \mid V_{t_{i-1}},V_{t_i},\theta\right), \tag{10}$$

implying that paths $\bar{V}_i$ may be sampled independently. Direct simulation of these bridges (or the projection thereof) is not feasible. Instead, we follow Ref. [12] and use an MH step for this simulation, whereby we propose paths from an alternative simpler model and accept them with the appropriate probability. Without loss of generality, focus on a single term of the form

$$p(\bar{V} \mid V_0,V_T,\theta) \tag{11}$$

from (10). This is a diffusion bridge, and in Ref. [12] it is shown that one can simulate these bridges with an MH step, using proposals from a Brownian bridge. Samples from the latter are easily generated, since the transition density of a Brownian bridge is for $t > s$,

$$p(V_t \mid V_s = v_s, V_0 = v_0, V_T = v_T) \sim \varphi\left(V_t \,\middle|\, v_s + \frac{t-s}{T-s}(v_T - v_s); \frac{(T-t)(t-s)}{T-s}\Gamma\right), \tag{12}$$

TABLE II. MH step for sampling from the diffusion bridge (11).

Initialize
(1) Initialize a skeleton path $(\bar{V}^M)_0$ according to (12), and compute the weight $\psi_0 = \psi((\bar{V}^M)_0,\theta)$, using (13).
Iterate
(2) Generate a proposal skeleton path, $\tilde{V}^M$ according to (12), and compute the weight $\tilde{\psi} = \psi(\tilde{V}^M,\theta)$, using (13).
(3) Let $(\bar{V}^M)_{k+1} = \begin{cases} \tilde{V}^M & \text{with prob. } \min\left(1, \frac{\tilde{\psi}}{\psi_k}\right), \\ (\bar{V}^M)_k & \text{otherwise.} \end{cases}$

where $\varphi(\cdot \mid \mu;\Omega)$ denotes the Gaussian density with mean μ and covariance matrix Ω. To each proposal a weight is assigned, the logarithm of which is given by

$$\log[\psi(\bar{V},\theta)] = \int_0^T b(\bar{V}_u,\theta)^T \Gamma^{-1} d\bar{V}_u - \frac{1}{2}\int_0^T b(\bar{V}_u,\theta)^T \Gamma^{-1} b(\bar{V}_u,\theta) du. \tag{13}$$

This is precisely (proportional to) the Radon-Nikodym derivative between the target diffusion bridge measure and the proposal Brownian bridge measure. The algorithm for simulating the bridge $\bar{V}$ is given in Table II. Note that step (2) in Table II involves an approximation of the integral in Eq. (13). More details on diffusion bridge simulation can be found in Ref. [13].

B. Sampling from $p(\theta \mid \bar{V}, D_n)$

After the reparametrization (7), the FHN model is linear in the drift parameter, and it can be written on the form (6). In this case, taking a Gaussian prior, the prior and the posterior distribution will be conjugate, i.e., the prior and the posterior belong to the same family of distributions. To see this, note that

$$p(\theta \mid \bar{V}, D_n) \propto p(\theta) p(V \mid \theta),$$

where V denotes the union of observed and imputed data. It follows directly from the Cameron-Martin-Girsanov theorem (see, e.g., Ref. [13]) that $p(V \mid \theta) = \psi(V,\theta)$, which, when the drift is as in Eq. (6), is exponentially quadratic in θ. We have

$$\log[\psi(\theta,V)] = -\tfrac{1}{2}(\theta^T R \theta + 2\theta^T F + R_{00} - 2I_0),$$

where $R = (R_{ij})_{i,j\geqslant 1}$, $F = (F_i)_{i\geqslant 1}$ and

$$R_{ij} = \int_0^T f_i(V_u)^T \Gamma^{-1} f_j(V_u) du, \tag{14}$$

$$I_i = \int_0^T f_i(V_u)^T \Gamma^{-1} dV_u, \quad F_i = I_i - R_{i0}, \quad i,j \geqslant 0 \tag{15}$$

Thus $p(\theta \mid \bar{V}, D_n)$ must be Gaussian. Assume that $p(\theta) = \varphi(\theta \mid \mu;\Phi)$ with Φ diagonal, such that the individual θ parameters *a priori* are independent. Then the posterior distribution of θ is Gaussian:

$$p(\theta \mid \bar{V}, D_n) \sim \varphi(\theta \mid (R+\Phi^{-1})^{-1}(F+\Phi^{-1}\mu); \quad (R+\Phi^{-1})^{-1}). \tag{16}$$

Having identified (16) is highly appealing, since it allows for direct sampling from $p(\theta \mid \bar{V}, D_n)$. In practice, the integrals in Eqs. (14) and (15) are approximated by Riemann sums, and the accuracy will depend on the sparsity of the imputed data.

In models where the assumptions of a Gaussian prior or linearity in the drift parameters are not met, the distribution $p(\theta \mid \bar{V}, D_n)$ could be approximated by an MH step. This approach is described in Appendix B.

Note that ε is assumed positive, therefore in principle the Gaussian prior should be truncated at 0. However, for small ε the effect of the truncation is negligible and can be omitted.

IV. ESTIMATION OF BOTH DRIFT AND DIFFUSION PARAMETERS

When estimating diffusion parameters an important point related to the dependence between parameters and imputed data must be made. The quadratic variation identity implies that for any $t > 0$,

$$\lim_{M\to\infty} \sum_{i=1}^{M} (V_{ti/M} - V_{t(i-1)/M})(V_{ti/M} - V_{t(i-1)/M})^T = t\,\Sigma(\sigma)\Sigma(\sigma)^T \quad \text{in probability.}$$

Thus, an observed path of V completely identifies σ. When dealing with discrete-time observations D_n, there is only finite information about σ, hence there will be statistical errors associated with its estimation. However, the identity implies that we cannot hope to apply a Gibbs sampler which iteratively would update paths and σ. Any value of σ would generate a path whose quadratic variation would return exactly the same value for σ, hence it will be impossible to explore the posterior distribution of σ in this way. Of course, in practice we only generate finite-dimensional projections of the paths, hence we would not observe this reducible behavior. Nevertheless, it is obvious, and actually proved in Ref. [12], that the Gibbs sampler which updates σ and a projection of the path based on M intermediate values for each pair of observations has a mixing time which is $O(M)$. Therefore it becomes worse as we reduce the approximation bias.

However, this problem is easy to overcome by a simple transformation as in Ref. [12]. The original article describes it for one-dimensional diffusions, but the extension is immediate for the multidimensional setting we consider here: We apply the one-to-one transformation

$$x \mapsto \Sigma^{-1}(\sigma)x$$

to the process V, and obtain a new diffusion Z which has the form

$$dZ_t = \alpha(Z_t,\theta,\sigma)dt + dB_t, \quad Z_0 = \Sigma^{-1}(\sigma)V_0, \tag{17}$$

where $\alpha(Z_t,\theta,\sigma) = \Sigma^{-1}(\sigma)b(\Sigma(\sigma)Z_t,\theta)$.

Sampling $\bar{V}$ is equivalent to sampling Z conditionally on $Z_{t_{i-1}} = \Sigma^{-1}(\sigma)V_{t_{i-1}}$ and $Z_{t_i} = \Sigma^{-1}(\sigma)V_{t_i}$, and the quadratic variation of Z is now independent of σ. However, there is again a perfect dependence between Z and σ via the endpoints of Z: For given σ, Z has σ-dependent endpoints, which then perfectly determine σ in the following iteration. Therefore we need to remove the dependence on the endpoints as well.

Define $\tilde{V}$ as

$$\tilde{V}_t = Z_t - \left(1 - \frac{t - t_{i-1}}{\Delta}\right) Z_{t_{i-1}} - \frac{t - t_{i-1}}{\Delta} Z_{t_i} \qquad (18)$$

for $t_{i-1} \leqslant t \leqslant t_i$. Note that $\tilde{V}_{t_{i-1}} = \tilde{V}_{t_i} = 0$, and $\bar{V}$ can be reconstructed from $\tilde{V}$ and σ by inverting the two transformations: adding first the endpoints to obtain Z and scaling by $\Sigma(\sigma)$ to obtain $\bar{V}$. To understand the intuition behind this transformation, consider the measure of the process in Eq. (17), without the drift α, but conditional on $Z_{t_{i-1}}$ and Z_{t_i}. Under this measure, Z is a Brownian bridge starting and ending at $Z_{t_{i-1}}$ and Z_{t_i}, respectively. Tilting it linearly as in Eq. (18) makes $\tilde{V}$ a standard Brownian bridge. This construction effectively allows us to sample Z from (17), using proposals from a Brownian bridge.

Next we describe the individual steps for the Gibbs sampler. We apply it to $(\theta,\sigma,\tilde{V})$ and not $(\theta,\sigma,\bar{V})$. This approach will ensure that $\tilde{V}$ and (θ,σ) are independent under the proposal for updating $\tilde{V}$, and circumvent the problem with reducible behavior of the Gibbs sampler. The parameters θ and σ are updated separately in order to take advantage of the simple Gaussian conditional posterior for θ.

A. Sampling from $p(\tilde{V} \mid \theta,\sigma,D_n)$

As in Sec. III A we write $p(\tilde{V} \mid \theta, D_n)$ as a product of densities and for simplicity we focus on a single term of the form $p(\tilde{V} \mid V_0,V_T,\theta,\sigma)$. Continuing along the lines of Sec. III A we note that $p(\tilde{V} \mid V_0,V_T,\theta,\sigma)$ can be simulated using an MH step with proposals from a Brownian bridge. Thus for each proposal $\tilde{V}$ we assign a weight ϕ given by

$$\log[\phi(\tilde{V},\theta,\sigma)] = \int_0^T \alpha(Z_s,\theta,\sigma)^T \, dZ_s - \frac{1}{2}\int_0^T \alpha(Z_s,\theta,\sigma)^T \alpha(Z_s,\theta,\sigma) ds, \qquad (19)$$

where $\tilde{V}$ and Z are linked by the relation (18).

The algorithm for updating the bridge $\tilde{V}$ is given in Table III.

B. Sampling from $p(\theta \mid \tilde{V}, D_n,\sigma)$

Sampling from $p(\theta \mid \tilde{V}, D_n,\sigma)$ is carried out in complete analogy to Sec. III B and (16), using the process $\Sigma(\sigma)Z_t$ instead of $\bar{V}_t$.

TABLE III. MH step for sampling from the diffusion bridge (17).

Initialize
(1) Initialize a skeleton path $(\tilde{V}^M)_0$, sampled as a standard Brownian bridge starting and ending at 0. Compute Z according to (18) and approximate the weight ϕ in Eq. (19), w_0.
Iterate
(2) Generate a proposal skeleton path, $\tilde{V}_P^M$ sampled as a standard Brownian bridge starting and ending at 0. Compute $\tilde{Z}$ according to (18) and approximate the weight ϕ in Eq. (19), $\tilde{w}$.
(3) Let $(\tilde{V}^M)_{k+1} = \begin{cases} \tilde{V}_P^M & \text{with prob. } \min\left(1, \frac{\tilde{w}}{w_k}\right), \\ (\tilde{V}^M)_k & \text{otherwise.} \end{cases}$

TABLE IV. MH step for sampling from $p(\sigma \mid \tilde{V}, D_n,\theta)$.

Initialize
(1) Initialize $\bar{\sigma}^{(0)}$.
Iterate
(2a) Generate a proposal $\tilde{\sigma}$ from $\varphi(\cdot \mid \bar{\sigma}^{(k)};\Omega)$.
(2b) Let $\bar{\sigma}^{(k+1)} = \begin{cases} \tilde{\sigma} & \text{with prob. } \min\left(1, \frac{s(\tilde{\sigma})}{s(\bar{\sigma}^{(k)})}\right), \\ \bar{\sigma}^{(k)} & \text{otherwise.} \end{cases}$

C. Sampling from $p(\sigma \mid \tilde{V}, D_n,\theta)$

The priors of θ and σ are assumed to be independent, and therefore

$$p(\sigma \mid \tilde{V}, D_n,\theta) \propto p(\tilde{V} \mid \theta,\sigma,D_n) p(D_n \mid \theta,\sigma) p(\sigma). \qquad (20)$$

Now define

$$\log[\phi_i(\tilde{V},\theta,\sigma)] = \int_{t_{i-1}}^{t_i} \alpha(Z_s,\theta,\sigma)^T \, dZ_s - \frac{1}{2}\int_{t_{i-1}}^{t_i} \alpha(Z_s,\theta,\sigma)^T \alpha(Z_s,\theta,\sigma) ds.$$

Using the identity (4.22) from Ref. [13] yields

$$p(\tilde{V} \mid \theta,\sigma,D_n) = \prod_{i=1}^n \frac{\varphi\left(\Sigma^{-1}(\sigma)V_{t_i} \mid \Sigma^{-1}(\sigma)V_{t_{i-1}}; \Delta I_2\right)}{p_{t_{i-1},t_i}\left(\Sigma^{-1}(\sigma)V_{t_{i-1}}, \Sigma^{-1}(\sigma)V_{t_i}\right)} \phi_i(\tilde{V},\theta,\sigma), \qquad (21)$$

where p is the transition density of (5), and I_2 is the two-dimensional identity matrix. Furthermore, with a change of variables,

$$p(D_n \mid \theta,\sigma) = |\det[\Sigma^{-1}(\sigma)]|^n \prod_{i=1}^n p_{t_{i-1},t_i}\left(\Sigma^{-1}(\sigma)V_{t_{i-1}}, \Sigma^{-1}(\sigma)V_{t_i}\right),$$

so we obtain

$$p(\sigma \mid \tilde{V}, D_n,\theta) \propto p(\sigma)|\det[\Sigma^{-1}(\sigma)]|^n \cdot \prod_{i=1}^n \varphi\left(\Sigma^{-1}(\sigma)V_{t_i} \mid \Sigma^{-1}(\sigma)V_{t_{i-1}}; \Delta I_2\right) \times \phi_i(\tilde{V},\theta,\sigma), \qquad (22)$$

which can be evaluated using a Riemann approximation of ϕ_i. Applying an MH step to a sample from the distribution proportional to (22) is straightforward. We use a Gaussian random walk, on the transformed diffusion parameter $\bar{\sigma} = \log(\sigma)$, in order to account for the restriction to positive values in the original parametrization. Therefore, define the proportional target function $s(\log(\sigma))$ as the right hand side of (22), and let $\varphi(\cdot \mid \bar{\sigma}^{(k)};\Omega)$ be the Gaussian proposal distribution, used to propose an update of $\bar{\sigma}$ from $\bar{\sigma}^{(k)}$ to $\bar{\sigma}^{(k+1)}$. The MH step is summarized in Table IV.

V. SIMULATION STUDY FOR THE FHN MODEL

Six data sets were generated, with parameter values resembling the situation in Fig. 2 with excitatory ($\beta = 1.4$) and oscillatory ($\beta = 0.6$) behavior, respectively. The choice of parameter values for the excitatory data is inspired by the values used in Ref. [19]. Parameters are given in Table V.

TABLE V. Parameter values for simulation study.

	ε	s	γ	β	σ_1	σ_2
Oscillatory	0.1	0.5	1.5	0.6	0.5	0.3
Excitatory	0.1	0.5	1.5	1.4	0.5	0.3

Data was generated by thinning simulations from an Euler-Maruyama scheme of the FHN model (3) and (4), with 20 000 observations and a time step of 0.001.

Two data sets were generated using subsamples from the FHN data for every 100th observation such that the sample size was $n = 200$ and the time step between consecutive observations was $\Delta = 0.1$, implying a sample interval length of 20 time units. To investigate large sample properties, an additional data set was created for the excitatory setting, using subsamples for every tenth observation, leading to a step size of $\Delta = 0.01$.

Finally, three data sets were generated to evaluate the estimation procedure for different values of ε. We used ε equal to 0.5, 0.05, 0.01 and sample size $n = 200$ and $\Delta = 0.1$. All other parameters were as in the excitatory data.

For all six data sets, four data points were imputed between consecutive observations ($M = 5$) and we used 100 000 iterations of the Gibbs sampler. In all settings, the prior for $\theta = (\tilde{\varepsilon},\tilde{s},\gamma,\beta)$ was taken to be independent Gaussian. In the estimation procedure, the prior of θ enters only in the posterior distribution of $(\theta \mid \bar{V}, D_n)$, and with the variance of the prior taken to be infinite, the prior contributes no information to the posterior. The prior for $\log(\sigma)$ was taken to be independent Gaussian with mean $[\log(0.3) + 2, \log(0.5) + 1]$ and variance (5,5).

For the MH step, the variance of the random walk proposal was set to (0.03,0.0075) for the data sets with sample size $n = 200$. For each iteration a proposal parameter is either accepted or rejected, and acceptance rates were 20% and 21% for oscillatory and excitatory data, respectively.

A. Estimation of the drift parameters

Figure 3 shows density plots of the marginal posterior of each of the four drift parameters, for the setting from Table V, with $n = 200$. The black curves represent the excitatory data, and the gray curves represent oscillatory data. The vertical lines denote the parameter values used to generate data. For the oscillatory data the posterior distribution is more narrow and the modes are closer to the true parameter values, indicating that the estimation procedure performs better for the oscillatory data. In the lower right panel the dashed gray line separates the domain of β that leads to either excitatory or oscillatory behavior.

Figure 4 shows trace plots of the posterior for θ for the oscillatory data ($n = 200$). The plot was thinned before plotting and contains only every 50th iteration of the Gibbs sampler. For all four parameters the chain quickly reaches a stable regime, and trace plots for the excitatory data show similar characteristics.

Figure 5 shows autocorrelation plots of the posterior for θ for the oscillatory data ($n = 200$). It is desirable that the autocorrelations die out quickly to obtain a variance close

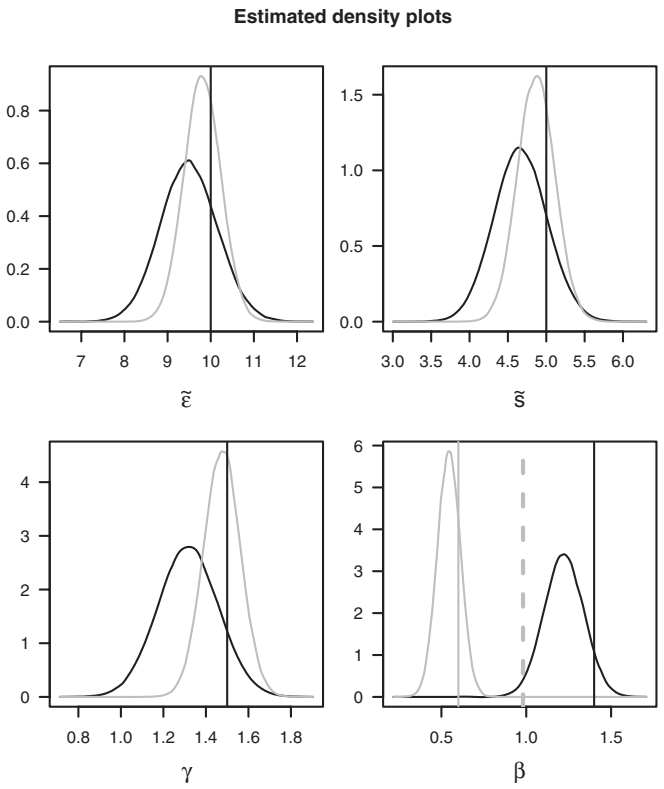

FIG. 3. Density plots of the sample posterior of the drift parameters ($n = 200$). The black and gray curves represent excitatory and oscillatory data and vertical lines denotes parameter values used to generate simulated data. The dashed line represents the parameter value where the regime changes between oscillatory and excitatory behavior. The burn-in period was 1000 iterations.

to that provided by independent sampling from the target. For γ and β the autocorrelation goes to zero very fast, but less so for $\tilde{\varepsilon}$ and $\tilde{s}$. The mixing of all parameters is much improved in the higher frequency dataset (not shown). For the excitatory data, the conclusions remain the same. Increasing the frequency of data ($n = 2000$) while keeping the sample length constant has little effect on the precision of the estimates of θ, as we expect from the theory anyway. Improved statistical precision for θ is achieved by increasing the time period of observation.

B. Estimation of the diffusion parameters

Figure 6 shows density plots of the posterior for σ for both oscillatory and excitatory data ($n = 200$). Also included is the situation where sampling frequency is increased to $n = 2000$ in the excitatory regime. The black and gray solid lines represent excitatory and oscillatory data, respectively, with $n = 200$. The black dashed line denotes excitatory data with $n = 2000$. For both data sets with $n = 200$, σ_2 is poorly estimated while for the oscillatory data the true parameter values are in the central part of the support of the posterior for σ_1. For increased sampling frequency ($n = 2000$) the posterior variance decreases and the generating parameter values are well within the support of the posterior for both σ_1 and σ_2.

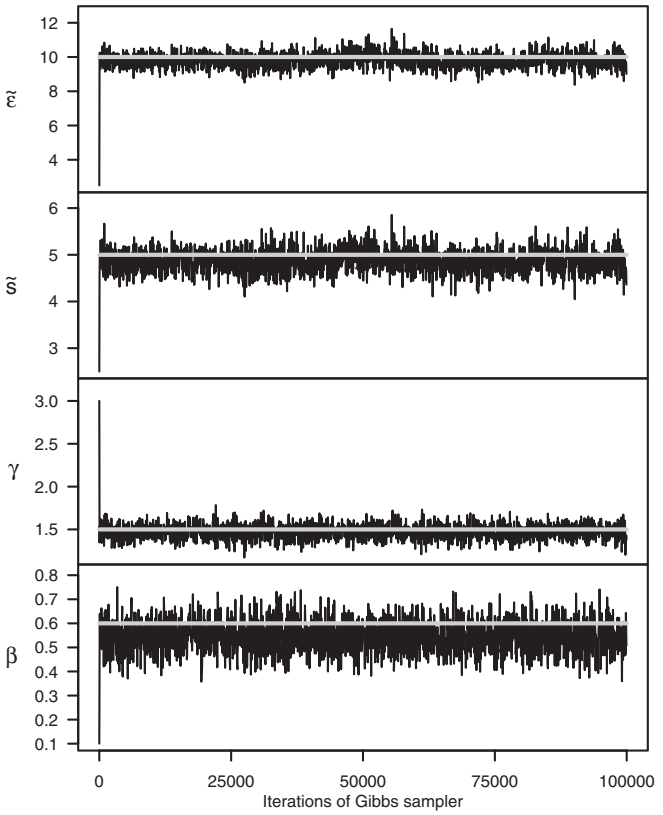

FIG. 4. Trace plot for the four drift parameters. The horizontal lines denote parameter values used to generate simulated data. Data was thinned before plotting.

C. Changing the time scale parameter ϵ

The performance of the algorithm depends strongly on the size of the time scale separation. If the separation is large, it may become difficult to extract information from both coordinates in the system. Figure 7 shows four density plots based on data sets with $\tilde{\varepsilon} = 2$, 10, 50, and 100, and all other parameters as in the excitatory setting from Table V and $n = 200$. Clearly the estimates get worse for large values of $\tilde{\varepsilon}$.

D. Practical comments

All computations were done using R version 2.14.2, and an R-package focusing on parameter estimation in the FHN model is currently in development.

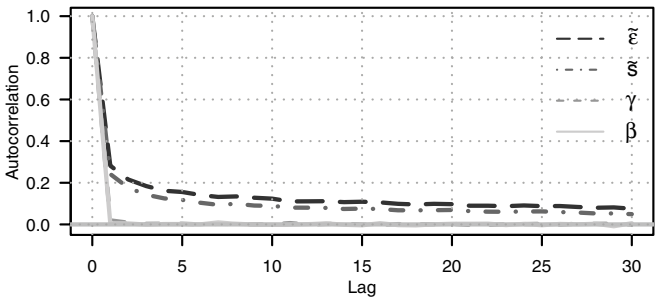

FIG. 5. Marginal autocorrelation plot for output of the Gibbs sampler. The burn-in period was 1000 iterations.

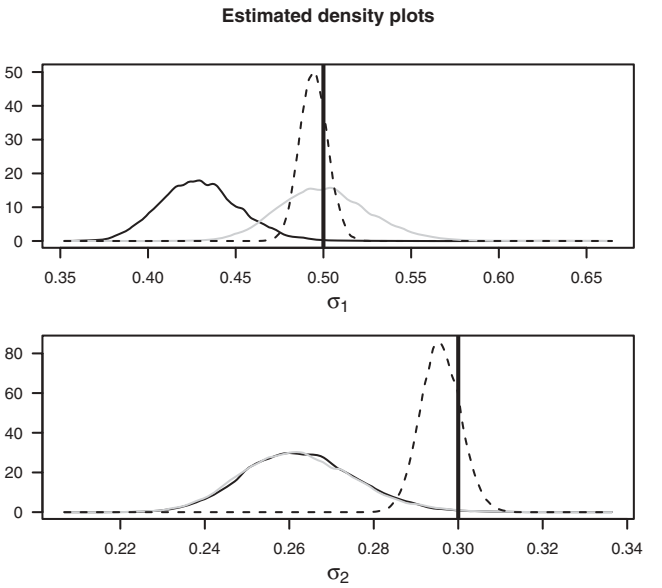

FIG. 6. Density plots of the sample posterior of the diffusion parameters. The solid and gray curves represent excitatory and oscillatory data, respectively, for $n = 200$. The black dashed curve represents excitatory data for $n = 2000$. The vertical lines denote true parameter values.

A Gaussian random walk was used as the proposal for updating the diffusion parameters. In order to tune the covariance matrix for the proposal, the Gibbs sampler ran for 10 000 iterations, with a unit proposal variance, leading to a very low acceptance rate for the parameters. Taking the diagonal of the empirical correlation matrix, $\hat{\Psi}$, after a suitable burn-in period, we obtained a rough relation between the diagonal elements of the covariance matrix. Thus, the covariance matrix was taken to be on the form $\lambda \cdot \mathrm{diag}(\hat{\Psi})$ for some $\lambda > 0$. Finally λ was tuned until the acceptance rate was relatively close to 0.23, as suggested in Ref. [23].

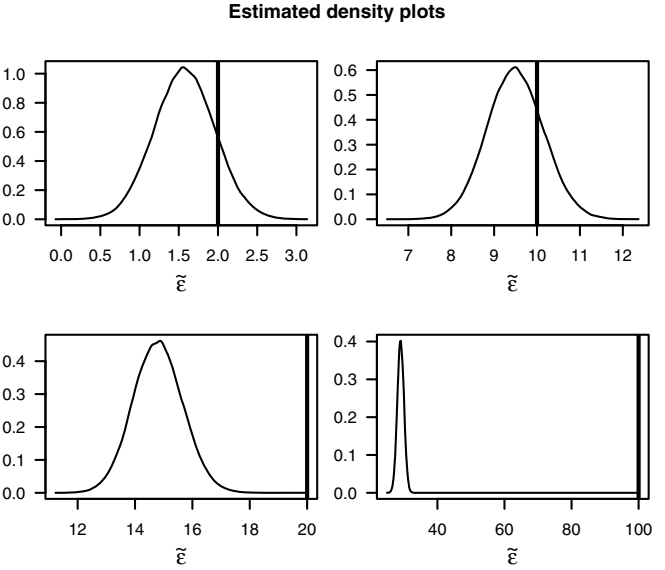

FIG. 7. Density plots for output of the Gibbs sampler for different values of $\tilde{\varepsilon}$. The vertical lines denote true values of $\tilde{\varepsilon}$.

VI. DISCUSSION

We have introduced a Bayesian approach to parameter estimation in multivariate diffusion models and the method has been applied to the FHN model for estimation of both drift and diffusion parameters. To the best of our knowledge, not many papers have previously focused on parameter estimation in the FHN model or other excitatory models.

A few comments regarding the performance of the algorithm must be made. First, it is sensitive to the size of the time scale separation, but it is expected that performance will improve further if the latent paths are updated using proposals that resemble true paths "better" than the Brownian bridge. Second, Fig. 3 suggests that the estimation procedure performs better for data in the oscillatory regime than data in the excitatory regime with respect to all four drift parameters. This may intuitively be explained by the fact that in the oscillatory setting, in the limit of no noise, the drift is observable, whereas in the excitatory regime, only the location of the steady state can be observed. Thus, less information about the drift is available in the latter case, even if the noise makes some inference possible.

In this paper we have focused on the setting where all coordinates are discretely observed without measurement noise, and the diffusion matrix Γ is of full rank. In some applications this is not the case. The methodology described here can be extended without too many difficulties to work when not all coordinates are observed, or the observations are contaminated by measurement noise, and this is work in progress. If only a subset of the coordinates includes noise (the hypoelliptic setting) the problem becomes much harder. The methodology breaks down, as it relies on the equivalence of (Gaussian) measures. To solve this problem, one could make an approximation that includes a small amount of noise, however, this may result in numerical instabilities when inverting the diffusion matrix Γ. To effectively deal with the hypoelliptic case, more sophisticated methods are required. See, for instance, Refs. [24,25].

ACKNOWLEDGMENTS

S.D. is supported by grants from the Danish Council for Independent Research | Natural Sciences. O.P. would like to acknowledge financial support by the Spanish government through a "Ramon y Cajal" fellowship and Grant No. MTM2009-09063. M.K. would like to acknowledge support from the Research Project No. P111109-0273-2610-Comunidad Castilla la Mancha, and from the Research Project No. 15400/PI/10-Seneca Fundation, Región de Murcia.

APPENDIX A: THE CONTINUOUS TIME LIKELIHOOD

The continuous time likelihood function is central to the described method of inference and it is therefore an important object to identify. It is defined as the Radon-Nikodym derivative between two equivalent measures, and for our purpose the Cameron-Martin-Girsanov theorem explicitly describes the Radon-Nikodym derivative between two equivalent Gaussian measures:

Theorem A.1. Let $(\Omega,\mathcal{F},\mathcal{F}_{t\geqslant 0},P_0)$ be a filtered probability space and define the Itô process

$$dV_t = \Sigma dB_t, \quad 0 \leqslant t \leqslant T.$$

Suppose there exist "suitable" processes $h(V_t,\theta),b(V_t,\theta)$ such that

$$\Sigma h(V_t,\theta) = -b(V_t,\theta).$$

Define for $0 \leqslant t \leqslant T$,

$$M_t = e^{-\int_0^t h(V_s,\theta)^T dB_s - \frac{1}{2}\int_0^t (h^T h)(V_s,\theta)ds},$$

and let

$$dP_b := M_T dP_0, \quad \text{on } \mathcal{F}_T.$$

If M_t is a martingale with regard to P_0 and $\mathcal{F}_t$, then P_b is a probability measure on $\mathcal{F}_T$, and

$$dV_t = b(V_t,\theta)dt + \Sigma d\tilde{B}_t, \tag{A1}$$

where $\tilde{B}_t$ is a Brownian motion with regard to P_b.

Proof. See Ref. [26], Theorem 8.6.6. ■

For a formal definition of "suitable" processes, see Definition 3.3.2, Ref. [26]. Note that the functions used here are "suitable" functions and that M_t is the continuous time likelihood relative to the measures P_b and P_0.

APPENDIX B: SAMPLING FROM $p(\theta \mid \bar{V}, D_n)$ WHEN DRIFT IS NOT LINEAR

If a reparametrization to obtain linearity in the drift parameters is not feasible, one can still obtain approximate samples from the distribution $p(\theta \mid \bar{V}, D_n)$. One approach is to use an MH step, though it requires additional computational time. We suppress σ in this section.

The interval $[t_i,t_{i+1}]$ is split into M subintervals defined by the time points $t_i^m := t_i + m\Delta/M, m = 0,\ldots,M$, such that $t_i^0 = t_i$ and $t_i^M = t_{i+1}$. Assume the imputation of $M-1$ data points between each pair of successive observations and denote the collection of imputed data in the interval (t_i,t_{i+1}) by $\bar{V}_i^M$. Thus we have $n+1$ observations, with $M-1$ imputed values in each of the n intervals. By the Markov property it follows that

$$\begin{aligned} p(\theta \mid \bar{V}, D_n) &\approx p\big(\theta \,\big|\, \cup_{i=0}^{n-1} \bar{V}_i^M, D_n\big) \\ &\propto p(\theta) p\big(\cup_{i=0}^{n-1} \bar{V}_i^M \,\big|\, D_n,\theta\big) \\ &= p(\theta) \prod_{i=0}^{n-1}\prod_{j=1}^{M} p\big(V_{t_i^j} \,\big|\, V_{t_i^{j-1}},\theta\big). \end{aligned} \tag{B1}$$

Compared to (8), each transition now occurs on the time scale of Δ/M instead of Δ and therefore, when M is large enough,

TABLE VI. MH step for sampling from $p(\theta \mid \bar{V}, D_n)$.

Initialize
(1) Initialize $\theta^{(0)}$.
Iterate
(2a) Generate a proposal $\tilde{\theta}$ from $\varphi(\tilde{\theta} \mid \theta^{(k)}; \Omega_2)$.
(2b) Let $\theta^{(k+1)} = \begin{cases} \tilde{\theta} & \text{with prob. } \min\left(1, \frac{f(\tilde{\theta}_1)}{f(\theta^{(k)})}\right), \\ \theta^{(k)} & \text{otherwise.} \end{cases}$

an Euler-Maruyama approximation is reasonable:

$$V_{t_i^j} \approx V_{t_i^{j-1}} + b\big(V_{t_i^{j-1}},\theta\big)\Delta + \Sigma\Delta B_{t_i^j},$$

where $\Delta B_{t_i^j} = B_{t_i^j} - B_{t_i^{j-1}} \sim \varphi(\cdot \mid 0; I_2\Delta)$. Therefore

$$p\big(V_{t_i^j} \mid V_{t_i^{j-1}},\theta\big) \approx \varphi\big(V_{t_i^j} \mid \mu_i; \Sigma\Sigma^T\Delta\big),$$

where $\mu_i = V_{t_i^{j-1}} + b(V_{t_i^{j-1}},\theta)\Delta$.

The density $p(\theta \mid \bar{V}, D_n)$ is approximately known up to a proportionality constant, and for simulations it is thus natural to use an MH step. Motivated by (B1) define the proportional target distribution f by

$$f(\theta) = p(\theta)\prod_{i=0}^{n-1}\prod_{j=1}^{M}\varphi\big(V_{t_i^j} \mid V_{t_i^{j-1}};\theta\big).$$

We suggest a Gaussian random walk for the proposal distribution $\varphi(\cdot \mid \theta^{(k-1)}; \Omega_2)$, to propose a new θ. Note that for this approach, parameters which are restricted to a true subset of $\mathbb{R}$ requires a reparameterization. This is indeed the case for the time scale separator ε, and since it is also smaller than 1, a logit transformation $x \mapsto \log[x/(1-x)]$ would allow for values on the entire real line. The MH step is given in Table VI.

[1] B. Lindner, J. Garcia-Ojalvo, L. Neiman, and A. Schimansky-Geier, Phys. Rep. **392**, 321 (2004).

[2] J. Keener and J. Sneyd, *Mathematical Physiology*, 2nd ed. (Springer, Berlin, 2009).

[3] N. Berglund and B. Gentz, *Noise-Induced Phenomena in Slow-Fast Dynamical Systems. A Sample Path Approach* (Springer, Berlin, 2010).

[4] R. FitzHugh, Biophys. J. **1**, 445 (1961).

[5] J. Nagumo, S. Animoto, and S. Yoshizawa, Proc. Inst. Radio Eng. **50**, 2061 (1962).

[6] A. L. Hodgkin and A. F. Huxley, J. Physiol. **117**, 500 (1952).

[7] W. Gerstner and W. Kistler, *Spiking Neuron Models* (Cambridge University Press, Cambridge, UK, 2002).

[8] S. Ditlevsen and P. Greenwood, J. Math. Biol. (2012), doi: 10.1007/s00285-012-0552-7.

[9] H. Sørensen, Int. Stat. Rev. **72**, 337 (2004).

[10] R. Hindriks, R. Jansen, F. Bijma, H. D. Mansvelder, M. C. M. de Gunst, and A. W. van der Vaart, Phys. Rev. E **84**, 021133 (2011).

[11] D. Kleinhans, Phys. Rev. E **85**, 026705 (2012).

[12] G. O. Roberts and O. Stramer, Biometrika **88**, 603 (2001).

[13] O. Papaspiliopoulos and G. O. Roberts, in *Statistical Methods for Stochastic Differential Equations*, Monographs on Statistics and Applied Probability (Chapman and Hall, London, 2012), pp. 311–337.

[14] A. Beskos, O. Papaspiliopoulos, G. Roberts, and P. Fearnhead, J. R. Stat. Soc. B **68**, 333 (2006).

[15] O. Papaspiliopoulos, Y. Pokern, G. O. Roberts, and A. M. Stuart, Biometrika **99**, 511 (2012).

[16] H. Wu and F. Noé, Phys. Rev. E **83**, 036705 (2011).

[17] Q. J. M. Huys, M. Ahrens, and L. Paninski, J. Neurophysiol. **96**, 872 (2006).

[18] Q. J. M. Huys and L. Paninski, PLoS Comput. Biol. **5**, e1000379 (2009).

[19] B. Lindner and L. Schimansky-Geier, Phys. Rev. E **60**, 7270 (1999).

[20] B. Lindner and L. Schimansky-Geier, Phys. Rev. E **61**, 6103 (2000).

[21] R. E. L. DeVille, E. Vanden-Eijnden, and C. B. Muratov, Phys. Rev. E **72**, 031105 (2005).

[22] E. Izhikevich, *Dynamical Models in Neuroscience* (MIT Press, Cambridge, MA, 2007).

[23] G. O. Roberts and J. S. Rosenthal, Stat. Sci. **16**, 351 (2001).

[24] Y. Pokern, A. Stuart, and P. Wiberg, J. R. Stat. Soc. B, **71**, 49 (2009).

[25] A. Samson and M. Thieullen, Stochastic Process. Appl. **122**, 2521 (2012).

[26] B. Øksendal, *Stochastic Differential Equations* (Springer, Berlin, 2007).