

Smoothing of, and Parameter Estimation from, Noisy Biophysical Recordings

Quentin J. M. Huys^{1,2*}, Liam Paninski^{2,3}

1 Gatsby Computational Neuroscience Unit, University College London, London, United Kingdom, **2** Center for Theoretical Neuroscience, Columbia University, New York, New York, United States of America, **3** Statistics Department, Columbia University, New York, New York, United States of America

Abstract

Biophysically detailed models of single cells are difficult to fit to real data. Recent advances in imaging techniques allow simultaneous access to various intracellular variables, and these data can be used to significantly facilitate the modelling task. These data, however, are noisy, and current approaches to building biophysically detailed models are not designed to deal with this. We extend previous techniques to take the noisy nature of the measurements into account. Sequential Monte Carlo (“particle filtering”) methods, in combination with a detailed biophysical description of a cell, are used for principled, model-based smoothing of noisy recording data. We also provide an alternative formulation of smoothing where the neural nonlinearities are estimated in a non-parametric manner. Biophysically important parameters of detailed models (such as channel densities, intercompartmental conductances, input resistances, and observation noise) are inferred automatically from noisy data via expectation-maximisation. Overall, we find that model-based smoothing is a powerful, robust technique for smoothing of noisy biophysical data and for inference of biophysical parameters in the face of recording noise.

Citation: Huys QJM, Paninski L (2009) Smoothing of, and Parameter Estimation from, Noisy Biophysical Recordings. *PLoS Comput Biol* 5(5): e1000379. doi:10.1371/journal.pcbi.1000379

Editor: Karl J. Friston, University College London, United Kingdom

Received: November 15, 2007; **Accepted:** April 1, 2009; **Published:** May 8, 2009

Copyright: © 2009 Huys, Paninski. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: QH: Gatsby Charitable Foundation, UCL Bogue Fellowship, Swartz Foundation; LP: NSF Career Award and Alfred P. Sloan Research Fellowship. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: qhuys@cantab.net

Introduction

Recent advances in imaging techniques allow measurements of time-varying biophysical quantities of interest at high spatial and temporal resolution. For example, voltage-sensitive dye imaging allows the observation of the backpropagation of individual action potentials up the dendritic tree [1–6]. Calcium imaging techniques similarly allow imaging of synaptic events in individual synapses. Such data are very well-suited to constrain biophysically detailed models of single cells. Both the dimensionality of the parameter space and the noisy and (temporally and spatially) undersampled nature of the observed data renders the use of statistical techniques desirable. Here, we here use sequential Monte Carlo methods (“particle filtering”) [7,8]—a standard machine-learning approach to hidden dynamical systems estimation—to automatically smooth the noisy data. In a first step, we will do this while inferring biophysically detailed models; in a second step, by inferring non-parametric models of the cellular nonlinearities.

Given the laborious nature of building biophysically detailed cellular models by hand [9–11], there has long been a strong emphasis on robust automatic methods [12–19]. Large-scale efforts (e.g. <http://microcircuit.epfl.ch>) have added to the need for such methods and yielded exciting advances. The Neurofitter [20] package, for example, provides tight integration with a number of standard simulation tools; implements a large number of search methods; and uses a combination of a wide variety of cost functions to measure the quality of a model’s fit to the data. These are, however, highly complex approaches that, while extremely

flexible, arguably make optimal use neither of the richness of the structure present in the statistical problem nor of the richness of new data emerging from imaging techniques. In the past, it has been shown by us and others [18,21–23] that knowledge of the true transmembrane voltage decouples a number of fundamental parameters, allowing simultaneous estimation of the spatial distribution of multiple kinetically differing conductances; of intercompartmental conductances; and of time-varying synaptic input. Importantly, this inference problem has the form of a constrained linear regression with a single, global optimum for all these parameters given the data.

None of these approaches, however, at present take the various noise sources (channel noise, unobserved variables etc.) in recording situations explicitly into account. Here, we extend the findings from [23], applying standard inference procedures to well-founded statistical descriptions of the recording situations in the hope that this more specifically tailored approach will provide computationally cheaper, more flexible, robust solutions, and that a probabilistic approach will allow noise to be addressed in a principled manner.

Specifically, we approach the issue of noisy observations and interpolation of undersampled data first in a model-based, and then in a model-free setting. We start by exploring how an accurate description of a cell can be used for optimal de-noising and to infer unobserved variables, such as Ca^{2+} concentration from voltage. We then proceed to show how an accurate model of a cell can be inferred from the noisy signals in the first place; this relies on using model-based smoothing as the first step of a standard, two-step, iterative machine learning algorithm known as

Author Summary

Cellular imaging techniques are maturing at a great pace, but are still plagued by high levels of noise. Here, we present two methods for smoothing individual, noisy traces. The first method fits a full, biophysically accurate description of the cell under study to the noisy data. This allows both smoothing of the data and inference of biophysically relevant parameters such as the density of (active) channels, input resistance, intercompartmental conductances, and noise levels; it does, however, depend on knowledge of active channel kinetics. The second method achieves smoothing of noisy traces by fitting arbitrary kinetics in a non-parametric manner. Both techniques can additionally be used to infer unobserved variables, for instance voltage from calcium concentration. This paper gives a detailed account of the methods and should allow for straightforward modification and inclusion of additional measurements.

Expectation-Maximisation [24,25]. The “Maximisation” step here turns out to be a weighted version of our previous regression-based inference method, which assumed exact knowledge of the biophysical signals.

Overview

The aim of this paper is to fit biophysically detailed models to noisy electrophysiological or imaging data. We first give an overview of the kinds of models we consider; which parameters in those models we seek to infer; how this inference is affected by the noise inherent in the measurements; and how standard machine learning techniques can be applied to this inference problem. The overview will be couched in terms of voltage measurements, but we later also consider measurements of calcium concentrations.

Compartmental models. Compartmental models are spatially discrete approximations to the cable equation [13,26,27] and allow the temporal evolution of a compartment's voltage to be written as

$$\begin{aligned} C_m dV_x &= \left[\sum \text{currents}_x(t) \right] dt + \sqrt{dt} \text{noise}_x \\ &= \left[\sum_i a_{x,i} J_{x,i}(t) \right] dt + \sqrt{dt} \sigma N_x(t) \end{aligned} \quad (1)$$

where $V_x(t)$ is the voltage in compartment x , C_m is the specific membrane capacitance, and $N_x(t)$ is current evolution noise (here assumed to be white and Gaussian). Note the important factor \sqrt{dt} which ensures that the noise variance grows linearly with time dt . The currents $a_{x,i} J_{x,i}(t)$ we will consider here are of three types:

- Axial currents along dendrites

$$I_{x,\text{axial}} = f_{xy} (V_y(t) - V_x(t)) \quad (2)$$

- Transmembrane currents from active (voltage-dependent), passive, or other (e.g. Ca^{2+} -dependent) membrane conductances

$$I_{x,\text{channel } c} = \bar{g}_{x,c} o_{x,c}(t) (E_c - V_x(t)) \quad (3)$$

- Experimentally injected currents

$$I_{x,\text{injected}} = R_m I_x(t) \quad (4)$$

where c indicates one particular current type (“channel”), E_c its reversal potential and $\bar{g}_{x,c}$ its maximal conductance in compartment x , R_m is the membrane resistivity and $I_x(t)$ is the current experimentally injected into that compartment. The variable $0 \leq o_{x,c}(t) \leq 1$ represents the time-varying open fraction of the conductance, and is typically given by complex, highly nonlinear functions of time and voltage. For example, for the Hodgkin and Huxley (HH) K^+ -channel, the kinetics are given by $o_c(t) = n^4(t)$, with

$$dn = (\alpha_n(V)(1-n) - \beta_n(V)n)dt + \sqrt{dt} \sigma_n N_n(t) \quad (5)$$

and $\alpha_n(V), \beta_n(V)$ themselves nonlinear functions of the voltage [28] and we again have an additive noise term. In practice, the gate noise is either drawn from a truncated Gaussian, or one can work with the transformed variable $\tilde{n}(t) = \tanh^{-1}(2n-1)$. Similar equations can be formulated for other variables such as the intracellular free Ca^{2+} concentration [27].

Noiseless observations. A detailed discussion of the case when the voltage is observed approximately noiselessly (such as with a patch-clamp electrode) is presented in [23] (see also [18,21,22]). We here give a short review over the material on which the present work will build. Let us henceforth assume that all the kinetics (such as $\alpha_n(V)$) of all conductances are known. Once the voltage is known, the kinetic equations can be evaluated to yield the open fraction $o_c(t)$ of each conductance c of interest. We further assume knowledge of the reversal potentials E_c , although this can be relaxed, and of the membrane specific capacitance C_m (which is henceforth neglected for notational clarity and fixed at 1 nF/cm^2 ; see [29] for a discussion of this assumption).

Knowledge of channel kinetics and voltage in each of the cell's compartments allows inference of the linear parameters $f_{xy}, \bar{g}_{c,x}, R_m$ and of the noise terms by constrained linear regression [23]. As an example, consider a single-compartment cell containing one active (Hodgkin-Huxley K^+) and one leak conductance and assume the voltage V_t has been recorded at sampling intervals dt for a time period of T_{total} . Let $T = T_{\text{total}}/dt$ be the number of data points and t index them successively $t = \{1, \dots, T\}$:

$$\begin{aligned} \underbrace{V_{t+1} - V_t}_{dV_t} &= \underbrace{\bar{g}_K}_{a_1} \underbrace{n_t^4 (E_K - V_t)}_{J_{t,1}} dt \\ &+ \underbrace{\bar{g}_L}_{a_2} \underbrace{(E_L - V_t)}_{J_{t,2}} dt \\ &+ \underbrace{R_m}_{a_3} \underbrace{I_t}_{J_{t,3}} dt + \sigma \sqrt{dt} N_{1,t} \\ &= \mathbf{J}_t \mathbf{a} + \sigma \sqrt{dt} N_{1,t} \end{aligned} \quad (6)$$

where we see that only \bar{g}_K, \bar{g}_L and R_m are now unknown; that they mediate the linear relationship between dV_t and \mathbf{J}_t ; and that these parameters can be concatenated into a vector \mathbf{a} as illustrated in equation 6. The maximum likelihood (ML) estimate of \mathbf{a} (in vectorized form) and of σ^2 are given by

$$\hat{\mathbf{a}}_{\text{ML}} \equiv \underset{\mathbf{a}}{\operatorname{argmax}} p(\{V_i\}_{i=1}^T | \mathbf{a}) = \underset{\mathbf{a}}{\operatorname{argmax}} \log p(\{V_i\}_{i=1}^T | \mathbf{a})$$

$$= \underset{\mathbf{a}}{\operatorname{argmax}} \left[- \sum_{i=1}^{T-1} \frac{(dV_i - \mathbf{J}_i \mathbf{a})^2}{2dt\sigma^2} + \text{const.} \right] \quad (7)$$

$$= \underset{\mathbf{a}}{\operatorname{argmin}} \|d\mathbf{V} - \mathbf{J}\mathbf{a}\|^2 \text{ s.t. } a_i \geq 0 \forall i$$

$$\hat{\sigma}_{\text{ML}}^2 = \frac{1}{T-1} \|d\mathbf{V} - \mathbf{J}\hat{\mathbf{a}}_{\text{ML}}\|^2 \quad (8)$$

where $\|\mathbf{x}\|^2 = \sum_i x_i^2$. Note that the last equality in equation 7 expresses the solution of the model fitting problem as a quadratic minimization with linear constraints on the parameters and is straightforwardly performed with standard packages such as `quadprog.m` in Matlab. The quadratic log-likelihood in equation 7 and therefore the linear form of the regression depends on the assumption that the evolution noise $N_{I,t}$ of the observed variable in equation 6 is Gaussian white noise. Parameters that can be simultaneously inferred in this manner from the true voltage trace are \bar{g} , f , R_m , time-varying synaptic input strengths and the evolution noise variances [23].

In the following, we will write all the dynamical equations as simultaneous equations

$$d\mathbf{h}_t = \boldsymbol{\psi}(\mathbf{h}_t)dt + \sqrt{\boldsymbol{\Sigma}dt}\mathbf{N}_t \quad (9)$$

where $\boldsymbol{\Sigma}_{ii}$ is the evolution noise variance of the i^{th} dynamic variable, $\boldsymbol{\Sigma}_{ij} = 0$ if $i \neq j$ and \mathbf{N}_t denotes a vector of independent, identically distributed (iid) random variables. These are Gaussian for unconstrained variables such as the voltage, and drawn from truncated Gaussians for constrained variables such as the gates. For the voltage we have $\boldsymbol{\psi}_V(\mathbf{h}_t) = \mathbf{J}_t \mathbf{a}/dt$ and we remind ourselves that \mathbf{J}_t is a function of \mathbf{h}_t (equation 6).

Observation noise. Most recording techniques yield estimates of the underlying variable of interest that are much more noisy than the essentially noise-free estimates patch-clamping can provide. Imaging techniques, for example, do not provide access to the true voltage which is necessary for the inference in equation 7. Figure 1 describes the hidden dynamical system setting that applies to this situation. Crucially, measurements \mathbf{y}_t are instantaneously related to the underlying voltage V_t by a probabilistic relationship (the turquoise arrows in Figure 1) which is dependent on the recording configuration. Together, the model of the observations, combined with the (Markovian) model of the dynamics given by the compartmental model define the following hidden dynamical system:

$$\text{Dynamics model } p(\mathbf{h}_t | \mathbf{h}_{t-1}, \boldsymbol{\theta}) = \mathcal{N}_{\mathbf{h}_t}(\mathbf{h}_{t-1} + \boldsymbol{\psi}(\mathbf{h}_t)dt, \boldsymbol{\Sigma}dt) \quad (10)$$

$$\text{Observation model } p(\mathbf{y}_t | \mathbf{h}_t, \boldsymbol{\theta}) = \mathcal{N}_{\mathbf{y}_t}(\mathbf{P}\mathbf{h}_t, \sigma_O^2 dt) \quad (11)$$

where $\mathcal{N}_x(m, v)$ denotes a Gaussian or truncated Gaussian distribution over x with mean m and variance v and $\mathbf{P}\mathbf{h}_t$ denotes the linear measurement process (in the following simply a linear projection such that $\mathbf{P}\mathbf{h}_t = V_t$ or $\mathbf{P}\mathbf{h}_t = [\text{Ca}^{2+}]_t$). We assume Gaussian noise both for the observations and the voltage; and truncated Gaussian noise for the gates. The Gaussian assumption on the evolution noise for the observed variable allows us to use a

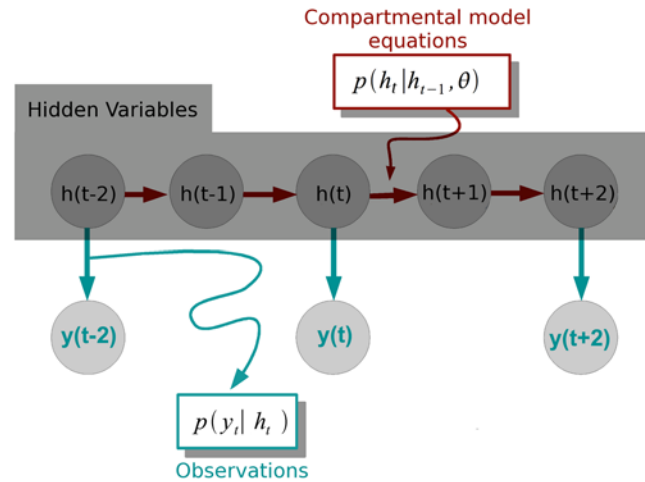


Figure 1. Hidden dynamical system. The dynamical system comprises the *hidden variables* $\mathbf{h}(t)$ and evolves as a Markov chain according to the compartmental model and kinetic equations. The dynamical system is hidden, because only noisy measurements of the true voltage are observed. To perform inference, one has to take the observation process $p(\mathbf{y}_t | \mathbf{h}_t, \boldsymbol{\theta})$ into account. Inference is now possible because the total likelihood of both observed and unobserved quantities given the parameters can be expressed in terms of these two probabilistic relations.

doi:10.1371/journal.pcbi.1000379.g001

simple regression (equation 7) in the inference of the channel densities. Note that although the noise processes are i.i.d., the fact that noise is injected into all gates means that the effective noise in the observations can show strong serial correlations.

Importantly, we do not assume that \mathbf{h} has the same dimensionality as \mathbf{y} ; in a typical cellular setting, there are several unobserved variables per compartment, only one or a few of them being measured. For Figure 2, which illustrates the particle filter for a single-compartment model with leak, Na^+ and K^+ Hodgkin-Huxley conductances, only $\mathbf{P}\mathbf{h}_t = V_t$ is measured, although the hidden variable $\mathbf{h}_t = \{V_t, h_t, m_t, n_t\}$ is 4-dimensional and includes the three gates for the Na^+ and K^+ channels in the classical Hodgkin-Huxley model. It is, however, possible to have \mathbf{y} of dimensionality equal to (or even greater than) \mathbf{h} . For example, [5] simultaneously image voltage- and $[\text{Ca}^{2+}]$ -sensitive dyes.

Expectation-Maximisation

Expectation-Maximisation (EM) is one standard machine-learning technique that allows estimation of parameters in precisely the circumstances just outlined, i.e. where inference depends on unobserved variables and certain expectations can be evaluated. The EM algorithm achieves a local maximisation of the data likelihood by iterating over two steps. For the case where voltage is recorded, it consists of:

- Expectation step (E-Step):** The parameters are fixed at their current estimate $\boldsymbol{\theta}_f = \hat{\boldsymbol{\theta}}_k$; based on this (initially inaccurate) parameter setting, the conditional distribution of the hidden variables $p(\mathbf{h}_t | \mathbf{y}_{1:T}; \boldsymbol{\theta}_f)$ (where $\mathbf{y}_{1:T} = \{\mathbf{y}_t\}_{t=1}^T$ are all the observations) is inferred. This effectively amounts to model-based smoothing of the noisy data and will be discussed in the first part of the paper.
- Maximisation step (M-Step):** Based on the model-based estimate of the hidden variables $p(\mathbf{h}_t | \mathbf{y}_{1:T}; \boldsymbol{\theta}_f)$, a new estimate of the parameters $\boldsymbol{\theta}_{k+1}$ is inferred, such that it maximises the expected joint log likelihood of the observations and the inferred

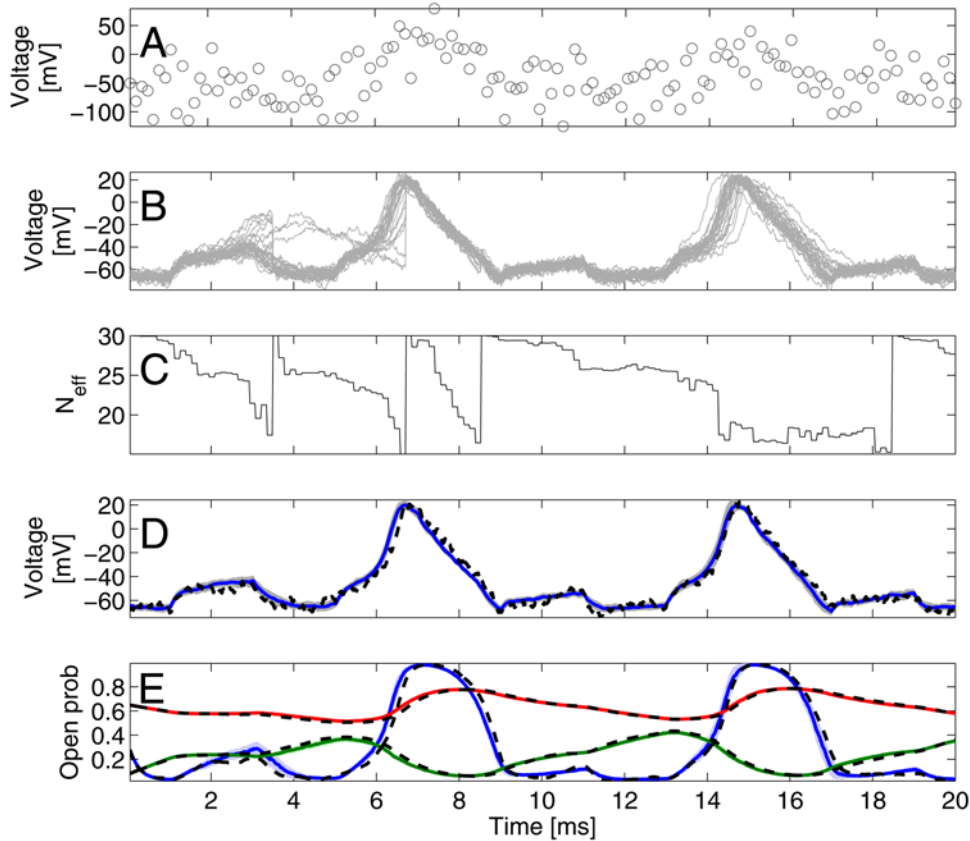


Figure 2. Model-based smoothing. **A:** Data; generated by adding Gaussian noise ($\sigma_O = 30$ mV) to the voltage trace and subsampling every seven timesteps ($\Delta = 0.02$ ms and $\Delta_s = 0.14$ ms). The voltage trace was generated by running the equation 1 for the single compartment with the correct parameters once and adding noise of variance σ_O . **B:** Voltage paths corresponding to the $N = 30$ particles which were run with the correct, known parameters. **C:** Effective particle number N_{eff} . As soon as enough particles have ‘drifted’ away from the data (N_{eff} reaches the threshold $N/2$), a resampling step eliminates the stray particles (they are reset to a particle with larger weight) all weights are reset to $1/N$ and the effective number returns to N . **D:** expected voltage trace $\hat{V}_t = \sum_i w_t^{(i)} V_t^{(i)} \pm 1$ st. dev. in shaded colours. The mean reproduces the underlying voltage trace with high accuracy. **E:** Conditional expectations for the gates of the particles (mean ± 1 st. dev.); blue: HH m -gate; green: HH h -gate; red: HH n -gate. Thus, using model-based smoothing, a highly accurate estimate of the underlying voltage and the gates can be recovered from very noisy, undersampled data. doi:10.1371/journal.pcbi.1000379.g002

distribution over the unobserved variables. This procedure is a generalisation of parameter inference in the case mentioned in equation 7, where the voltage was observed noiselessly.

The EM algorithm can be shown to increase the likelihood of the parameters at each iteration [24,25,30,31], and will typically converge to a local maximum. Although in combination with the Monte-Carlo estimation these guarantees no longer hold, in practice, we have never encountered nonglobal optima.

Methods

Model-based smoothing

We first assume that the true parameters θ are known, and in the E-step infer the conditional marginal distributions $p(\mathbf{h}_t | \mathbf{y}_{1:T}, \theta)$ for all times t . The conditional mean $\langle \mathbf{h}_t \rangle = \int d\mathbf{h}_t \mathbf{h}_t p(\mathbf{h}_t | \mathbf{y}_{1:T}, \theta)$ is a model-based, smoothed estimate of the true underlying signal \mathbf{h}_t at each point in time t which is optimal under mean squared error. The E-step is implemented using standard sequential Monte Carlo techniques [7]. Here we present the detailed equations as applied to noisy recordings of cellular dynamic variables such as the transmembrane voltage or intracellular calcium concentration.

The smoothed distribution $p(\mathbf{h}_t | \mathbf{y}_{1:T}, \theta)$ is computed via a backward recursion which relies on the filtering distribution

$p(\mathbf{h}_t | \mathbf{y}_{1:t}, \theta)$, which in turn is inferred by writing the following recursion (suppressing the dependence on θ for clarity):

$$\begin{aligned} p(\mathbf{h}_t | \mathbf{y}_{1:t}) &\propto p(\mathbf{y}_t | \mathbf{h}_t) p(\mathbf{h}_t | \mathbf{y}_{1:t-1}) \\ &= p(\mathbf{y}_t | \mathbf{h}_t) \int d\mathbf{h}_{t-1} p(\mathbf{h}_t | \mathbf{h}_{t-1}) p(\mathbf{h}_{t-1} | \mathbf{y}_{1:t-1}) \end{aligned} \quad (12)$$

This recursion relies on the fact that the hidden variables are Markovian

$$p(\mathbf{h}_t | \mathbf{h}_1, \dots, \mathbf{h}_{t-1}) = p(\mathbf{h}_t | \mathbf{h}_{t-1}) \quad (13)$$

Based on this, the smoothed distribution, which gives estimates of the hidden variables that incorporate all, not just the past, observations, can then be inferred by starting with $p(\mathbf{h}_T | \mathbf{y}_{1:T})$ and iterating backwards:

$$\begin{aligned} p(\mathbf{h}_t | \mathbf{y}_{1:T}) &= \int d\mathbf{h}_{t+1} p(\mathbf{h}_t, \mathbf{h}_{t+1} | \mathbf{y}_{1:T}) \\ &= \int d\mathbf{h}_{t+1} p(\mathbf{h}_{t+1} | \mathbf{y}_{1:T}) \frac{p(\mathbf{h}_{t+1} | \mathbf{h}_t) p(\mathbf{h}_t | \mathbf{y}_{1:t})}{\int d\mathbf{h}'_t p(\mathbf{h}_{t+1} | \mathbf{h}'_t) p(\mathbf{h}'_t | \mathbf{y}_{1:t})} \end{aligned} \quad (14)$$

where all quantities inside the integral are now known.

Sequential Monte Carlo

The filtering and smoothing equations demand integrals over the hidden variables. In the present case, these integrals are not analytically tractable, because of the complex nonlinearities in the kinetics $\psi(\cdot)$. They can, however, be approximated using Sequential Monte Carlo methods. Such methods (also known as “particle filters”) are a special version of importance sampling, in which distributions and expectations are represented by weighted samples $\{x^{(i)}\}_{i=1}^N$

$$p(x) \approx \sum_{i=1}^N w^{(i)} \delta(x - x^{(i)})$$

$$\int dx p(x)x = \langle x \rangle_{p(x)} \approx \sum_i w^{(i)} x^{(i)}$$

with $0 \leq w^{(i)} \leq 1$, $\sum_i w^{(i)} = 1$. If samples are drawn from the distribution $p(x)$ directly, the weights $w = w^{(i)} = 1/N \forall i$. In the present case, this would mean drawing samples from the distributions $p(\mathbf{h}_t | \mathbf{y}_{1:t})$ and $p(\mathbf{h}_t | \mathbf{y}_{1:T})$, which is not possible because they themselves depend on integrals at adjacent timesteps which are hard to evaluate exactly. Instead, importance sampling allows sampling from a different “proposal” distribution $x^{(i)} \sim q(x)$ and compensating by setting $w^{(i)} = p(x^{(i)})/q(x^{(i)})$. Here, we first seek samples and forward filtering weights $w_f^{(i)}$ such that

$$p(\mathbf{h}_t | \mathbf{y}_{1:t}) \approx \sum_i w_{f,t}^{(i)} \delta(\mathbf{h}_t^{(i)} - \mathbf{h}_t) \quad (15)$$

and based on these will then derive backwards, smoothing weights such that

$$p(\mathbf{h}_t | \mathbf{y}_{1:T}) \approx \sum_i w_{s,t}^{(i)} \delta(\mathbf{h}_t^{(i)} - \mathbf{h}_t). \quad (16)$$

Substituting the desideratum in equation 15 for time $t-1$ into equation 12

$$p(\mathbf{h}_t | \mathbf{y}_{1:t}) = p(\mathbf{y}_t | \mathbf{h}_t) \int d\mathbf{h}_{t-1} p(\mathbf{h}_t | \mathbf{h}_{t-1}) p(\mathbf{h}_{t-1} | \mathbf{y}_{1:t-1})$$

$$\approx p(\mathbf{y}_t | \mathbf{h}_t) \sum_j w_{f,t-1}^{(j)} p(\mathbf{h}_t | \mathbf{h}_{t-1}^{(j)}) \quad (17)$$

As a proposal distribution for our setting we use the one-step predictive probability distribution (derived from the Markov property in equation 13):

$$\mathbf{h}_t^{(i)} \sim q(\mathbf{h}_t) = p(\mathbf{h}_t | \mathbf{h}_{t-1}^{(i)}) \quad (18)$$

where $\mathbf{h}_{i,t}^{(i)}$ is termed the i^{th} “particle”. The samples are made to reflect the conditional distribution by adjusting the weights, for which the probabilities $p(\mathbf{h}_t^{(i)} | \mathbf{y}_{1:t})$ need to be computed. These are given by

$$p(\mathbf{h}_t^{(i)} | \mathbf{y}_{1:t}) \approx p(\mathbf{y}_t | \mathbf{h}_t^{(i)}) \sum_j w_{f,t-1}^{(j)} p(\mathbf{h}_t^{(i)} | \mathbf{h}_{t-1}^{(j)})$$

which involves a sum over $p(\mathbf{h}_t^{(i)} | \mathbf{h}_{t-1}^{(j)})$ that is quadratic in N . We approximate this by

$$p(\mathbf{h}_t^{(i)} | \mathbf{y}_{1:t}) \approx p(\mathbf{y}_t | \mathbf{h}_t^{(i)}) p(\mathbf{h}_t^{(i)} | \mathbf{h}_{t-1}^{(i)}) w_{f,t-1}^{(i)} \quad (19)$$

which neglects the probability that the particle i at time t could in fact have arisen from particle j at time $t-1$. The weights for each of the particles are then given by a simple update equation:

$$w_{f,t}^{*(i)} = w_{f,t-1}^{(i)} p(\mathbf{y}_t | \mathbf{h}_t^{(i)}) \quad (20)$$

$$w_{f,t}^{(i)} = \frac{w_{f,t}^{*(i)}}{\sum_j w_{f,t}^{*(j)}} \quad (21)$$

One well-known consequence of the approximation in equations 19–21 is that over time, the variance of the weights becomes large; this means that most particles have negligible weight, and only one particle is used to represent a whole distribution. Classically, this problem is prevented by resampling, and we here use stratified resampling [8]. This procedure, illustrated in Figure 2, results in eliminating particles that assign little, and duplicating particles that assign large likelihood to the data whenever the effective number of particles N_{eff} drops below some threshold, here $N_{\text{eff}} = N/2$.

It should be pointed out that it is also possible to interpolate between observations, or to do learning (see below) from subsampled traces. For example, assume we have a recording frequency of $1/\Delta_s$ but wish to infer the underlying signal at a higher frequency $1/\Delta$, with $\Delta_s \leq \Delta$. At time points without observation the likelihood term in equation 21 is uninformative (flat) and we therefore set

$$w_{f,t}^{(i)} = w_{f,t-1}^{(i)} \quad (22)$$

keeping equation 21 for the remainder of times. In this paper, we will run compartmental models (equation 1) at sampling intervals Δ , and recover signals to that same temporal precision from data subsampled at intervals $\Delta_s \geq \Delta$. See e.g. [32] for further details on incorporating intermittently-sampled observations into the alternative predictive distribution $p(\mathbf{h}_t | \mathbf{h}_{t-1}^{(i)}, \mathbf{y}_t)$.

We have so far derived the filtering weights such that particles are representative of the distribution conditioned on the past data $p(\mathbf{h}_t | \mathbf{y}_{1:t}, \boldsymbol{\theta})$. It often is more appropriate to condition on the entire set of measurements, i.e. represent the distribution $p(\mathbf{h}_t | \mathbf{y}_{1:T}, \boldsymbol{\theta})$. We will see that this is also necessary for the parameter inference in the M-step. Substituting equations 15 and 16 into equation 14, we arrive at the updates for the smoothing weights

$$w_{s,t+1}^{(j)} = w_{s,t}^{(i)} \frac{p(\mathbf{h}_{t+1}^{(j)} | \mathbf{h}_t^{(i)}) w_{f,t}^{(j)}}{\sum_k p(\mathbf{h}_{t+1}^{(i)} | \mathbf{h}_t^{(k)}) w_{f,t}^{(k)}}$$

$$w_{s,t}^{(j)} = \sum_i w_{s,t+1}^{(ij)}$$

where the weights $w_{s,t+1}^{(ij)}$ now represent the joint distribution of the hidden variables at adjacent timesteps:

$$p(\mathbf{h}_t, \mathbf{h}_{t+1} | \mathbf{y}_{1:T}) \approx \sum_i \sum_j w_{s,t+1,t}^{(ij)} \delta(\mathbf{h}_{t+1}^{(i)} - \mathbf{h}_{t+1}) \delta(\mathbf{h}_t^{(j)} - \mathbf{h}_t).$$

Parameter inference

The maximum likelihood estimate of the parameters can be inferred via a maximisation of an expectation over the hidden variables:

$$\hat{\boldsymbol{\theta}}_{ML} \equiv \operatorname{argmax}_{\boldsymbol{\theta}} p(\mathbf{y}_{1:T} | \boldsymbol{\theta}) = \operatorname{argmax}_{\boldsymbol{\theta}} \int d\mathbf{h}_{1:T} p(\mathbf{y}_{1:T}, \mathbf{h}_{1:T} | \boldsymbol{\theta}),$$

where $\mathbf{h}_{1:T} = \{\mathbf{h}_t\}_{t=1}^T$. This is achieved by iterating over the two steps of the EM algorithm. In the M-step of the k^{th} iteration, the likelihood of the entire set of measurements $\mathbf{y}_{1:T}$ with respect to the parameters $\boldsymbol{\theta}$ is maximised by maximising the expected total log likelihood [25]

$$\hat{\boldsymbol{\theta}}_{k+1} = \operatorname{argmax}_{\boldsymbol{\theta}} \langle \log p(\mathbf{y}_{1:T}, \mathbf{h}_{1:T} | \boldsymbol{\theta}) \rangle_{p(\mathbf{h}_{1:T} | \mathbf{y}_{1:T}, \boldsymbol{\theta}_k)},$$

which is achieved by setting the gradients with respect to $\boldsymbol{\theta}$ to zero (see [31,33] for alternative approaches). For the main linear parameters we seek to infer in the compartmental model ($\mathbf{a} = \{f_{xy}, \bar{g}, R_m\}$), these equations are solved by performing a constrained linear regression, akin to that in equation 7. We write the total likelihood in terms of the dynamic and the observation models (equations 10 and 11):

$$p(\mathbf{y}_{1:T}, \mathbf{h}_{1:T} | \boldsymbol{\theta}) = p(\mathbf{h}_1 | \boldsymbol{\theta}) \left(\prod_{t=1}^{T-1} p(\mathbf{h}_{t+1} | \mathbf{h}_t, \boldsymbol{\theta}) \right) \left(\prod_{t=1}^T p(\mathbf{y}_t | \mathbf{h}_t, \boldsymbol{\theta}) \right)$$

Let us assume that we have noisy measurements of the voltage. Because the parametrisation of the evolution of the voltage is linear, but that of the other hidden variables is not, we separate the two as $\mathbf{h} = [V, \tilde{\mathbf{h}}]$ where $\tilde{\mathbf{h}}$ are the gates of the conductances affecting the voltage (a similar formulation can be written for $[\text{Ca}^{2+}]$ observations). Approximating the expectations by the weighted sums of the particles defined in the previous section, we arrive at

$$\begin{aligned} \langle \log p(\mathbf{y}_{1:T}, \mathbf{h}_{1:T} | \boldsymbol{\theta}) \rangle_{p(\mathbf{h}_{1:T} | \mathbf{y}_{1:T}, \boldsymbol{\theta}_k)} &\approx -\frac{1}{2} \sum_i w_{s,1}^{(i)} \left\| \mathbf{h}_1^{(i)} - \boldsymbol{\mu}_1 \right\|_{\Sigma_1}^2 \\ &- \frac{1}{2} \sum_{t=1}^{T-1} \sum_{ij} w_{s,t+1,t}^{(ij)} \left\| V_{t+1}^{(i)} - V_t^{(j)} - \mathbf{J}_t^{(j)} \mathbf{a} \right\|_{\sigma^2}^2 \\ &- \frac{1}{2} \sum_{t=1}^{T-1} \sum_i w_{s,t}^{(i)} \left\| V_t^{(i)} - \mathbf{y}_t \right\|_{\sigma_o^2}^2 \\ &- \frac{1}{2} \log |\Sigma_1| - \frac{T-1}{2} \log \sigma^2 - \frac{T}{2} \log \sigma_o^2 + \text{const.} \end{aligned} \quad (23)$$

where $\|\mathbf{x}\|_{\mathbf{C}}^2 = \frac{1}{2} \mathbf{x}^T \mathbf{C}^{-1} \mathbf{x}$, $\boldsymbol{\mu}_1$ and Σ_1 parametrise the distribution $p(\mathbf{h}_1)$ over the initial hidden variables at time $t=1$, and $\mathbf{J}_t^{(j)}$ is the t^{th} row of the matrix $\mathbf{J}^{(j)}$ derived from particle j . Note that because we are not inferring the kinetics of the channels, the evolution term for the gates (a sum over terms of the form

$w_{s,t+1,t}^{(ij)} \left\| \tilde{\mathbf{h}}_{t+1}^{(i)} - \tilde{\mathbf{h}}_t^{(j)} - \boldsymbol{\psi}(\mathbf{h}_t^{(j)}) \right\|_{\Sigma}^2$) is a constant and can be neglected. Now setting the gradients of equation 23 with respect to the parameters to zero, we find that the linear parameters can be written, as in equation 7, as a straightforward quadratic minimisation with linear constraints

$$\hat{\mathbf{a}} = \operatorname{argmin}_{\mathbf{a}} \mathbf{a}^T \mathbf{H} \mathbf{a} - 2\mathbf{f}^T \mathbf{a} \quad \text{s.t.} \quad a_i \geq 0 \forall i$$

$$\mathbf{H} = \sum_{t=1}^{T-1} \sum_j w_{s,t}^{(j)} \left(\mathbf{J}_t^{(j)} \right)^T \mathbf{J}_t^{(j)} \quad (24)$$

$$\mathbf{f} = \sum_{t=1}^{T-1} \sum_{ij} w_{s,t+1,t}^{(ij)} \left(V_{t+1}^{(i)} - V_t^{(j)} \right)^T \mathbf{J}_t^{(j)}$$

where we see that the Hessian \mathbf{H} and the linear term \mathbf{f} of the problem are given by an expectation involving the particles. Importantly, this is still a quadratic optimisation problem with linear constraints, and which is efficiently solved by standard packages. Similarly, the initialisation parameters for the unobserved hidden variables are given by

$$\begin{aligned} \hat{\boldsymbol{\mu}}_1 &= \sum_i w_{s,1}^{(i)} \mathbf{h}_1^{(i)} \\ \hat{\Sigma}_1 &= \sum_i w_{s,1}^{(i)} \left(\mathbf{h}_1^{(i)} - \hat{\boldsymbol{\mu}}_1 \right) \left(\mathbf{h}_1^{(i)} - \hat{\boldsymbol{\mu}}_1 \right)^T \end{aligned}$$

which are just the conditional mean and variance of the particles at time $t=1$; and the evolution and observation noise terms finally by

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{(T-1)dt} \sum_{t=1}^{T-1} \sum_{ij} w_{s,t+1,t}^{(ij)} \left(V_{t+1}^{(i)} - V_t^{(j)} - \mathbf{J}_t^{(j)} \mathbf{a} \right)^2 \\ \hat{\sigma}_o^2 &= \frac{1}{T} \sum_{t=1}^T \sum_i w_{s,t}^{(i)} \left(V_t^{(i)} - \mathbf{y}_t \right)^2 \end{aligned}$$

Thus, the procedure iterates over running the particle smoother in section *Sequential Monte Carlo* and then inferring the optimal parameters from the smoothed estimates of the unobserved variables.

Results

Model-based smoothing

We first present results on model-based smoothing. Here, we assume that we have a *correct* description of the parameters of the cell under scrutiny, and use this description to infer the true underlying signal from noisy measurements. These results may be considered as one possible application of a detailed model. Figure 2A shows the data, which was generated from a known, single-compartment cell with Hodgkin-Huxley-like conductances by adding Gaussian noise. The variance of the noise was chosen to replicate typical signal-to-noise ratios from voltage-dye experiments [2]. Figure 2B shows the $N=30$ particles used here, and Figure 2C the number of particles with non-negligible weights (the “effective” number N_{eff} of particles). We see that when N_{eff} hits a threshold of $N/2$, resampling results in large jumps in some particles. At around 3 ms, we see that some particles, which produced a spike at a time when there is little evidence for it in the data, are re-set to a value that is in better accord with the data.

Figure 2D shows the close match between the true underlying signal and the inferred mean $\hat{V}_t = \sum_i w_{f,t}^{(i)} V_t^{(i)}$, while Figure 2E shows that even the unobserved channel open fractions are inferred very accurately. The match for both the voltage and the open channel fractions improves with the number of particles. Code for the implementation of this smoothing step is available online at <http://www.gatsby.ucl.ac.uk/~qhuys/code.html>.

For imaging data, the laser often has to be moved between recording locations, leading to intermittent sampling at any one location (see [34–36]). Figure 3 illustrates the performance of the model-based smoother both for varying noise levels and for temporal subsampling. We see that even for very noisy and highly subsampled data, the spikes can be recovered very well.

Figure 4 shows a different aspect of the same issue, whereby the laser moves linearly across an extended linear dendrite. Here, samples are taken every Δ_s timesteps, but samples from each individual compartment are only obtained each $N_{\text{comp}}\Delta_s$. The true voltage across the entire passive dendrite is shown in Figure 4A, and the sparse data points distributed over the dendrite are shown in panel B. The inferred mean in panel C matches the true voltage very well. For this *passive*, linear example, the equations for the hidden dynamical system are exactly those of a Kalman smoother [37]; thus the standard Kalman smoother performs the correct spatial and temporal smoothing once the parameters are known, with no need for the more general (but more computationally costly) particle smoother introduced above. More precisely, in this case the integrals in equations 12 and 14 can be evaluated analytically, and no sampling is necessary.

The supplemental video S1 shows the results of a similar linear (passive-membrane) simulation, performed on a branched simulated dendrite (instead of the linear dendritic segment illustrated in Figure 4).

We emphasize that the strong performance of the particle smoother and the Kalman smoother here should not be surprising, since the data were generated from a known model and in these cases these methods perform smoothing in a statistically optimal manner. Rather, these results should illustrate the power of using an exact, correct description of the cell and its dynamics.

EM – inferring cellular parameters

We have so far shown model-based filtering assuming that a full model of the cell under scrutiny is available. Here, we instead infer some of the main parameters from the data; specifically the linear parameters f, \bar{g}_L, R_m , the observation noise σ_O and the evolution noise σ . We continue to assume, however, that the kinetics of all channels that may be present in the cell are known exactly (see [23] for a discussion of this assumption).

Figure 5 illustrates the inference for a passive multicompartamental model, similar to that in Figure 4, but driven by a square current injection into the second compartment. Figure 5B shows statistics of the inference of the leak conductance maximal density \bar{g}_L , the intercompartmental conductance f , the input resistance R_m and the observation noise σ_O across 50 different randomly generated noisy voltage traces. All the parameters are reliably recovered from 2 seconds of data at a 1 ms sampling frequency.

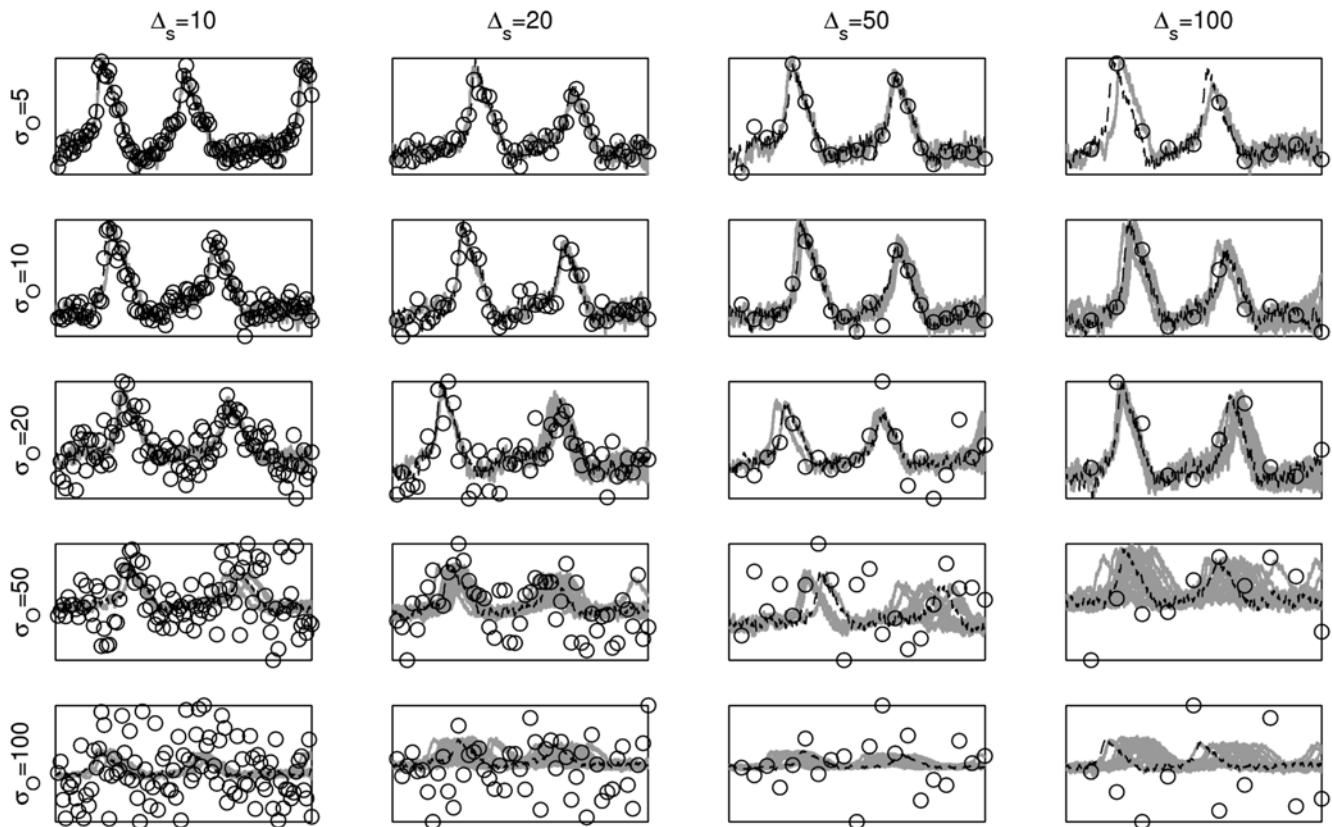


Figure 3. Performance of the model-based smoother with varying observation noise σ_O and temporal subsampling Δ_s . True underlying voltage trace in dashed black lines, the $N=20$ particles in gray and the data in black circles. Accurate inference of underlying voltage signals, and thus of spike times, is possible with accurate descriptions of the cell, over a wide range of noise levels and even at low sampling frequencies.

doi:10.1371/journal.pcbi.1000379.g003

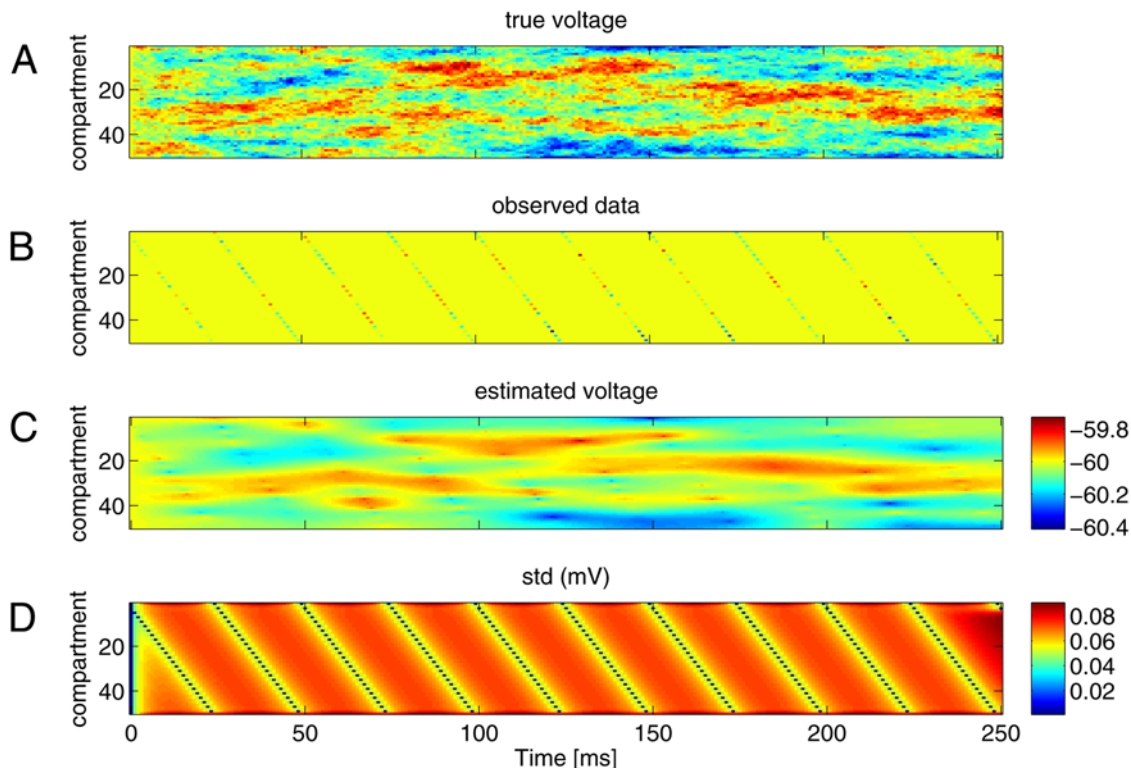


Figure 4. Inferring spatiotemporal voltage distribution from scanning, intermittent samples. **A:** True underlying voltage signal as a function of time for all 15 compartments. This was generated by injecting white noise current into a passive cell containing 50 linearly arranged compartments. **B:** Samples obtained by scanning repeatedly along the dendrite. The samples are seen as diagonal lines extending downwards, i.e. each compartment was sampled in sequence, overall 10 times and 25 ms apart. Note that the samples were noisy ($\sigma_O = 3.16$ mV). **C:** Conditional expected voltage time course for all compartments reconstructed by Kalman smoothing. The colorbar indicates the voltage for all three panels. Note that even though there is only sparse data over time and space, a smooth version of the full spatiotemporal pattern is recovered. **D:** Variance of estimated voltage. It is smallest at the observation times and rapidly reaches a steady state between observations. Due to the smoothing, which takes future data into account, the variance diminishes *ahead* of observations.
doi:10.1371/journal.pcbi.1000379.g004

We now proceed to infer channel densities and observation noise from active compartments with either four or eight channels. Figure 6 shows an example trace and inference for the four channel case (using Hodgkin-Huxley like channel kinetics). Again, we stimulated with square current pulses. Only 10 ms of data were recorded, but at a very high temporal resolution $\Delta_s = \Delta = 0.02$ ms. We see that both the underlying voltage trace and the channel and input resistance are recovered with high accuracy. Figure 7 presents batch data over 50 runs with varying levels of observation noise σ_O . The observation noise here has two effects: first, it slows down the inference (as every data point is less informative), but secondly the variance across runs increases with increasing noise (although the mean is still accurate). For illustration purposes, we started the maximal K^+ conductance at its correct value. As can be seen, however, the inference initially moves \bar{g}_K away from the optimum, to compensate for the other conductance misestimations. (This nonmonotonic behavior in \bar{g}_K is a result of the fact that the EM algorithm is searching for an optimal setting of all of the cell's conductance parameters, not just a single parameter; we will return to this issue below.)

Parametric inference here has so far employed densely sampled traces (see Figure 6A). The algorithm however applies equally to subsampled traces (see equation 22). Figure 8 shows the effect of subsampling. We see that subsampling, just as noise, slows down the inference, until the active conductances are no longer inferred accurately (the yellow trace for $\Delta_s = 0.5$ ms). In this case, the total

recording length of 10 ms meant that inference had to be done based on one single spike. For longer recordings, information about multiple spikes can of course be combined, partially alleviating this problem; however, we have found that in highly active membranes, sampling frequencies below about 1 KHz led to inaccurate estimates of sodium channel densities (since at slower sampling rates we will typically miss significant portions of the upswing of the action potential, leading the EM algorithm to underestimate the sodium channel density). Note that we kept the length of the recording in Figure 8 constant, and thus subsampling reduced the total number of measurements.

As with any importance sampling method, particle filtering is known to suffer in higher dimensions [38]. To investigate the dependence of the particle smoother's accuracy on the dimensionality of the state space, we applied the method to a compartment with a larger number of channels: fast (Na_f) and persistent Na^+ (Na_p) channels in addition to leak (L) and delayed rectifier (K_{DR}), A-type (K_A), K2-type (K_2) and M-type (K_M) K^+ channels (channel kinetics from ModelDB [39], from [9,40]). Figure 9 shows the evolution of the channel intensities during inference. Estimates of most channel densities are correct up to a factor of approximately 2. Unlike in the previous, smaller example, as either observation noise or subsampling increase, significant biases in the estimation of channel densities appear. For instance, the density of the fast sodium channel observed with noise of standard deviation 20 mV is only about half the true value.

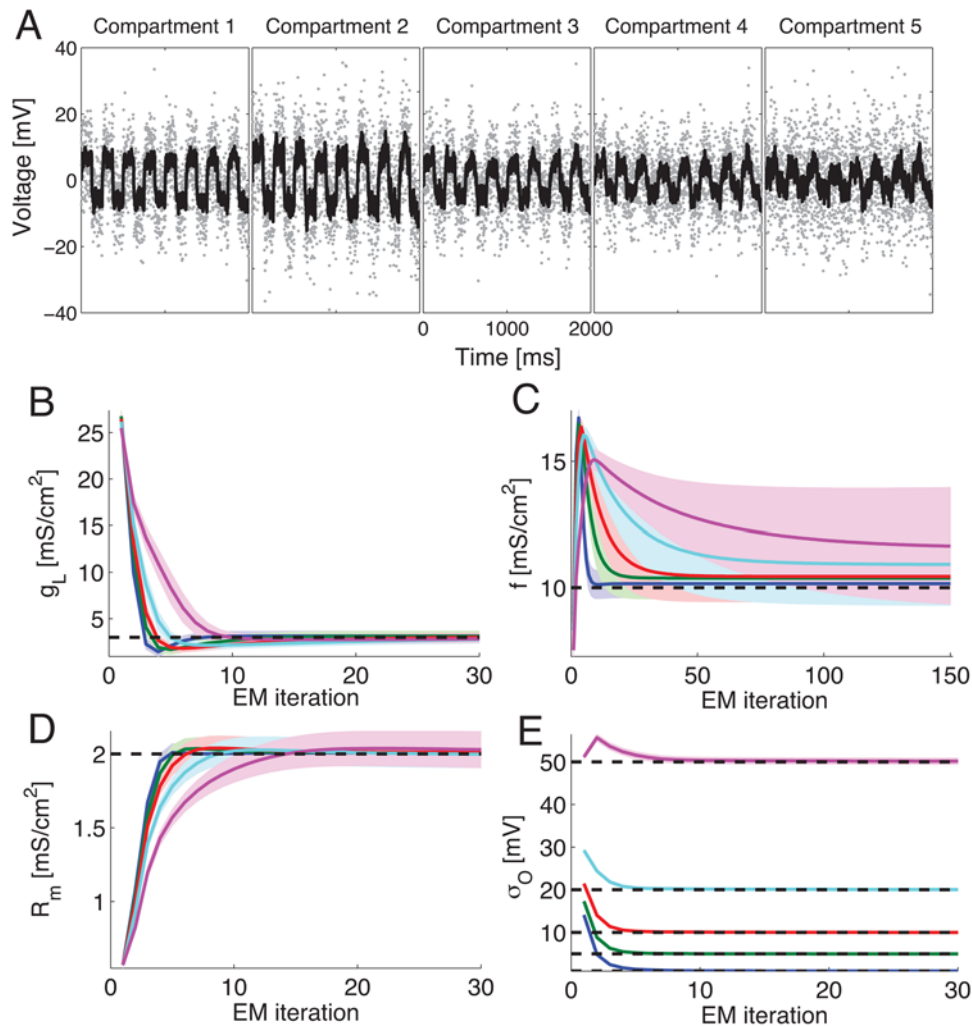


Figure 5. Inferring biophysical parameters from noisy measurements in a passive cell. **A:** True voltage (black) and noisy data (grey dots) from the 5 compartments of the cell with noise level $\sigma_O = 10$ mV. **B–E:** Parameter inference with EM. Each panel shows the average inference time course \pm one st. dev. of one of the cellular parameters. **B:** Leak conductance; **C:** intercompartmental conductance; **D:** input resistivity; **E:** Observation noise variance. The grey dotted line shows the true values. The coloured lines show the inference for varying levels of noise σ_O . Blue: $\sigma_O = 1$ mV, Green: $\sigma_O = 5$ mV, Red: $\sigma_O = 10$ mV, Cyan: $\sigma_O = 20$ mV, Magenta: $\sigma_O = 50$ mV. Throughout $\Delta_s = 1$ ms = 10Δ . Note that accurate estimation of the leak, input resistance and noise levels is even possible when the noise is five times as large as that shown in panel A. Inference of the intercompartmental conductance suffers most from the added noise because the small intercompartmental currents have to be distinguished from the apparent currents arising from noise fluctuations in the observations from neighbouring compartments. Throughout, the underlying voltage was estimated highly accurately (data not shown), which is also reflected in the accurate estimates of σ_O .
doi:10.1371/journal.pcbi.1000379.g005

It is worth noting that this bias problem is not observed in the passive linear case, where the analytic Kalman smoother suffices to perform the inference: we can infer the linear dynamical parameters of neurons with many compartments, as long as we sample information from each compartment [23]. Instead, the difficulty here is due to multicollinearity of the regression performed in the M-step of the EM algorithm and to the fact that the particle smoother leads to biased estimation of covariance parameters in high-dimensional cases [38]. We will discuss some possible remedies for these biases below.

Somewhat surprisingly, however, these observed estimation biases do not prove catastrophic if we care about predicting or smoothing the subthreshold voltage. Figure 10A compares the response to a new, random, input current of a compartment with the true parameters to that of a compartment with parameters as estimated during EM inference, while Figure 10B shows an

example prediction with $\langle |V - V_{est}| \rangle \approx 3$ mV. Note the large plateau potentials after the spikes due to the persistent sodium channel Na_p . We see that even the parameters as estimated under high noise accurately come to predict the response to a new, previously unseen, input current. The asymptote in Figure 10A is determined by the true evolution noise level (here $\sigma = 1$ mV); the more inherent noise, the less a response to a specific input is actually predictable.

Some further insight into the problem can be gained by looking at the structure of the Hessian of the total likelihood \mathbf{H} around the true parameters. We estimate \mathbf{H} by running the particle smoother with a large number of particles once at the true parameter value; more generally, one could perform a similar analysis about the inferred parameter setting to obtain a parametric bootstrap estimate of the posterior uncertainty. Figure 11 shows that, around the true value, changes in either the fast Na^+ or the

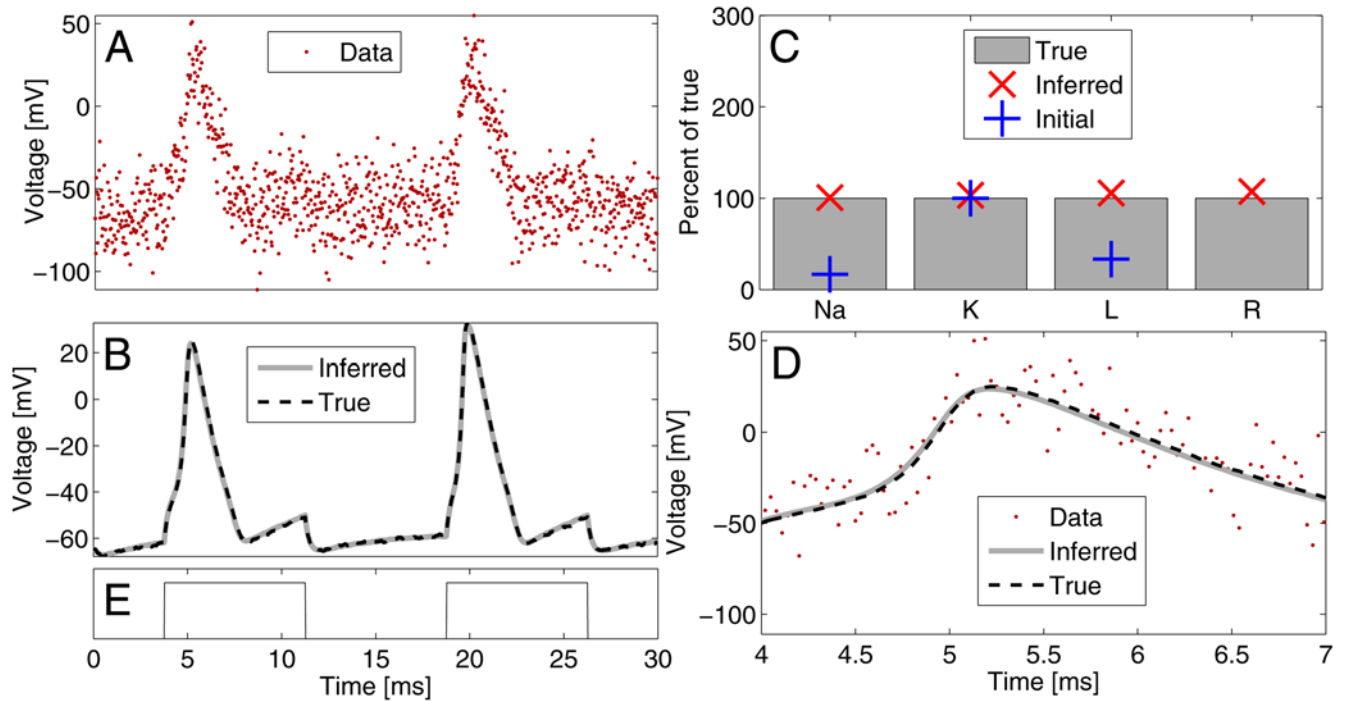


Figure 6. Example inference for single compartment with active conductances. **A:** Noisy data, $\sigma_O = 10$ mV; **B:** True underlying voltage (black dashed line) resulting from current pulse injection shown in **E**. The gray trace shows the mean inferred voltage after inferring the parameter values in **C**. **C:** Initial (blue +) and inferred parameter values (red x) in percent relative to true values (gray bars $\hat{g}_{Na} = 120$ mS/cm², $\hat{g}_K = 20$ mS/cm², $\hat{g}_{Leak} = 3$ mS/cm², $R_m = 1$ mS/cm²). At the initial values the cell was non-spiking. **D:** Magnified view showing data, inferred and true voltage traces for the first spike. Thus, despite the very high noise levels and an initially inaccurate, non-spiking model of the cell, knowledge of the channel kinetics allows accurate inference of the channel densities and very precise reconstruction of the underlying voltage trace.
doi:10.1371/journal.pcbi.1000379.g006

K_2 -type K^+ channel have the least effect; i.e., the curvature in the loglikelihood is smallest in these directions, indicating that the observed data does not adequately constrain our parameter estimates in these directions, and prior information must be used to constrain these estimates instead. This explains why these

channels showed disproportionately large amounts of inference variability, and why the prediction error did not suffer catastrophically from their relatively inaccurate inference (Figure 10A). See [23] for further discussion of this multicollinearity issue in large multichannel models.

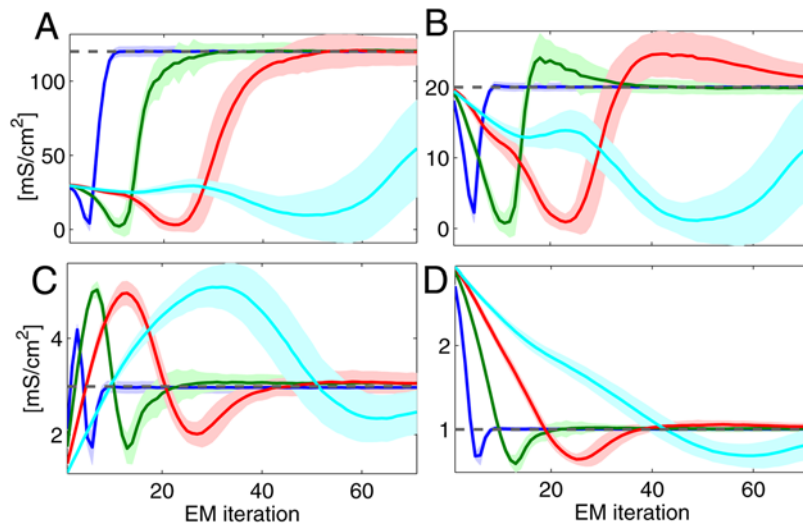


Figure 7. Time course of parameter estimation with HH channels. The four panels show, respectively, the inference for the conductance parameters **A:** \hat{g}_{Na} **B:** \hat{g}_K **C:** \hat{g}_L and **D:** R_m . The thick coloured lines indicate the mean over 50 data samples and the shaded areas 1 st. dev. The colours indicate varying noise levels σ_O . Blue: $\sigma_O = 1$ mV, Green: $\sigma_O = 5$ mV, Red: $\sigma_O = 10$ mV, Cyan: $\sigma_O = 20$ mV. The true parameters are indicated by the horizontal gray dashed lines. Throughout $\Delta t = \Delta = 0.02$ ms. The main effect of increasing observation noise is to slow down the inference. In addition, larger observation noise also adds variance to the parameter estimates. Throughout, only 10 ms of data were used.
doi:10.1371/journal.pcbi.1000379.g007

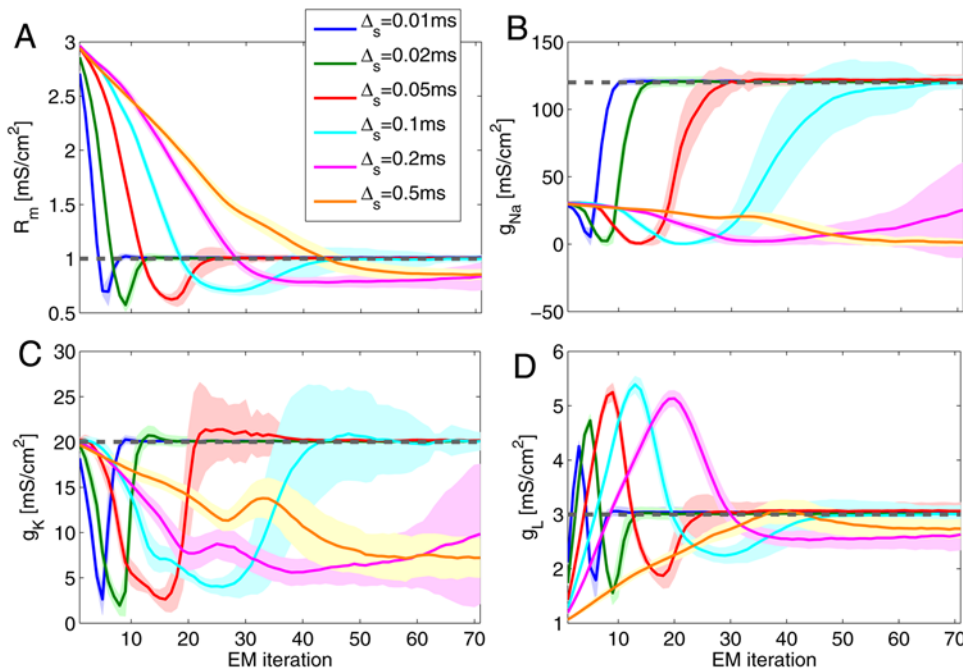


Figure 8. Subsampling slows down parametric inference. Inference of the same parameters as in previous Figure (**A**: R_m , **B**: \bar{g}_{Na} , **C**: \bar{g}_K , **D**: \bar{g}_{Leak}), but the different colours now indicate increasing subsampling. Particles evolved at timesteps of $\Delta = 0.04$ ms. The coloured traces inference with show sampling timesteps of $\Delta_s = \{0.01, 0.02, 0.05, 0.1, 0.5\}$ ms respectively. All particles were run with a $\Delta = 0.01$ ms timestep, and the total recording was always 10 ms long, meaning that progressive subsampling decreased the total number of data points. Thus, it can be seen that parameter inference is quite relatively to undersampling. At very large subsampling times, 10 ms of data supplied too few observations during a spike to justify inference of high levels of Na^+ and K^+ conductances, but the input resistance and the leak were still reliably and accurately inferred. doi:10.1371/journal.pcbi.1000379.g008

Estimation of subthreshold nonlinearity by nonparametric EM

We saw in the last section that as the dimensionality of the state vector \mathbf{h}_t grows, we may lose the ability to simultaneously estimate all of the system parameters. How can we deal with this issue? One approach is to take a step back: in many statistical settings we do not care primarily about estimating the underlying model parameters accurately, but rather we just need a model that predicts the data well. It is worth emphasizing that the methods we have introduced here can be quite useful in this setting as well. As an important example, consider the problem of estimating the subthreshold voltage given noisy observations. In many applications, we are more interested in a method which will reliably extract the subthreshold voltage than in the parameters underlying the method. For example, if a linear smoother (e.g., the Kalman smoother discussed above) works well, it might be more efficient and stable to stick with this simpler method, rather than attempting to estimate the parameters defining the cell's full complement of active membrane channels (indeed, depending on the signal-to-noise ratio and the collinearity structure of the problem, the latter goal may not be tractable, even in cases where the voltage may be reliably measured [23]).

Of course, in many cases linear smoothers are not appropriate. For example, the linear (Kalman) model typically leads to oversmoothing if the voltage dynamics are sufficiently nonlinear (data not shown), because action potentials take place on a much faster timescale than the passive membrane time constant. Thus it is worth looking for a method which can incorporate a flexible nonlinearity and whose parameters may not be directly interpretable biophysically but which nonetheless leads to good estimation

of the signal of interest. We could just throw a lot of channels into the mix, but this increases the dimensionality of the state space, hurting the performance of the particle smoother and leading to multicollinearity problems in the M-step, as illustrated in the last subsection.

A more promising approach is to fit nonlinear dynamics directly, while keeping the dimensionality of the state space as small as possible. This has been a major theme in computational neuroscience, where the reduction of complicated multichannel models into low-dimensional models, useful for phase plane analysis, has led to great insights into qualitative neural dynamics [26,41].

As a concrete example, we generated data from a strongly nonlinear (Fitzhugh-Nagumo) two-dimensional model, and then attempted to perform optimal smoothing, without prior knowledge of the underlying voltage nonlinearity. We initialized our analysis with a linear model, and then fit the nonlinearity nonparametrically via a straightforward nonparametric modification of the EM approach developed above.

In more detail, we generated data from the following model [41]:

$$dV_t = (1/\tau_V)(f(V_t) - u_t + I_t)dt + \sqrt{d_t}\sigma_V N_V(t) \quad (25)$$

$$du_t = -(1/\tau_w)u_t dt + \sqrt{d_t}\sigma_u N_u(t), \quad (26)$$

where the nonlinear function $f(V)$ is cubic in this case, and $N_u(t)$ and $N_V(t)$ denote independent white Gaussian noise processes.

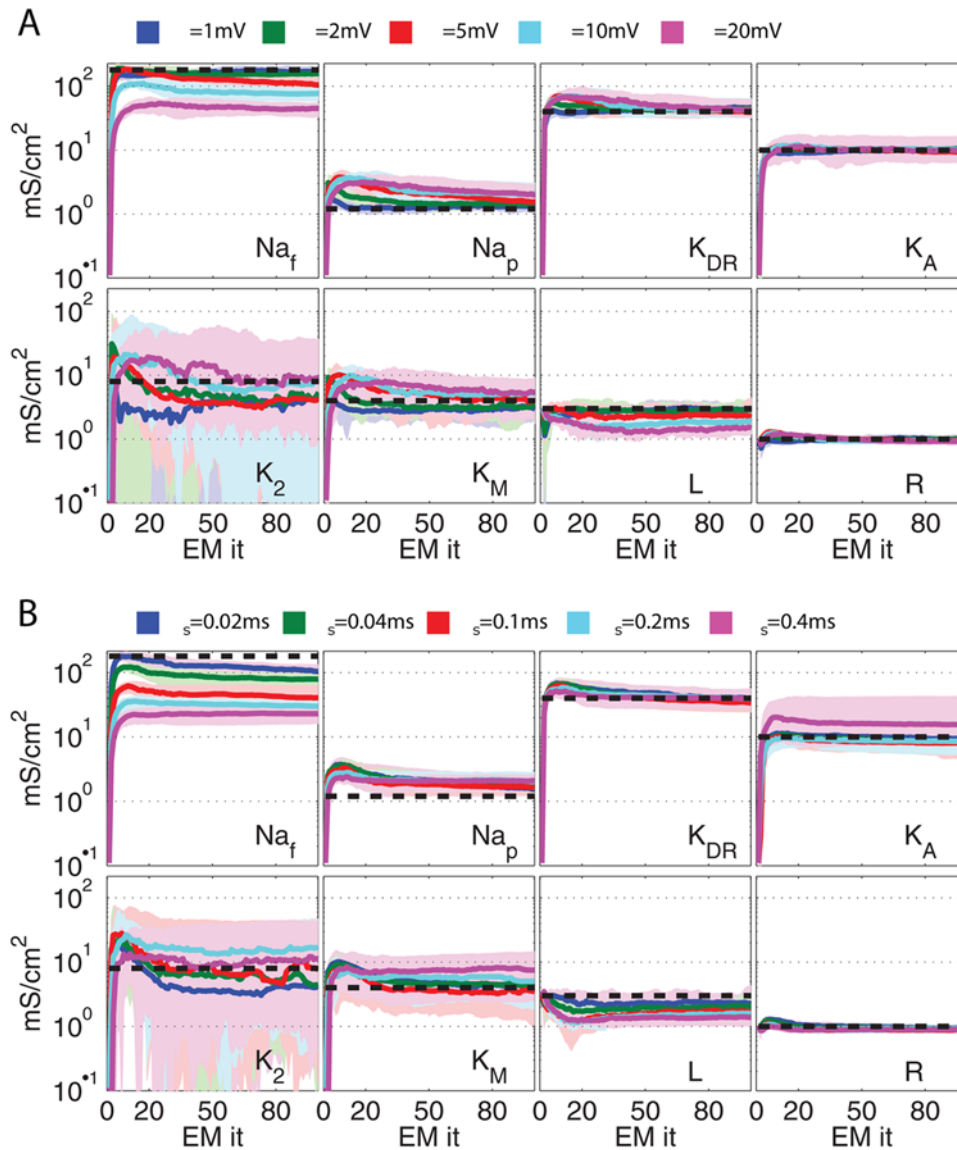


Figure 9. Time course of parameter estimation in a model with eight conductances. Evolution of estimates of channel densities for compartment with eight channels. Colours show inference with changes in the observation noise σ_O and the subsampling Δ_s . True levels are indicated by dotted gray lines. **A:** $\Delta_s = \{.02, .04, .1, .2, .4\}$ ms respectively for blue, green, red, cyan and purple lines **B:** $\sigma_O = \{1, 2, 5, 10, 20\}$ mV, $\Delta_s = \{.02, .04, .1, .2, .4\}$ ms again for blue, green, red, cyan and purple lines respectively. Thick lines show median, thin lines show 10 and 90% quantiles of distribution across 50 runs.

doi:10.1371/journal.pcbi.1000379.g009

Then, given noisy observations of the voltage V_t (Figure 12, left middle panel), we used a nonparametric version of our EM algorithm to estimate $f(V)$. The E-step of the EM algorithm is unchanged in this context: we compute $E(V_t|Y)$ and $E(u_t|Y)$, along with the other pairwise sufficient statistics, using our standard particle forward-backward smoother, given our current estimate of $f(V)$. The M-step here is performed using a penalized spline method [42]: we represent $f(V)$ as a linearly weighted combination of fixed basis functions $f_k(V)$:

$$f(V) = \sum_k \theta_k f_k(V),$$

and then determine the optimal weights θ by maximum penalized likelihood:

$$\hat{\theta} = \arg \min_{\theta} \frac{1}{2\sigma_V^2 dt} \sum_t \sum_{ij} w_{s,t+1,t}^{(ij)} \left[V_{t+1}^{(i)} - V_t^{(j)} - \frac{1}{\tau_V} \left(-u_t^{(j)} + I_t + \sum_k \theta_k f_k(V_t^{(j)}) \right) dt \right]^2 + \lambda \int \left(\frac{d}{dV} \sum_k \theta_k f_k(V) \right)^2 dV.$$

The first term here corresponds to the expected complete loglikelihood (as in equation (23)), while the second term is a penalty which serves to smooth the inferred function $f(V)$ (by penalizing non-smooth solutions, i.e., functions $f(V)$ whose

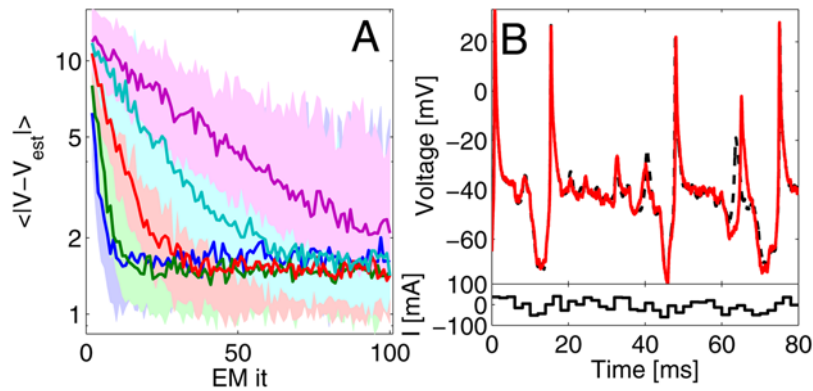


Figure 10. Predictive performance of inferred parameter settings on new input current. **A:** Parameter estimates as shown in Figure 9A were used to predict response to a new input stimulus. The plot shows the absolute error averaged over the entire trace (3000 timesteps, $\Delta_t = .02$ ms), for 40 runs. Thick lines show the median, shaded areas 10 and 90% quantiles over the same 40 runs as in Figure 9. Blue: $\sigma_O = 1$ mV, Green: $\sigma_O = 2$ mV, Red: $\sigma_O = 5$ mV, Cyan: $\sigma_O = 10$ mV, Magenta: $\sigma_O = 20$ mV. Note logarithmic y axis. **B:** Example prediction trace. The dashed black line shows the response of the cell with the true parameters, the red line that with the inferred parameters. The observation noise was $\sigma_O = 20$ mV, while the average error for this trace $\langle |V - V_{est}| \rangle = 2.96$ mV. doi:10.1371/journal.pcbi.1000379.g010

derivative has a large squared norm); the scalar λ serves to set the balance between the smoothness of $f(V)$ and the fit to the data. Despite its apparent complexity, in fact this expression is just a quadratic function of θ (just like equation (24)), and the update $\hat{\theta}$ may be obtained by solving a simple linear equation. If the basis functions $f_k(V)$ have limited overlap, then the Hessian of this objective function with respect to θ is banded (with bandwidth equal to the degree of overlap in the basis functions $f_k(V)$), and therefore this linear equation can be solved quickly using sparse banded matrix solvers [42,43]. We used 50 nonoverlapping simple step functions to represent $f(V)$ in Figures. 12–13, and each

M-step took negligible time ($\ll 1$ sec). The penalty term λ was fit crudely by eye here (we chose a λ that led to a reasonable fit to the data, without drastically oversmoothing $f(V)$); this could be done more systematically by model selection criteria such as maximum marginal likelihood or cross-validation, but the results were relatively insensitive to the precise choice of λ . Finally, it is worth noting that the EM algorithm for maximum penalized likelihood estimation is guaranteed to (locally) optimize the penalized likelihood, just as the standard EM algorithm (locally) optimizes the unpenalized likelihood.

Results are shown in Figures 12 and 13. In Figure 12, we observe a noisy version of the voltage V_t , iterate the nonparametric penalized

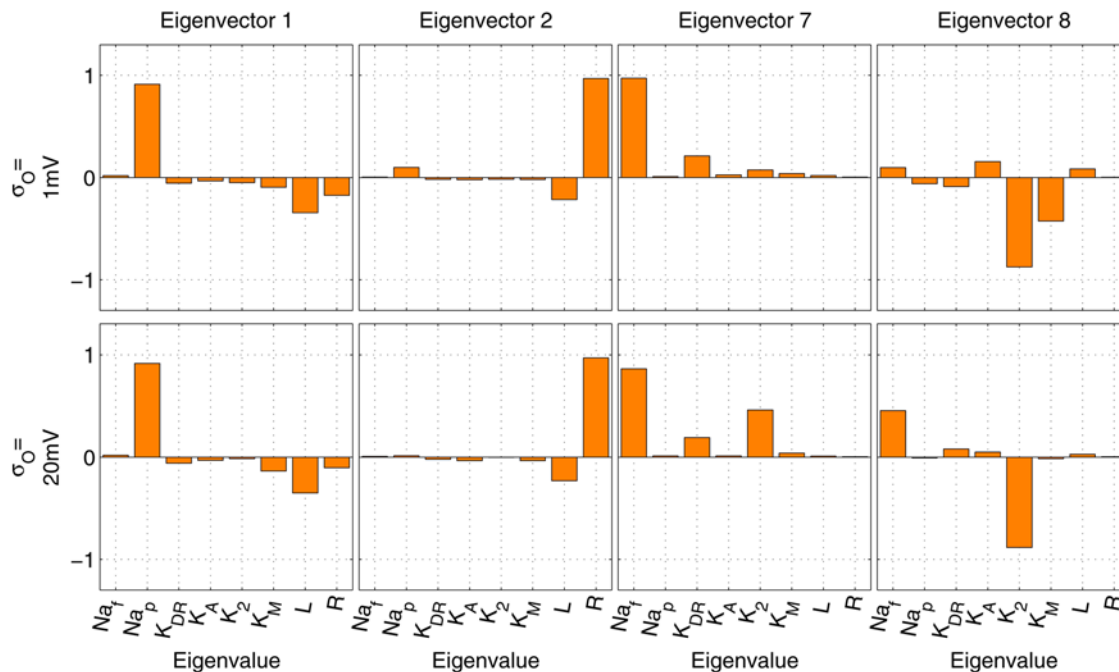


Figure 11. Eigenstructure of Hessian H with varying observation noise. Eigenvector 1 has the largest ($>10^4$), and eigenvector 8 respectively the smallest eigenvalue (~ 0.5). Independently of the noise, the smallest eigenvectors involve those channels for which inference in Figure 9 appeared least reliable: the fast Na^+ and the K_2 -type K^+ channel. doi:10.1371/journal.pcbi.1000379.g011

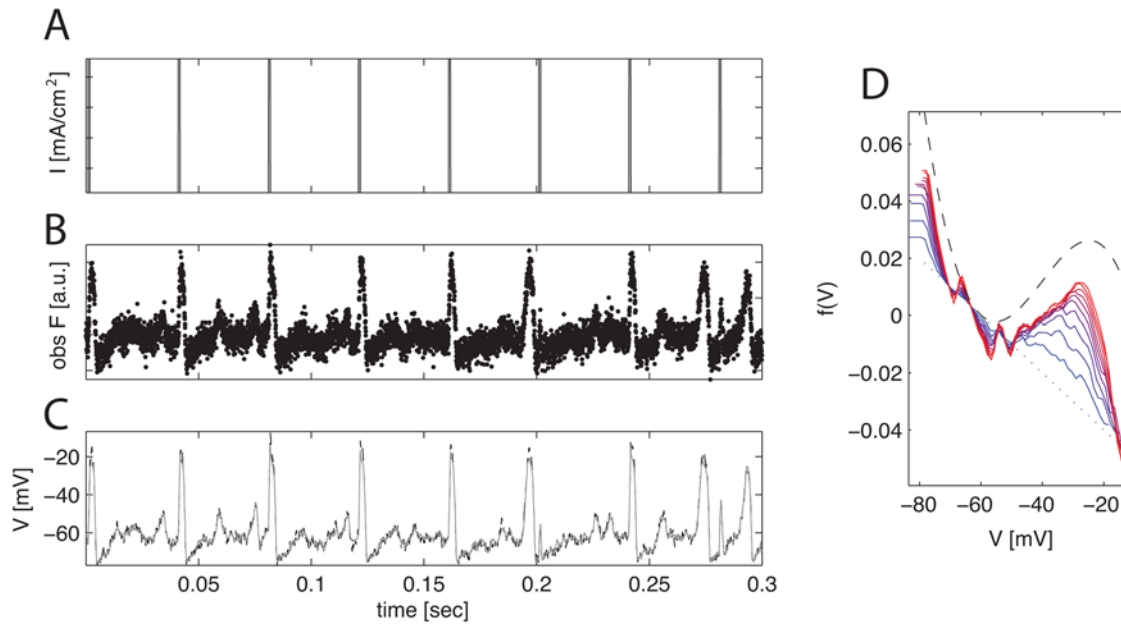


Figure 12. Estimating subthreshold nonlinearity via nonparametric EM, given noisy voltage measurements. **A, B:** input current and observed noisy voltage fluorescence data. **C:** inferred and true voltage trace. Black dashed trace: true voltage; gray solid trace: voltage inferred using nonlinearity given tenth EM iteration (red trace from right panel). Note that voltage is inferred quite accurately, despite the significant observation noise. **D:** voltage nonlinearity estimated over ten iterations of nonparametric EM. Black dashed trace: true nonlinearity; blue dotted trace: original estimate (linear initialization); solid traces: estimated nonlinearity. Color indicates iteration number: blue trace is first and red trace is last. doi:10.1371/journal.pcbi.1000379.g012

EM algorithm ten times to estimate $f(V)$, then compute the inferred voltage $E(V_t|Y)$. In Figure 13, instead of observing the noise-contaminated voltage directly, we observe the internal calcium concentration. This calcium concentration variable C_t followed its own noisy dynamics,

$$dC_t = (-C_t/\tau_C + r(V_t))dt + \sqrt{dt}\sigma_C N_C(t),$$

where $N_C(t)$ denotes white Gaussian noise, and the $r(V_t)$ term represents a fast voltage-activated inward calcium current which activates at -20 mV (i.e., this current is negligible at rest; it is effectively only activated during spiking). We then observed a noisy fluorescence signal F_t which was linearly related to the calcium concentration C_t [32]. Since the informative signal in F_t is not its absolute magnitude but rather how quickly it is currently changing

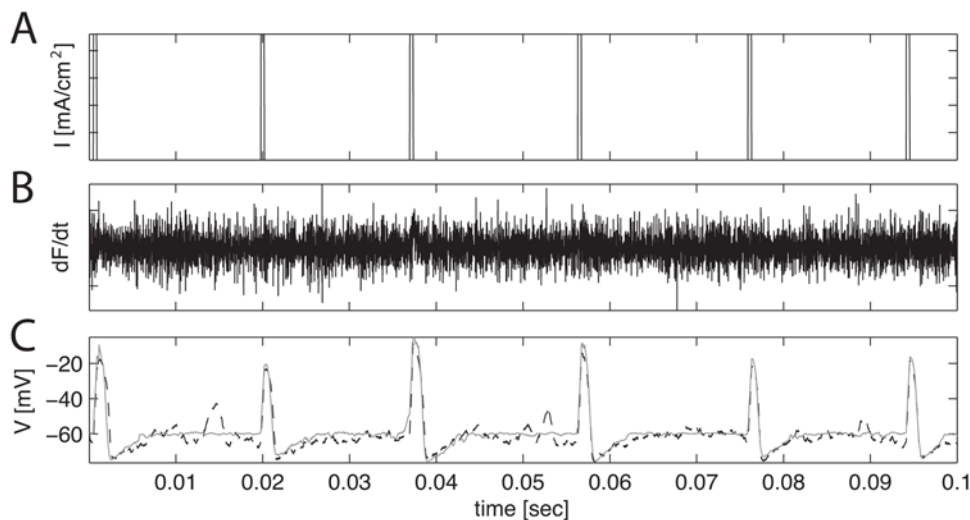


Figure 13. Estimating voltage given noisy calcium measurements, with nonlinearity estimated via nonparametric EM. **A:** Input current. **B:** Observed time derivative of calcium-sensitive fluorescence. Note the low SNR. **C:** True and inferred voltage. Black dashed trace: true voltage; gray solid trace: voltage inferred using nonlinearity following five EM iterations. Here the voltage-dependent calcium current had an activation potential at -20 mV (i.e., the calcium current is effectively zero at voltages significantly below -20 mV; at voltages >10 mV the current is ohmic). The superthreshold voltage behavior is captured fairly well, as are the post-spike hyperpolarized dynamics, but the details of the resting subthreshold behavior are lost. doi:10.1371/journal.pcbi.1000379.g013

(dC_i/dt is dominated by $r(V_i)$ during an action potential), we plot the time derivative dF_i/dt in Figure 13; note that the effective signal-to-noise in both Figures 12 and 13 is quite low.

The nonparametric EM-smoothing method effectively estimates the subthreshold voltage V_i in each case, despite the low observation SNR. In Figure 12, our estimate of $f(V)$ is biased towards a constant by our smoothing prior; this low-SNR data is not informative enough to overcome the effect of the smoothing penalty term here; indeed, since this oversmoothed estimate of $f(V)$ is sufficient to explain the data well, as seen in the left panels of Figure 12, the smoother estimate is preferred by the optimizer. With more data, or a higher SNR, the estimated $f(V)$ becomes more accurate (data not shown). It is also worth noting that if we attempt to estimate V_i from dF_i/dt using a linear smoother in Figure 13, we completely miss the hyperpolarization following each action potential; this further illustrates the advantages of the model-based approach in the context of these highly nonlinear dynamical observations.

Discussion

This paper applied standard machine learning techniques to the problem of inferring biophysically detailed models of single neurones automatically and directly from single-trial imaging data. In the first part, the paper presented techniques for the use of detailed models to filter noisy and temporally and spatially subsampled data in a principled way. The second part of the paper used this approach to infer unknown parameters by EM.

Our approach is somewhat different from standard approaches in the cellular computational neuroscience literature ([12,14,15,19], although see [18]), in that we argue that the inference problem posed is equivalent to problems in many other statistical situations. We thus postulate a full probabilistic model of the observations and then use standard machine learning tools to do inference about biophysically relevant parameters. This is an approach that is more standard in other, closely related fields in neuroscience [44,45]. Importantly, we attempt to use the description of the problem in detail to arrive at as efficient as possible a method of using the data. This implies that we directly compare recording traces (the voltage or calcium trace), rather than attempting to fit measures of the traces such as the ISI distribution, and the sufficient statistics that are used for the parameter inference involves aspects of the data these parameters influence directly. One alternative is to include a combination of such physiologically relevant objective functions and to apply more general fitting routines [46,47]. A key assumption in our approach is that accurately fitting the voltage trace will lead to accurate fits of such other measures derived from the voltage trace, such as the inter-spike interval distribution. In the present approach this means that variability is explicitly captured by parameters internal to the model. In our experience, this is important to avoid both overfitting individual traces and neglecting the inherently stochastic nature of neural responses.

A number of possible alternatives to sequential Monte Carlo methods exist, such as variations of Kalman filtering like extended or unscented Kalman filters [48,49], variational approaches (see [50]) and approximate innovation methods [45,51,52]. We here opted for a sequential Monte Carlo method because it has the advantage of allowing the approximation of arbitrary distributions and expectations. This is of particular importance in the problem at hand because a) we specifically wish to capture the nonlinearities in the problem as well as possible and b) the distributions over the unobserved states are highly non-Gaussian, due to both the nonlinearities but also due to unit bounds on the gates.

Model-based smoothing thus provides a well-founded alternative to standard smoothing techniques, and, importantly, allows smoothing of data without any averaging over either multiple cells or multiple trials [53]. This allows the inference of unobserved variables that have an effect on the observed variable. For example, just as one can infer the channels' open fractions, one can estimate the voltage from pure $[Ca^{2+}]$ recordings (data not shown). The formulation presented makes it also straightforward to combine measurements from various variables, say $[Ca^{2+}]$ and transmembrane voltage, simply by appropriately defining the observation density $p(\mathbf{y}_t|\mathbf{h}_t,\boldsymbol{\theta})$. We should emphasize, though, that the techniques themselves are not novel. Rather, this paper aims to point out to what extent these techniques are promising for cellular imaging.

The demand, when smoothing, for an accurate knowledge of the cell's parameters is addressed in the learning part of the paper where some of the important parameters are inferred accurately from small amounts of data. One instructive finding is that adding noise to the observations did not hurt our inference on average, though it did make it slower and more variable (note the wider error bars in Figure 7). In the higher-dimensional cases, we found that the dimensions in parameter space which have least effect on the models' behavior were also least well inferred. This may replicate the reports of significant flat (although not disconnected) regions in parameter space revealed in extensive parametric fits using other methods [19]. A number of parameters also remain beyond the reach of the methods discussed here, notably the kinetic channel parameters; this is the objective of the nonparametric inference in the last section of the *Results*, and also of further ongoing work.

A number of additional questions remain open. Perhaps the fundamental direction for future research involves the analysis of models in which the nonlinear hidden variable \mathbf{h}_t is high-dimensional. As we saw in section *EM – inferring cellular parameters*, our basic particle smoothing-EM methodology can break down in this high-dimensional setting. The statistical literature suggests two standard options here. First, we could replace the particle smoothing method with more general (but more computationally expensive) Markov chain Monte Carlo (MCMC) methods [54] for computing the necessary sufficient statistics for inference in our model. Designing efficient MCMC techniques suitable for high-dimensional multicompartmental neural models remains a completely open research topic. Second, to combat the multicollinearity diagnosed in Figure 11 (see also Figure 6 of [23]), we could replace the maximum-likelihood estimates considered here with maximum a posteriori (maximum penalized likelihood) estimates, by incorporating terms in our objective function (7) to penalize parameter settings which are believed to be unlikely a priori. As discussed in section *Estimation of subthreshold nonlinearity by nonparametric EM*, the EM algorithm for maximum penalized likelihood estimation follows exactly the same structure as the standard EM algorithm for maximum likelihood estimation, and therefore our methodology may easily be adapted for this case. Finally, a third option is to proceed along the direction indicated in section *Estimation of subthreshold nonlinearity by nonparametric EM*: instead of attempting to fit the parameters of our model perfectly, in many cases we can develop good voltage smoothers using a cruder, approximate model whose parameters may be estimated much more tractably. We expect that a combination of these three strategies will prove to be crucial as optimal filtering of nonlinear voltage- and calcium-sensitive dendritic imaging data becomes more prevalent as a basic tool in systems neuroscience.

Supporting Information

Video S1 Kalman smoother video. The video shows the inference of the underlying voltage in a passive cell from intermittent recordings along the dendrites. The left panel shows the true voltage; the middle panel the measurements (black means no measurement from that dendritic location at that time, cf. Figure 4); and the right panel the reconstructed voltage in the entire cell.

Found at: doi:10.1371/journal.pcbi.1000379.s001 (1.63 MB MOV)

References

- Chien CB, Pine J (1991) Voltage-sensitive dye recording of action potentials and synaptic potentials from sympathetic microcultures. *Biophys J* 60: 697–711.
- Djurisic M, Antic S, Chen W, Zecevic D (2004) Voltage imaging from dendrites of mitral cells: EPSP attenuation and spike trigger zones. *J Neurosci* 24: 6703–6714.
- Baker BJ, Kosmidis EK, Vucinic D, Falk CX, Cohen LB, et al. (2005) Imaging brain activity with voltage- and calcium-sensitive dyes. *Cell Mol Neurobiol* 25: 245–82.
- Nuriya M, Jiang J, Nemet B, Eisenthal KB, Yuste R (2006) Imaging membrane potential in dendritic spines. *Proc Natl Acad Sci USA* 103: 786–9.
- Canepari M, Djurisic M, Zecevic D (2007) Dendritic signals from rat hippocampal CA1 pyramidal neurons during coincident pre- and post-synaptic activity: a combined voltage- and calciumimaging study. *J Physiol* 580: 463–484.
- Stuart GJ, Palmer LM (2006) Imaging membrane potential in dendrites and axons of single neurons. *Pflugers Arch* 453: 403–410.
- Doucet A, Godsill S, Andrieu C (2000) On sequential Monte Carlo sampling methods for Bayesian filtering. *Stat Comput* 10: 197–208.
- Douc R, Cappe O, Moulines E (2005) Comparison of Resampling Schemes for Particle Filtering. *Image and Signal Processing and Analysis, 2005 ISPA 2005 Proceedings of the 4th International Symposium on*. pp 64–69.
- Traub RD, Wong RK, Miles R, Michelson H (1991) A model of a CA3 hippocampal pyramidal neuron incorporating voltage-clamp data on intrinsic conductances. *J Neurophysiol* 66: 635–650.
- Poirazi P, Brannon T, Mel BW (2003) Arithmetic of subthreshold synaptic summation in a model CA1 pyramidal cell. *Neuron* 37: 977–87.
- Schaefer AT, Larkum ME, Sakmann B, Roth A (2003) Coincidence detection in pyramidal neurons is tuned by their dendritic branching pattern. *J Neurophysiol* 89: 3143–54.
- Bhalla US, Bower JM (1993) Exploring parameter space in detailed single neuron models: Simulations of the mitral and granule cells of the olfactory bulb. *J Neurophysiol* 69: 1948–1965.
- Baldi P, Vanier M, Bower JM (1998) On the use of Bayesian methods for evaluating compartmental neural models. *J Comput Neurosci* 5: 285–314.
- Vanier MC, Bower JM (1999) A comparative survey of automated parameter-search methods for compartmental neural models. *J Comput Neurosci* 7: 149–171.
- Prinz AA, Billimoria CP, Marder E (2003) Hand-tuning conductance-based models: Construction and analysis of databases of model neurons. *J Neurophysiol* 90: 3998–4015.
- Jolivet R, Lewis TJ, Gerstner W (2004) Generalized integrate-and-fire models of neuronal activity approximate spike trains of a detailed model to a high degree of accuracy. *J Neurophysiol* 92: 959–976.
- Keren N, Peled N, Korngreen A (2005) Constraining compartmental models using multiple voltage recordings and genetic algorithms. *J Neurophysiol* 94: 3730–42.
- Bush K, Knight J, Anderson C (2005) Optimizing conductance parameters of cortical neural models via electrotonic partitions. *Neural Netw* 18: 488–496.
- Achard P, Schutter ED (2006) Complex parameter landscape for a complex neuron model. *PLoS Comput Biol* 2: e94. doi:10.1371/journal.pcbi.0020094.
- Geit WV, Achard P, Schutter ED (2007) Neurofitter: a parameter tuning package for a wide range of electrophysiological neuron models. *Front Neuroinformatics* 1: 1.
- Morse TM, Davison AP, Hines ML (2001) Parameter space reduction in neuron model optimization through minimization of residual voltage clamp current. *Soc Neurosci, Online Viewer Abstract, Program No 6065*.
- Wood R, Gurney KN, Wilson CJ (2004) A novel parameter optimisation technique for compartmental models applied to a model of a striatal medium spiny neuron. *Neurocomputing* 58–60: 1109–1116.
- Huys QJM, Ahrens MB, Paninski L (2006) Efficient estimation of detailed single-neuron models. *J Neurophysiol* 96: 872–890.
- Dempster A, Laird N, Rubin D (1977) Maximum Likelihood from Incomplete Data via the EM Algorithm. *J Royal Stat Soc Series B (Methodological)* 39: 1–38.
- MacKay DJ (2003) *Information theory, inference and learning algorithms*. Cambridge, UK: Cambridge University Press.
- Koch C (1999) *Biophysics of Computation. Information processing in single neurons*. OUP.
- Dayan P, Abbott LF (2001) *Theoretical Neuroscience. Computational Neuroscience*. MIT Press.
- Hodgkin AL, Huxley AF (1952) A quantitative description of membrane current and its application to conduction and excitation in nerve. *J Physiol* 117: 500–544.
- Roth A, Häusser M (2001) Compartmental models of rat cerebellar purkinje cells based on simultaneous somatic and dendritic patch-clamp recordings. *J Physiol* 535: 445–472.
- Roweis S, Ghahramani Z (1998) Learning nonlinear dynamical systems using an EM algorithm. In: *Advances in Neural Information Processing Systems (NIPS) 11*. Cambridge, MA: MIT Press.
- Salakhutdinov R, Roweis S, Ghahramani Z (2003) Optimization with EM and Expectation-Conjugate-Gradient. In: *International Conference on Machine Learning*. volume 20. pp 672–679.
- Vogelstein J, Watson B, Packer A, Jedynak B, Yuste R, et al. (2008) Spike inference from calcium imaging using sequential monte carlo methods. *Biophys J*. In Press.
- Olsson R, Petersen K, Lehn-Schioler T (2007) State-Space Models: From the EM Algorithm to a Gradient Approach. *Neural Computation* 19: 1097.
- Djurisic M, Zecevic D (2005) Imaging of spiking and subthreshold activity of mitral cells with voltage-sensitive dyes. *Ann N Y Acad Sci* 1048: 92–102.
- Iyer V, Hoogland TM, Saggau P (2005) Fast functional imaging of single neurons using randomaccess multiphoton (RAMP) microscopy. *J Neurophysiol*.
- Saggau P (2006) New methods and uses for fast optical scanning. *Curr Opin Neurobiol* 16: 543–550.
- Durbin J, Koopman S (2001) *Time Series Analysis by State Space Methods*. Oxford University Press.
- Bickel P, Li B, Bengtsson T (2008) Sharp failure rates for the bootstrap particle filter in high dimensions. *IMS Collections* 3: 318–329.
- Hines ML, Morse T, Migliore M, Carnevale NT, Shepherd GM (2004) Modeldb: A database to support computational neuroscience. *J Comput Neurosci* 17: 7–11.
- Royeck M, Horstmann MT, Remy S, Reitze M, Yaari Y, et al. (2008) Role of axonal NaV1.6 sodium channels in action potential initiation of CA1 pyramidal neurons. *J Neurophysiol* 100: 2361–80.
- Gerstner W, Kistler WM (2002) *Spiking Neuron Models: Single Neurons, Populations, Plasticity*. Cambridge, UK: Cambridge University Press.
- Green PJ, Silverman BW (1994) *Nonparametric regression and generalized linear models*. Chapman & Hall.
- Paninski L, Ahmadian Y, Ferreira D, Koyama S, Rad KR, et al. (2009) A new look at state-space models for neural data. *J Comp Neuro Under Review*.
- Friston K, Mattout J, Trujillo-Barreto N, Ashburner J, Penny W (2007) Variational free energy and the Laplace approximation. *Neuroimage* 34: 220–234.
- Sotero R, Trujillo-Barreto N, Jimenez J, Carbonell F, Rodriguez-Rojas R (2008) Identification and comparison of stochastic metabolic/hemodynamic models (sMHM) for the generation of the BOLD signal. *J Comput Neurosci*.
- Druckmann S, Banitt Y, Gidon A, Schrmann F, Markram H, et al. (2007) A novel multiple objective optimization framework for constraining conductance-based neuron models by experimental data. *Front Neurosci* 1: 7–18.
- Druckmann S, Berger TK, Hill S, Schrmann F, Markram H, et al. (2008) Evaluating automated parameter constraining procedures of neuron models by experimental and surrogate data. *Biol Cybern* 99: 371–379.
- Wan E, van der Merve R (2001) The Unscented Kalman Filter. *Kalman Filtering and Neural Networks*. pp 221–280.
- Julier S, Uhlmann J (1997) A new extension of the Kalman filter to nonlinear systems. In: *Int. Symp. Aerospace/Defense Sensing, Simul. and Controls*. volume 3.
- Penny W, Kiebel S, Friston K (2003) Variational Bayesian inference for fMRI time series. *Neuroimage* 19: 727–741.
- Jimenez J, Ozaki T (2006) An approximate innovation method for the estimation of diffusion processes from discrete data. *Journal of Time Series Analysis* 27: 77–97.
- Riera JJ, Watanabe J, Kazuki I, Naoki M, Aubert E, et al. (2004) A state-space model of the hemodynamic approach: nonlinear filtering of BOLD signals. *Neuroimage* 21: 547–567.
- Golowasch J, Goldman MS, Abbott LF, Marder E (2002) Failure of averaging in the construction of a conductance-based neuron model. *J Neurophysiol* 87: 1129–1131.
- Liu J (2002) *Monte Carlo Strategies in Scientific Computing*. Springer.
- Huys QJM, Paninski L (2006) Model-based optimal interpolation and filtering for noisy, intermittent biophysical recordings. *Fifteenth Annual Computational Neuroscience Meeting*.

Acknowledgments

We would like to thank Misha Ahrens, Arnd Roth and Joshua Vogelstein for very valuable discussions and detailed comments. A preliminary version of this work has appeared in abstract form as [55].

Author Contributions

Conceived and designed the experiments: QJH LP. Performed the experiments: QJH LP. Analyzed the data: QJH LP. Wrote the paper: QJH LP.