J Comput Neurosci (2008) 24:179–194
DOI 10.1007/s10827-007-0047-5

Parameter estimation for a leaky integrate-and-fire neuronal model from ISI data

Paul Mullowney · Satish Iyengar

Received: 24 January 2007 / Revised: 12 June 2007 / Accepted: 13 June 2007 / Published online: 28 July 2007

Abstract The Ornstein-Uhlenbeck process has been proposed as a model for the spontaneous activity of a neuron. In this model, the firing of the neuron corresponds to the first passage of the process to a constant boundary, or threshold. While the Laplace transform of the first-passage time distribution is available, the probability distribution function has not been obtained in any tractable form. We address the problem of estimating the parameters of the process when the only available data from a neuron are the interspike intervals, or the times between firings. In particular, we give an algorithm for computing maximum likelihood estimates and their corresponding confidence regions for the three identifiable (of the five model) parameters by numerically inverting the Laplace transform. A comparison of the two-parameter algorithm (where the time constant τ is known *a priori*) to the three-parameter algorithm shows that significantly more data is required in the latter case to achieve comparable parameter resolution as measured by 95% confidence intervals widths. The computational methods described here are a efficient alternative to other well known estimation techniques for leaky integrate-and-fire models. Moreover, it could serve as a template for performing parameter inference on more complex integrate-and-fire neuronal models.

Keywords Ornstein–Uhlenbeck process · Parameter inference · Inverse Laplace transform

Action Editor: Wulfram Gerstner

P. Mullowney (✉)
Tech-X Corporation, 5621 Arapahoe Avenue,
Suite A, Boulder, CO 80303, USA
e-mail: paulm@txcorp.com

P. Mullowney
Department of Mathematics and Statistics,
University of Canterbury, Private Bag 4800,
Christchurch, New Zealand

S. Iyengar
Department of Statistics, University of Pittsburgh,
2730 Cathedral of Learning, Pittsburgh, PA 15260, USA
e-mail: si@stat.pitt.edu

1 Introduction

There are a variety of mathematical techniques for modelling neuronal behavior observed at the individual, cellular level (Rabinovich et al. 2006). The predictive power of these techniques ranges from purely qualitative, which is often the result of simple phenomenological models, to quantitatively precise, which arises from comprehensive descriptions of the subcellular biophysics. The seminal work on quantitatively precise models is due to Hodgkin and Huxley (HH) (Hodgkin and Huxley 1952). The HH model and its many simplifications (Morris and Lecar 1981), provide accurate predictions of the neuronal dynamics leading to action potential generation for a broad range of input currents. While the HH model's success is due to its detailed biophysical accounting, the intricate detail also yields a very large parameter space making any robust analysis difficult and a complete understanding unfeasible. With that in mind, simpler descriptions of neuronal dynamics have been pursued. Integrate-and-fire (IF) models are one approach that attempt to describe broader features of neuronal dynamics while

maintaining sufficient simplicity to enable deeper analytic understanding.

The leaky integrate-and-fire (LIF) model (Stein 1965; Burkitt 2006a) has received significant attention recently because it provides quantitative accuracy (Jolivet et al. 2006) while simultaneously maintaining a degree of simplicity that HH models lack. LIF models represent neurons as single compartment entities in which the spatial structure of the neuron is neglected. The time evolution of the membrane potential is formulated as an electric circuit containing a resistor and capacitor in parallel. There are several accounts of the neurophysiological background leading to these models (Ricciardi and Sacerdote 1977; Tuckwell 1988). In the absence of synaptic input, the membrane potential decays to its rest state with a characteristic time constant; this is referred to as the leakage. Synaptic input is usually accounted for in two ways. current-based or conductance-based. The former typically yields models allowing for greater analytic development; the latter are more biophysically realistic because they include the effects of reversal potentials (Lánský et al. 1995). The input can be modelled as temporally homogenous Poisson events. In limit of a large number of synapses, the diffusion models like the Ornstein–Uhlenbeck (OU) emerge (Uhlenbeck and Ornstein 1954). Time-dependent synaptic current has also been considered recently (Shimokawa et al. 1999); a thorough review of this variation on the LIF is given in Burkitt (2006b).

Action potentials or spikes are generated by imposing a threshold condition. The simplest of these conditions corresponds to resetting the membrane potential to its rest state once it reaches a certain value. More complex conditions that compare the time rate of change of the membrane potential to the time rate of change of the threshold value have also been considered (Jolivet et al. 2006). Many neurons undergo refractory periods after spiking–this phenomenon can be adopted into the LIF model.

An important aspect in the analysis and validation of LIF models has focused on parameter estimation from a given data set (Keat et al. 2001). The methods are distinguishable by the following charateristics: the types of data available for the estimate, the model being considered, the corresponding computational approach, and the target parameter set under consideration. The types of data come in three categories: first-passage time or interspike interval (ISI) data, subthreshold voltage traces gathered through patch-clamp techniques (Sharp et al. 1993a,b), and input stimulus current traces. The models can have different threshold conditions or stimulus current while the computational approaches are varied. The parameters sets are grouped into intrinsic variables like the time constant, threshold value, the reset value, and the reversal potentials. Then, there are parameters associated with the input to the neuron.

One of the first attempts was made by Ricciardi and Sato (1988), who made a detailed study of a simplified version of an LIF model by assuming that the mean of the input current is constant. This assumption reduces the LIF model to an OU model. They were able to provide useful approximations to some of the models parameters. However their analysis was incomplete in that they could not estimate the leakage parameter. This was also the case in more recent studies (Inoue et al. 1995; Ditlevsen and Ditlevsen 2006).

More sophisticated techniques have surfaced recently in the literature for more general LIF models with time-dependent stochastic input (Plesser and Tanaka 1997; Burkitt and Clark 2000). These studies provide a simple algorithm for computing the first-passage time density by numerically solving a Volterra integral equation. In particular, Plesser and Tanaka (1997) showed that an optimal signal to noise ratio (SNR) can be estimated from ISI data only. This optimal measure is computed from the power spectral density which is related to the Fourier transform of the first-passage time distribution. It was shown that the optimal SNR occurs for strictly positive noise coefficients thus giving rise to the phenomenon known as "stochastic resonance". However, this approach is limited because it assumes a simple form for the input stimulus current.

Alternatively, (Pillow et al. 2004; Paninski et al. 2004) employed a technique based on both ISI data and time-dependent stimulus current traces. First, any time-dependent stimulus current was decomposed into a linear combination of basis functions. Then, they numerically solved a Chapman–Kolmogorov (Karlin and Taylor 1981) partial differential equation representing the transition probability density of the membrane potential. The solution was then used to construct a likelihood or cost function which was then optimized through an ascent based technique over a high dimensional parameter space; ≈ 15 parameters for the stimulus current decomposition and 3 for the intrinsic parameters of the LIF-threshold model. Properties of the likelihood function imply that the result of their optimization was unique and that their method was robust.

Lánský et al. (2006) devised a technique based on subthreshold voltage traces and ISI data. Like Ricciardi and Sato (1988), they assumed that the mean of the input stimulus current was constant, thus their LIF model reduces to OU. Parameter estimations were then

performed using regression and maximum-likelihood techniques. Their optimization techniques were computationally simple due to the availability of the subthreshold voltage traces, and the existence of an exact solution for the subthreshold membrane potential for the OU process. In general, they found that the regression technique produced more reliable results when comparing experimentally gathered voltage traces to simulated voltage traces using the optimized parameters.

Then, there are computational methods based on the availability of all three data types: subthreshold voltage traces, stimulus current traces, and ISI data (Jolivet et al. 2004, 2006). In these studies, the parameters of the spike response model (SRM) and a nonlinear leaky integrate-and-fire (NLIF) models were estimated. They compared the optimized SRM and NLIF models to a full conductance based (HH) model. Remarkably, they found that the simple models reproduced 70-80 percent of the spikes generated by the full HH models. These results suggest that NLIFs can accurately reproduce the dynamics of more complex neuronal models, and that a simple representation of the stimulus current is not always such a bad assumption.

In the present paper, we outline an algorithm for computing maximum likelihood (ML) estimates for parameters of the OU model. We assume that the input stimulus current has constant mean, thus the LIF model reduces to OU. In addition, we assume that ISIs are the only available data and show that OU has three identifiable parameters, which are themselves functions of the five OU model parameters. The main difficulty with the study of first-passage times using the OU model is that the probability density function (pdf) of these times is not analytically tractable. We use the properties of the Laplace transform (Darling and Siegert 1953) and known inversion algorithms to compute the first-passage time density and its parameter partial derivatives. We assume that the firing threshold is constant, and that each time the neuron fires, it is instantly reset to its resting potential; that is, we ignore the refractory period so the spike train forms a renewal process. This implies that the spike occurs fast enough to be approximated as a point event.

In Section 2, we summarize the known properties of the OU process and we extract the estimable or identifiable parameters of the process. Then, we give an expression for the first-passage time density in terms of its Laplace transform and the identifiable parameters, which themselves are combinations of the intrinsic variables of the neuron and the input to the neuron. In Section 3, the algorithm for generating ML estimates (MLEs) is described in detail. This includes an overview of the ML method for known densities Section 3.1 and techniques for calculating the densities by inverting the Laplace transform Section 3.2. This section also discusses methods for maximizing the efficiency Section 3.3 and robustness of the algorithm Section 3.4 and finishes with a comparison of the algorithm in a case where the exact pdf is known. Section 4 opens with a description of techniques for calculating confidence intervals and regions for the estimations. This is followed by numerical results for two and three-parameter algorithms. We conclude with a discussion of the main results in Section 5.

2 SDE, first-passage time

Let $\{X(t), t \geq 0, X(0) = X_0\}$ be a stationary, Gaussian, Markovian process that is continuous in probability. Any such process is referred to as an OU diffusion and it satisfies a linear stochastic differential equation (SDE) of the form (Arnold 1974):

$$\begin{aligned} \mathbf{d}X(t) &= \left(\mu - \frac{X(t)}{\tau}\right)\mathbf{d}t + \sigma \mathbf{d}W(t) \\ X(0) &= X_0 \,. \end{aligned} \tag{1}$$

Here, $\{W(t), t \geq 0\}$ denotes Brownian motion with zero mean and unit variance and $\mathbf{d}W(t)$ represents white noise. OU diffusions have continuous sample paths. That is, for any $\epsilon > 0$,

$$\lim_{h\to 0} \mathrm{P}\{|X(t+h) - X(t)| > \epsilon\} = 0 \,,$$

where P is the probability measure. The infinitesimal drift and variance, μ and σ^2, are defined through expectation operators:

$$\mu = \lim_{h\to 0} \frac{1}{h} E\big\{X(t+h) - X(t)\big\}$$

$$\sigma^2 = \lim_{h\to 0} \frac{1}{h} E\Big\{[X(t+h) - X(t)]^2\Big\} \,.$$

More general diffusions can have both spatial and temporal dependence in the drift and variance, $\mu(X(t), t)$ and $\sigma^2(X(t), t)$ (Karlin and Taylor 1981), although this study focuses on the case where μ and σ^2 are constant. In the context of a neuron, $X(t)$ represents the depolarization or membrane potential at time t. μ captures the mean excitatory ($\mu > 0$) or inhibitory ($\mu < 0$) input from the surrounding neural network while the net variance in that signal is stored within σ^2. The remaining parameter, τ, is a physiological parameter of the cell. It defines the rate at which $X(t)$ decays to its resting

potential, X_0, in the absence of external stimuli ($\mu = 0, \sigma^2 = 0$). We can further simplify by assuming $X_0 = 0$ by translation. OU processes are commonly referred to as mean-reverting due to the presence of the attracting equilibrium, $X = \mu\tau$, in the deterministic limit of the SDE ($\sigma \to 0$). The solution to (1) is given in terms of $W(t)$:

$$X(t) = \mu\tau\left(1 - e^{-t/\tau}\right) + X_0 e^{-t/\tau} + \sigma\sqrt{\frac{\tau}{2}} e^{-t/\tau} W\left(e^{2t/\tau} - 1\right) . \quad (2)$$

Note that in the limit $\sigma \to 0$, (2) converges to the solution of the deterministic case.

The first-passage time of the OU process to a horizontal barrier, $X(t) = X_f$, is a random variable T:

$$T = \inf\left\{t > 0 : X(t) = X_f\right\} = \inf\left\{t > 0 : \theta_2 = \theta_1 e^{-t/\theta_3} + \frac{e^{-t/\theta_3}}{\sqrt{2}} W(e^{2t/\theta_3} - 1)\right\}, \quad (3)$$

where θ_i ($i = 1, 2, 3$) are the *identifiable* parameters:

$$\Theta = (\theta_1, \theta_2, \theta_3)^T \quad (4)$$

$$= \left(\frac{X_0 - \mu\tau}{\sigma\sqrt{\tau}}, \frac{X_f - \mu\tau}{\sigma\sqrt{\tau}}, \tau\right)^T . \quad (5)$$

Given only a set of first-passage time data $\vec{T} = \{T_1, \cdots, T_n\}$ and a method for calculating the pdf of (3), estimates of Θ can be generated using the ML method. It is not possible though, to estimate the full set of parameters $\{X_f, X_0, \mu, \sigma, \tau\}$ given only first-passage time data. This would require auxiliary experiments to fix two of the parameters to determine the full parameter set.

Tractable expressions for the pdf of T exist only in the special case $X_f = \mu\tau$ (Darling and Siegert 1953):

$$p(t, \zeta, \theta_3) = \frac{2\zeta e^{2t/\theta_3}}{\sqrt{\pi}\theta_3\left(e^{2t/\theta_3} - 1\right)^{3/2}} \exp\left[-\frac{\zeta^2}{e^{2t/\theta_3} - 1}\right]$$

$$\zeta = \frac{X_f - X_0}{\sigma\sqrt{\tau}} \quad (6)$$

For all other parameter values, exact representations for $p(t, \Theta)$ are more difficult to generate. However, progress can be made by considering the Laplace transform of the pdf, $\hat{p}(\nu, \Theta) = \mathcal{L}(p(t, \Theta))$, which is computed from the Chapman–Kolmogorov backward equation for the transition probability density for the OU process(Arnold 1974):

$$\frac{\partial f}{\partial t} = \mu(X)\frac{\partial f}{\partial x} + \sigma^2(X)\frac{\partial^2 f}{\partial x^2} . \quad (7)$$

Since $\mu(X) = \mu - X/\tau$ and $\sigma(X) = \sigma$ are time-independent, the transition probability is stationary in time and thus satisfies the strong Markov property:

$$f(t, X_0, X_f) = \int_0^t p(s, X_0, X) \times f(t - s, X, X_f)\mathbf{d}s \quad X_0 < X < X_f .$$

Here $p(t, X_0, X)$ is the first-passage time pdf. This convolution can be solved via Laplace transforms; the result is given in terms of the identifiable parameters, Θ:

$$\hat{p}(\nu, \Theta) = \frac{\hat{f}(\nu\theta_3, \theta_1, \theta)}{\hat{f}(\nu\theta_3, \theta_2, \theta)} \quad \forall\theta > \theta_{1,2}$$

A tedious but manageable computation of $\hat{f}$ yields (Darling and Siegert 1953; Siegert 1951):

$$\hat{p}(\nu, \Theta) = \frac{H_{\nu\theta_3}(\theta_1)}{H_{\nu\theta_3}(\theta_2)} . \quad (8)$$

Here, $H_\nu(\theta)$ are parabolic cylinder or Hermite functions that satisfy the ODE,

$$H_\nu''(\theta) - 2\theta H_\nu'(\theta) - 2\nu H_\nu(\theta) = 0 , \quad (9)$$

and the derivative are with respect to θ. Solutions of (9) are uniformly convergent power series for all θ (Lebedev 1972):

$$H_\nu(\theta) = \sum_{n=0}^{\infty} \frac{\Gamma\left(\frac{n+\nu}{2}\right)(2\theta)^n}{\Gamma\left(\frac{\nu}{2}\right)n!} . \quad (10)$$

ν is the variable of the Laplace transform domain and is called the order parameter. When $\nu \in \mathbb{Z}^+$, the Hermite functions reduce to the Hermite polynomials. The real density, $p(t, \Theta)$, can then be obtained through an inverse Laplace transform, also known as the Bromwich integral (Churchill 1981):

$$p(t, \Theta) = \mathcal{L}^{-1}\left(\hat{p}(\nu, \Theta)\right) = \frac{1}{2\pi i}\int_{\sigma_0 - i\infty}^{\sigma_0 + i\infty} e^{t\nu}\hat{p}(\nu, \Theta)\mathbf{d}\nu . \quad (11)$$

The contour integration in (11) extends ν to the complex plane ($\nu = \nu_{real} + i\nu_{imag}$). The integration path ($\nu_{real} = \sigma_0, -\infty < \nu_{imag} < \infty$) is chosen such that the abscissa of convergence, σ_0, is greater than the real part of all singularities of $\hat{p}(\nu, \Theta)$. Properties of the

Hermite functions simplify this contour integration. For instance, $H_\nu(\theta)$ is an entire function of $\nu \in \mathbb{C}$ and does not vanish for $\nu_{real} \geq 0$. This implies $\sigma_0 \geq 0$. Moreover, $H_\nu(\theta) = 0$ at a countable set of points residing along the line $\nu_{real} < 0$, $\nu_{imag} = 0$. Therefore, $\hat{p}(\nu, \Theta)$ has a countable set of simple poles at the zeros of the Hermite function $H_{\nu\theta_3}(\theta_2)$. If one extends the contour to contain these poles, then residue theory can be applied to generate a formal expression for $p(t, \Theta)$ [see Eq. (28) in Ricciardi and Sato (1988)]; however, these representations are cumbersome.

Despite the existence of the formal representation for the real density, it is not used in the algorithm discussed in Section 3 and Section 4. The decision was motivated by computational efficiency and implementation considerations. The formal representation for $p(t, \Theta)$ is an infinite series whose coefficients are also power series in the parameters $\theta_{1,2,3}$. The asymptotic behavior of these series is not completely understood, therefore robust and efficient calculation for large parameter ranges could prove difficult. More importantly, the MLE will be generated via Newton's method and requires all parameter derivatives out to second order. Generating and implementing these derivatives would be difficult for the formal representation. However, these calculations are comparatively simple in the Laplace transform domain.

3 Algorithm

We begin by describing the ML method for parameter estimation. This method assumes that the pdf and its partial derivatives with respect to the parameters can be readily computed. This is followed by a summary of the algorithms for generating a function when only its Laplace transform is available. Then, the focus turns to the computational efficiency of the ML method. Next, we discuss techniques for maximizing the algorithm's effectiveness over the broadest possible parameter space. Finally, we give results of the algorithm for the specific case where the pdf is known.

3.1 ML estimation

The method of ML provides estimates of parameters of a distribution. The goal of ML is to maximize the likelihood function,

$$L(\Theta|\vec{T}) = \prod_{i=1}^{n} p(T_i, \Theta) \,, \tag{12}$$

with respect to the identifiable parameters Θ. Here, the switch in the order of the arguments, $\vec{T}$ and Θ, reflects the idea that the observed data $\vec{T}$ now acts as the parameters of the function. Maximizing (12) is equivalent to maximizing the log-likelihood function,

$$\ln L(\Theta|\vec{T}) = \sum_{i=1}^{n} \ln p(T_i, \Theta) \,, \tag{13}$$

with respect to Θ. Moreover, ML assumes that the pdf is available and can be evaluated at the data points $p(T_i, \Theta)$. With sufficient smoothness, the MLE, $\hat{\Theta} = \left(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3\right)^T$, is then the solution to the nonlinear system of equations:

$$\begin{aligned} \frac{\partial}{\partial\theta_1} \ln L(\Theta|\vec{T}) &= 0 \\ \frac{\partial}{\partial\theta_2} \ln L(\Theta|\vec{T}) &= 0 \\ \frac{\partial}{\partial\theta_3} \ln L(\Theta|\vec{T}) &= 0 \,. \end{aligned} \tag{14}$$

This system can be solved numerically with a multidimensional root-finding algorithm. If the derivatives of $p(T_i, \Theta)$ with respect to $\theta_{1,2,3}$ are available out to second order, then application of Newton's method gives an algorithm that is asymptotically efficient (Lehmann 1983):

$$\begin{aligned} \Theta^{(n+1)} &= \Theta^{(n)} - H(\Theta^{(n)}) F(\Theta^{(n)}) \\ \Theta^{(n)} &= \left(\theta_1^{(n)}, \theta_2^{(n)}, \theta_3^{(n)}\right)^T \,. \end{aligned}$$

Here, $H(\Theta^{(n)})$ is the Hessian matrix containing the second derivatives of $\ln L(\Theta|\vec{T})$, and $F(\Theta^{(n)})$ is the system (14). Then, the MLE is the limiting value of the sequence:

$$\hat{\Theta} = \lim_{n\to\infty} \Theta^{(n)} = \lim_{n\to\infty} \left(\theta_1^{(n)}, \theta_2^{(n)}, \theta_3^{(n)}\right)^T = \left(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3\right)^T \,.$$

Typically it is found that only a few iterations are needed for convergence. This is due to the fact that $\ln L(\Theta|\vec{T})$ is a log-concave function of Θ (Iyengar and Mullowney, submitted for publication). Moreover, convergence of the algorithm requires an initial guess within the basin of attraction. Estimates for θ_1 and θ_2 can be generated through the method of moments (Inoue et al. 1995), although no such technique exists for θ_3 . For the typical neuron, θ_3 is on the order of $5 - 20$ milliseconds. This range can be used to formulate an

appropriate initial guess, although a bit of guesswork is certainly required.

3.2 Inversion of the Laplace transform

In this section, we consider techniques for computing the inverse Laplace transform of a function (11). There are a number of methods available with each having their own benefits and drawbacks. In Weeks' method (Weeks 1966), the desired function is written as an expansion in Laguerre polynomials. This routine can be computationally efficient if one needs to perform multiple evaluations in the time domain. However, the implementation is not as straightforward as the other methods. Alternatively, the Post–Widder algorithm (Kano et al. 2005) has a simple form for its expansion and is easy to use. However the convergence is slow and the method is computationally prohibitive when multiple time domain evaluations are needed. The Fourier Series technique of De Hoog (De Hoog et al. 1982; D'Amore et al. 1999) is a good choice because it is highly efficient for multiple time evaluations and straightforward to program. In this algorithm, the path of the contour integration of (11) is discretized as:

$$\nu_k = \sigma_0 + ik\Delta\nu \quad 0 \le k \le N\,. \tag{15}$$

The grid spacing parameter is given by $\Delta\nu = \pi/T_{max}$ where T_{max} is a parameter to be determined. Any numerical integration technique could be employed at this juncture; the trapezoidal rule gives a simple Fourier series expansion:

$$\tilde{p}_N(t,\Theta) = \frac{e^{\sigma_0 t}}{T_{max}}\mathbf{Re}\left[\frac{\hat{p}(\sigma_0,\Theta)}{2}+\sum_{k=1}^{N}\hat{p}\left(\sigma_0+\frac{ik\pi}{T_{max}},\Theta\right)e^{\frac{ik\pi}{T_{max}}}\right]. \tag{16}$$

The error in the approximation, $|\tilde{p}_N(t,\Theta) - p(t,\Theta)|$ converges to the trapezoidal discretization error as $N \to \infty$. This holds for $t \in [0, T_{max}]$ provided $p(t,\Theta)$ has period T_{max} . Otherwise, the method suffers from the Runge phenomenon at the boundaries of the interval. In the context of the problem at hand, $\sigma_0 = 0$ due to the properties of the Hermite functions. T_{max} is chosen to be greater than the maximum stopping time $\max\{T_1, \cdots, T_n\}$ while N is chosen sufficiently large to achieve a desired accuracy in the Fourier Series approximation. Equation (16) also requires the evaluation of the Laplace transform, $\hat{p}(\nu,\Theta)$, at an evenly spaced grid on the complex ν-axis.

3.3 Efficiency

Given this technique for inverting Laplace transform, we now consider the requirements of the ML algorithm:

$$\begin{aligned}
\textbf{density}: &\quad p(t,\Theta)\\
\textbf{first derivatives}: &\quad \partial_{\theta_1}p(t,\Theta),\ \partial_{\theta_2}p(t,\Theta),\ \partial_{\theta_3}p(t,\Theta)\\
\textbf{second derivatives}: &\quad \partial^2_{\theta_1}p(t,\Theta),\ \partial^2_{\theta_2}p(t,\Theta),\ \partial^2_{\theta_3}p(t,\Theta),\\
&\quad \partial^2_{\theta_1\theta_2}p(t,\Theta),\ \partial^2_{\theta_1\theta_3}p(t,\Theta),\ \partial^2_{\theta_2\theta_3}p(t,\Theta)\,.
\end{aligned}$$

Although the formal representation of $p(t,\Theta)$ exists and the required derivatives could be analytically computed, such expressions are impractical. Alternatively, one can work in the Laplace transform domain where the density and its derivatives have simple representations that can be computed efficiently and accurately for all parameter values. Then, the Fourier series inversion method can be used to generate the real density and its parameter derivatives. For $p(t,\Theta)$, a careful application of (16) gives the real density. For the parameter derivatives more work is required. In Iyengar and Mullowney (submitted for publication), it has been shown that the order of the differentiation and contour integration can be switched for any parameter partial derivative of $p(t,\Theta)$, i.e.

$$\begin{aligned}
\frac{\partial}{\partial\theta_i}p(t,\Theta) &= \frac{\partial}{\partial\theta_i}\left(\frac{1}{2\pi i}\int_{\sigma_0-i\infty}^{\sigma_0+i\infty}e^{t\nu}\hat{p}(\nu,\Theta)\mathbf{d}\nu\right)\\
&= \frac{1}{2\pi i}\int_{\sigma_0-i\infty}^{\sigma_0+i\infty}e^{t\nu}\frac{\partial}{\partial\theta_i}\hat{p}(\nu,\Theta)\mathbf{d}\nu
\end{aligned}$$

The efficiency of the algorithm would be significantly hindered in the event that this switch were not possible. First, the parameter derivatives would need to be approximated using finite differences. Second order accuracy in all the parameter partial derivatives using the traditional $[1, -2, 1]$ stencils or some variation therein would require 27 separate inversions of $\hat{p}(\nu,\Theta)$ for parameter values near $\Theta^{(n)}$. This would be necessary at each ML iteration. Moreover, it would introduce another source of error in the computations. Although one could use Broyden's method (Broyden 1965), this would require a second initial approximation to the parameters which is not readily available. For these reasons, the ML method is optimal provided the parameter partial derivatives of $\hat{p}(\nu,\Theta)$ can be generated in a time-efficient manner.

3.4 Algorithmic parameter range

In order for the ML algorithm to be robust, accurate representations of $\hat{p}(\nu,\Theta)$ and all its parameter partial derivatives are necessary for the widest possible

parameter range: $-\infty < \theta_2 < \theta_1 < \infty, 0 < \theta_3 < \infty$. This is a nontrivial computational task whose details merits some explanation. Consider first, the computation of $\hat{p}(\nu, \Theta)$. Recall from (8) that $\hat{p}(\nu, \Theta)$ is defined as a ratio of Hermite functions that are uniformly convergent power series in their arguments, $\theta_{1,2}$, and that the order parameter always has the form $\nu\theta_3$. In addition, recall that the Fourier inversion algorithm (16) requires evaluations of $\hat{p}(\nu, \Theta)$ along the complex axis of the order parameter. When the argument of the Hermite function is small ($|\theta_{1,2}| < 2$ roughly), the power series can be used to calculate H and thus $\hat{p}(\nu, \Theta)$ for all $\nu\theta_3$ values on the complex axis with approximately 100 terms. However, for larger $|\theta_{1,2}|$, the power series suffers from catastrophic cancellation as $\nu\theta_3 \to \infty$ (on the complex axis). Catastrophic cancellation occurs when the sum of an alternating series results in cancellation across orders of magnitude greater than the accuracy of the chosen floating point representation. The onset of the phenomenon is observed for smaller $\nu\theta_3$ with increasing $|\theta_{1,2}|$ until eventually, the power series gives unreliable results.

The observations above indicate that limiting representations of the Hermite functions are needed for $|\theta_{1,2}| \to \infty$ and $|\nu\theta_3| \to \infty$. Using the known relationships between the Hermite and parabolic cylinder functions (Lebedev 1972; Abramowitz and Stegun 1972), the following asymptotic representation can be derived for the case $\xi = \sqrt{x^2 + 4\nu\theta_3 - 2} \to \infty$:

$$H_{\nu\theta_3}(x) = 2^{3/4}\sqrt{\frac{\Gamma\left(\frac{\nu\theta_3+1}{2}\right)}{\Gamma\left(\frac{\nu\theta_3}{2}\right)}} \times \exp\left[\frac{x^2}{2} \pm \vartheta + G\left(\nu\theta_3 - \frac{1}{2}, x\sqrt{2}\right)\right]$$

$$\vartheta = \frac{x\xi}{4} + \left(\nu\theta_3 - \frac{1}{2}\right)\ln\left(\frac{x+\xi}{2\sqrt{\nu\theta_3 - \frac{1}{2}}}\right)$$

$$G\left(\nu\theta_3 - \frac{1}{2}, x\sqrt{2}\right) = -\frac{\ln\xi}{2} + \frac{g_3}{\xi^3} + \frac{g_6}{\xi^6} + \frac{g_9}{\xi^9} + \frac{g_{12}}{\xi^{12}} + O\left(\frac{1}{|\xi|^{15}}\right) \quad (17)$$

The exact form of the coefficients $g_3, g_6, \cdots$ is given by (19.10.13) in Abramowitz and Stegun (1972); note that each is a function of x and $\nu\theta_3$.

Now, consider a situation where the power series representation of H gives reliable results for small $\nu\theta_3$ before succumbing to catastrophic cancellation. (17) is particularly useful in this scenario because it can accurately compute H well before the onset of the numerical instability. Thereafter, the approximation improves for larger $\nu\theta_3$ thus giving a precise representation of H over the entire integration range. Next, consider the situation where the power series never gives reliable results ($|x| \to \infty$). Equation (17) gives reasonable approximations of H when $|\nu\theta_3| << |x|$. In this regime though, another expansion is far more accurate (Lebedev 1972):

$$H_{\nu\theta_3}(x) = \frac{2}{\Gamma\left(\frac{\nu\theta_3}{2}\right)}\left\{(-2x)^{-\nu\theta_3}\left[\sum_{k=0}^{n}\frac{(-1)^k\Gamma(\nu\theta_3+2k)}{k!(2x)^{2k}} + O\left(|x|^{-2n-2}\right)\right] + h(x)\sqrt{\pi}e^{x^2}x^{\nu\theta_3-1}\left[\sum_{k=0}^{n}\frac{\Gamma(1-\nu\theta_3+2k)}{\Gamma(1-\nu\theta_3)k!(2x)^{2k}} + O\left(|x|^{-2n-2}\right)\right]\right\}, \quad (18)$$

where $h(x)$ is the Heaviside function. For larger $\nu\theta_3$, (18) loses validity while the (17) approximation improves. Thus, one can generate an accurate computation of H over the entire integration range for any parameter values by piecing together the different representations.

Figure 1 shows the relative error between the three representations of $H_{\nu\theta_3}(x)$ for $\nu\theta_3$ along the complex axis. In this case, x is moderately large ($x = -4.47$) and the power series suffers from catastrophic cancellation giving at least $O(1)$ errors for $|\nu\theta_3| > 3$. The dotted trajectory shows that (18) gives an excellent approximation of (10) as $|\nu\theta_3| \to 0$. As expected though, the approximation diminishes as $|\nu\theta_3|$ increases. On the other hand, (17) gives a poor approximation for small $|\nu\theta_3|$ but improves significantly for larger values. Eventually though the approximation degrades due to the onset of the cancellation in (10). For $|x| < 2$, this error would go to 0 because the power series does not suffer from cancellation in this regime. Finally, the dashed line gives the error between the two asymptotic approximations. While (17) should improve with increasing $|\nu\theta_3|$ for any x, (18) loses validity when $|x| \sim |\nu\theta_3|$. This picture is typical for any x value, the only difference between the location where the different representations gain/lose validity.

We now address the computation of the parameter derivatives of H and thus $\hat{p}(\nu, \Theta)$. In the power series representation (10), uniform convergence implies that

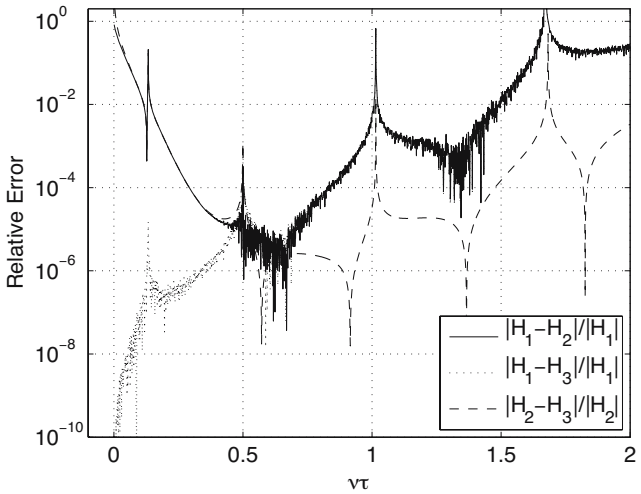

Fig. 1 Relative error versus $\nu\theta_3$ for the three Hermite function representations for $\theta_3 = 5$, $\theta = -4.47$. The *solid line* gives the error between the power series (10) and the Darwin expansion (17), the *dotted line* compares (10) and (18), and the *dashed line* measures the error between (17) and (18)

the infinite series can be differentiated term by term to give a series that is also uniformly convergent:

$$\frac{\partial}{\partial x} H_{\nu\theta_3}(x) = 2\nu\theta_3 H_{\nu\theta_3+1}(x) \ .$$

The same holds for the second derivatives with respect to x although for computation purposes, it would be easier to use (9). For the θ_3 derivatives, direct differentiation of (10) also yields convergent power series.

Parameter derivatives of all types can be computed from the asymptotic expansion (17). Here, convergence is guaranteed because it is generated as a solution to a differential equation (Miller 1955; Bender and Orzag 1978). This is not the case when differentiating (18) with respect to the parameters because it was derived as an approximation to an integral (Lebedev 1972). In practice though, it has never been shown to fail.

Recall the special case, $\theta_2 = 0$ or $X_f = \mu\theta_3$, where the exact pdf is given by (6). Figure 2 displays (6) and the numerically inverted pdf from (16) on a semilog scale. The numerical inversion does a very good job across most of the distribution although we note discrepancies for small t. In this regime though, the probability density is very small, $O(10^{-7})$ and smaller. To the eye, (16) accurately computes the exact distribution.

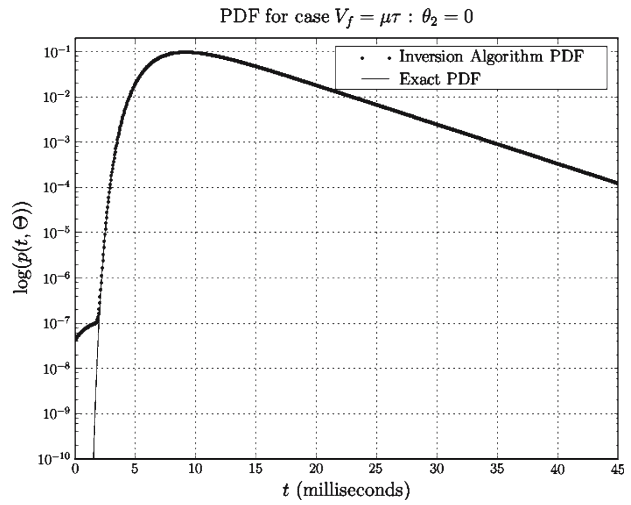

Fig. 2 Probability distribution function (pdf) versus t for $\{X_0, X_f, \mu, \sigma, \tau\} = \{0, 15, 3, 1.5, 5\}$. *Solid line* gives the exact representation, (6) while the markers (*dots*) give the pdf through (16)

4 Numerical results

In this section, numerical results for the algorithm are given. First, we consider the case of estimating the two parameters, (θ_1, θ_2), when θ_3 is fixed. In particular, we focus on the construction and testing of approximate confidence intervals and regions for the estimates. Then, results for the full three-parameter estimation $\theta_{1,2,3}$ are given. Comparisons with the two-parameter case will be discussed. Here, we note that simulated data from (2) is used to generate the first-passage time samples for all parameter estimation studies. The data sets are generated via discrete mapping in time:

$$\begin{aligned} t_i &= i\Delta t \\ W_i &= \sqrt{\frac{2\Delta t}{\theta_3}} e^{t_i/\theta_3} Z_i + W_{i-1} \\ X_i &= (X_0 - \mu\theta_3)e^{-t_i/\theta_3} + \mu\theta_3 + \sigma\sqrt{\frac{\theta_3}{2}} e^{-t_i/\theta_3} W_i \ . \end{aligned} \tag{19}$$

Here, Z_i are standard, independent Gaussian variates and W_i is the position of the Wiener process with $W_0 = 0$. Since the probability that $X_n = X_f$ at some $t_n = n\Delta t$ is zero, linear interpolation is used to find the time of the boundary crossing once X_n first exceeds X_f. It is important to emphasize that the reliability of the sample in representing a true OU diffusion is inversely proportional to Δt. This fact will have an important effect on both the quality of the estimates as well as the confidence intervals and regions as we shall see in the two-parameter estimation case study.

In order to construct approximate confidence intervals and regions for the estimates, we make use of the Hessian evaluated at the MLE, $H(\hat{\Theta})$, or the observed Fisher information matrix:

$$J\left(\hat{\Theta}\right) = -H\left(\hat{\Theta}\right) \ .$$

It has been shown in Iyengar and Mullowney (submitted for publication) that the MLEs of the embedded parameters Θ are asymptotically normal as the number of first-passage time samples, n, tends to ∞, i.e.:

$$\sqrt{J\left(\hat{\Theta}_n\right)}\left(\hat{\Theta}_n - \Theta\right) \to N(0, I) \tag{20}$$

$$J\left(\hat{\Theta}_n\right)^{-1} = \begin{bmatrix} \mathbf{Var}\left(\hat{\theta}_{1,n}\right) & \mathbf{Cov}\left(\hat{\theta}_{1,n}, \hat{\theta}_{2,n}\right) & \mathbf{Cov}\left(\hat{\theta}_{1,n}, \hat{\theta}_{3,n}\right) \\ \mathbf{Cov}\left(\hat{\theta}_{2,n}, \hat{\theta}_{1,n}\right) & \mathbf{Var}\left(\hat{\theta}_{2,n}\right) & \mathbf{Cov}\left(\hat{\theta}_{2,n}, \hat{\theta}_{3,n}\right) \\ \mathbf{Cov}\left(\hat{\theta}_{3,n}, \hat{\theta}_{1,n}\right) & \mathbf{Cov}\left(\hat{\theta}_{3,n}, \hat{\theta}_{2,n}\right) & \mathbf{Var}\left(\hat{\theta}_{3,n}\right) \end{bmatrix}. \tag{21}$$

Here, we adopt the notation $\hat{\Theta}_n = \left(\hat{\theta}_{1,n}, \hat{\theta}_{2,n}, \hat{\theta}_{3,n}\right)^T$ to specify the sample size dependence of the estimate thereby distinguishing it from the iteration of the ML algorithm given by the superscript $\hat{\Theta}^{(n)}$. Typically J is referred to as the covariance or error matrix. I is a vector of ones whose dimension equals the number of parameters being estimated. Using (20) and (21), one can construct individual confidence intervals for each parameter in the usual manner via:

$$\hat{\theta}_{1,n} \pm z_{\frac{1-\alpha}{2}} \sqrt{J(\hat{\Theta}_n)^{-1}_{11}} \tag{22}$$

$$\hat{\theta}_{2,n} \pm z_{\frac{1-\alpha}{2}} \sqrt{J(\hat{\Theta}_n)^{-1}_{22}} \tag{23}$$

$$\hat{\theta}_{3,n} \pm z_{\frac{1-\alpha}{2}} \sqrt{J(\hat{\Theta}_n)^{-1}_{33}} \tag{24}$$

where $z_{\frac{1-\alpha}{2}} = 1.96$ when $\alpha = .95$.[1] The intersection of these regions in the $\theta_1\theta_2\theta_3$ space defines an α percent confidence box. For (22–24), note that the square root operation is taken after computing $J(\hat{\Theta}_n)^1_{ii}$

Alternatively, one can construct more restrictive confidence ellipsoids. To do this, we assume that (13) can be modeled with an ellipsoid around the MLE:

$$F(\theta_1, \theta_2, \theta_3) = a_1\theta_1^2 + a_2\theta_2^2 + a_3\theta_3^2 + 2a_4\theta_1\theta_2 + 2a_5\theta_1\theta_3 + 2a_6\theta_2\theta_3 + 2a_7\theta_1 + 2a_8\theta_2 + 2a_9\theta_3 + a_{10}. \tag{25}$$

Then, use the following facts to determine the parameters, a_i :

- (13) is maximized at $\hat{\Theta}_n$ and thus F is as well. This gives 3 constraints.
- The second derivatives of (13), i.e. the Hessian, must coincide with the second derivatives of F. This gives 6 constraints since the Hessian is symmetric.
- Equating the two functions at the MLE gives 1 constraint and determines the center of the ellipsoid:

$$\ln L(\hat{\Theta}_n|\vec{T}) = F(\hat{\theta}_{1,n}, \hat{\theta}_{2,n}, \hat{\theta}_{3,n}) .$$

Given this, we must now determine which contour of F defines the α% confidence region. To do this, we first compute the square root of the covariance matrix. This is done via eigenvalue decomposition since the covariance matrix is always positive definite. Let $\{\lambda_{max}, v_{max}\}$ be the eigenvalue-eigenvector pair corresponding to the maximum eigenvalue. v_{max} defines the direction of the ellipsoids' semi-major axis. The other eigenvectors define the minor axes. Then, the contour of F, $z_{\frac{1-\alpha}{4}}\sqrt{\lambda_{max}}$ units from the center of the region in the direction of v_{vmax} yields the approximate α% confidence ellipsoid. This contour could equally be determined using the other eigenvalue-eigenvector pairs. Notice here that $z_{\frac{1-\alpha}{4}}$ is used rather than $z_{\frac{1-\alpha}{2}}$.[2] This arises from the fact that confidence ellipsoids measure the joint probability of the true parameters residing within the region. Thus, a larger z value must be used to accommodate the combined variation. For uncorrelated parameters, $z_{\frac{1-\alpha}{4}}$ is sufficient for a reasonable approximation. For highly correlated parameters, $z > z_{\frac{1-\alpha}{4}}$ is necessary.

4.1 2-D results: fixed θ_3

The results in of the previous section instill confidence in the accuracy and reliability of the inversion algorithm. We now proceed with an analysis of the MLEs in the case where θ_3 is fixed. Figure 3(a) shows a histogram of a simulated first-passage time sample for $\theta_1 \approx -3.354$ and $\theta_2 \approx 1.118$. Figure 3(b) displays the true pdf evaluated on a grid of points in t with the exact parameters used to generate the sample in Fig. 3(a). It also shows the ML pdf evaluated at the first-passage sample shown in Fig. 3(a) using the MLE parameters: $\hat{\theta}_{1,n} \approx -3.231$, $\hat{\theta}_{2,n} \approx 1.134$. Each density is calculated from the inversion technique (16). To the eye, the ML pdf does a good job of approximating the true distribution although deviations are certainly evident especially near the maximum.

[1] The decimal and percentage interpretations for α are used interchangeably hereafter.

[2] For $\alpha = .95$, $z_{\frac{1-\alpha}{4}} = 2.243$

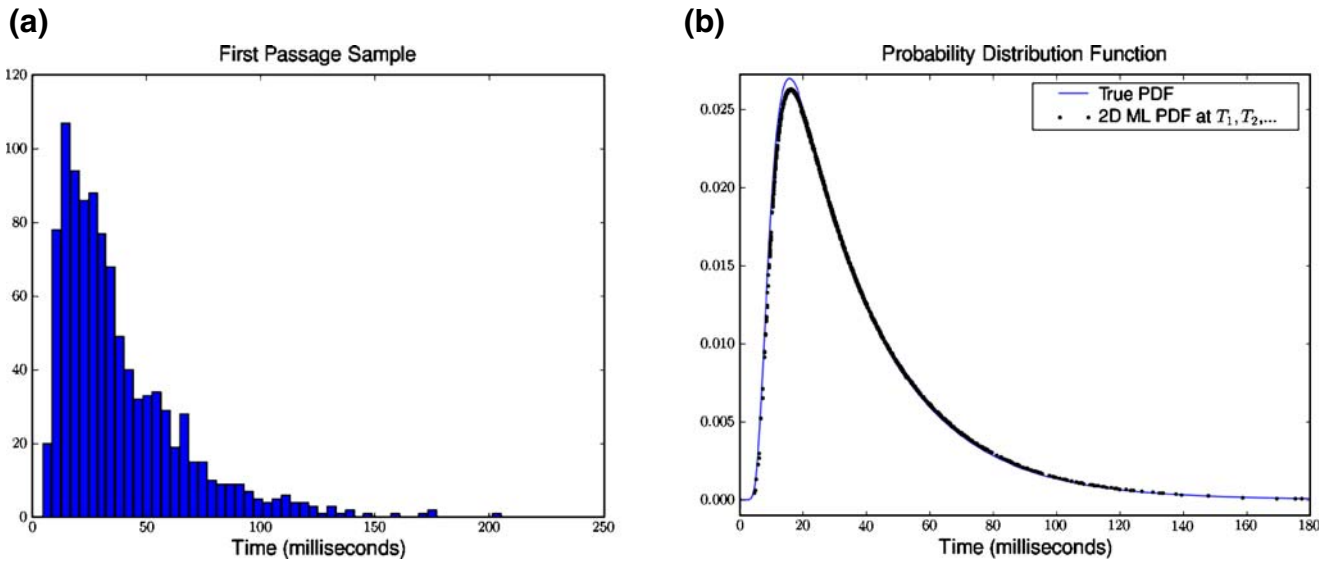

Fig. 3 $\{X_0, X_f, \mu, \sigma, \tau\} = \{0, 20, 3, 2, 5\} \Leftrightarrow \Theta = (-3.354, 1.118, 5)^T$. Simulation parameters: $n = 1000$ samples with time step $\Delta t = 10^{-4}$. (**a**) Distribution of first-passage times and (**b**) ML pdf using $\hat{\theta}_{1,n} \approx -3.231, \hat{\theta}_{2,n} \approx 1.134$ (*dots*) versus true pdf (*solid line*). Each distribution was generated from (16)

Contours of the log-likelihood function (13) are shown in Fig. 4 in the $\theta_1\theta_2$ plane for two cases of θ_2. The true values of the parameters, (θ_1, θ_2), and the estimate $(\hat{\theta}_{1,n}, \hat{\theta}_{2,n})$ are also given for reference. The 95% and 99% confidence regions/ellipses are plotted as thick, dark contours. The 95% and 99% confidence intervals/boxes are also shown to contrast the ellipses. At first we glance, we note that the contours are approximately elliptical and that the MLE resides at the center of the concentric ellipses. Note the different scales of the axes in the two pictures. The ratio of the ellipse's major to minor axes is roughly 10 : 1 in Fig. 4(a) and 4 : 1 in Fig. 4(b). These pictures are representative of the log-likelihood (13) when $\theta_2 > 0$ (Fig. 4(a)) and $\theta_2 < 0$ (Fig. 4(b)). In Fig. 4(a), $\ln L(\Theta|\vec{T})$ is broad in the θ_1 direction and sharp in the θ_2 direction. These features are a byproduct of the mean reverting nature of the OU process. As $\theta_1 \to -\infty$, the process has a stronger inclination to revert back to $\mu\theta_3$ as indicated by (1). Therefore, large changes in θ_1 in either direction should have little effect on the first-passage time distribution. On the other hand, large positive increases in θ_2 yields significantly longer first-passage times since the natural tendency of the process is to revert back to $\mu\theta_3$ (recall that $\theta_2 > 0$ gives a threshold greater than $\mu\theta_3$). Then given enough samples ($n = 1000$ for both cases), it is not surprising that one has enough information to make a precise estimate of the θ_2 as indicated by the length of the semi-minor axis of the 95% confidence ellipsoid or the size of the 95 confidence interval in θ_2. If one chooses $X_f < \mu\theta_3$ as in Fig. 4(b), the aspect ratio decreases dramatically as noted above, however, the correlation between θ_1 and θ_2 increases as measured by the rotation angle of the ellipse. The correlation increases as $\theta_2 \to \theta_1$ and the ellipses tend to circles. Due to this higher correlation, it is likely that the calculated 95% confidence ellipse in Fig. 4(b) underestimates the true region.[3]

Now we consider the quality of the estimations with regard to the confidence intervals and regions. In particular, we expect asymptotic normality in the parameters for large sample sizes and we expect an α% confidence interval/region to contain the true parameter in α% of a set of trials. To test this, 100 simulations are run with each consisting of 1000 first-passage times. Asymptotic normality is measured by looking at the distribution of (20) using the Anderson-Darling test at the 95% level (Stephens 1974). The confidence regions/intervals can be tested by counting the number of times the true parameter(s) fall within the appropriate regions for all simulations. These tests are performed as a function of the simulation time increment Δt. The results are shown in Table 1 Clearly, refining the simulations with smaller times steps yields statistics consistent with the theoretical predictions when considering the asymptotic normality of the estimates, and the predictive power of the confidence intervals and regions/ellipses. More importantly, the data suggests

[3]The first-passage sample in Fig. 3 gives the MLEs and contours in Fig. 4(a). A different first-passage sample was used in Fig. 4(b).

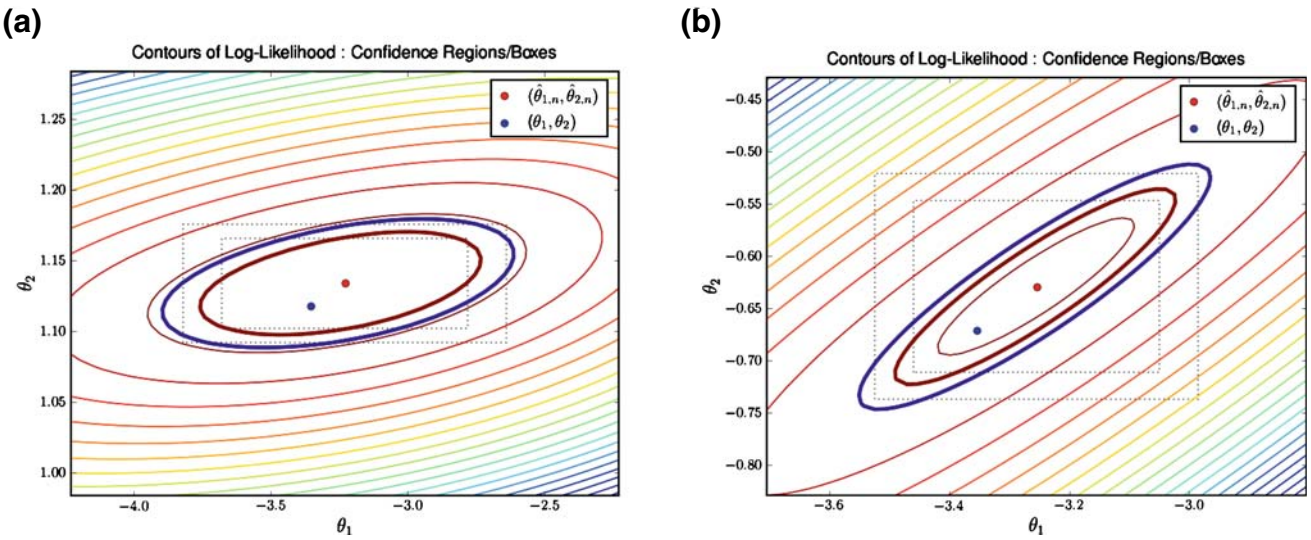

Fig. 4 $\{X_0, \mu, \sigma, \tau\} = \{0, 3, 2, 5\} \Rightarrow \theta_1 = -3.354$. Simulation parameters: $n = 1000$ samples with time step $\Delta t = 10^{-4}$. Contours of log-likelihood function (13) in the $\theta_1\theta_2$ plane. The thick contours give the 95% and 99% confidence ellipses. (**a**) $X_f = 20 \Leftrightarrow \theta_2 = 1.118$ and (**b**) $X_f = 12 \Leftrightarrow \theta_2 = -.67082$

that simulated data is only reliable when small time steps are taken. It should also be mentioned that the results in Table 1 are consistent for different values of θ_1 and θ_2 in all aspects

4.2 3-D results

In the previous section, estimates for θ_1 and θ_2 were computed from first-passage time samples for fixed θ_3. This may incorrectly constrain the optimization since (3) is clearly a function of three identifiable parameters: $\theta_{1,2,3}$. Thus, we now allow for variation in θ_3 and see what effect this has on the estimates. First, the ML pdf is computed from the sample shown in Fig. 3(a). The algorithm yields the estimates: $\hat{\Theta}_n = (-2.653, 1.035, 5.984)^T$. The associated pdf is shown in Fig. 5. To the eye there is little discernible difference with the two-parameter estimation given in Fig. 3(b). Although, once again, there are deviations from the true pdf around the maximum. The degree of similarity can be quantified by computing the r.m.s error between the true pdf evaluated at the first-passage times and the ML pdf for both the two and three parameter estimations, i.e.:

$$\epsilon_j = \sqrt{\sum_{i=1}^{n} \left[p(T_i, \Theta) - p_j(T_i, \hat{\Theta}_n) \right]^2}$$

The subscript j denotes the two and three parameter estimates respectively and $p(T_i, \Theta)$ is the true pdf with parameters $\Theta = (-3.354, 1.118, 5)^T$. For the data given in Fig. 3 and Fig. 5, we find that $\epsilon_2 \approx 3.7 * 10^{-4}$ and $\epsilon_3 \approx 6.7 * 10^{-4}$. Both errors are of the same order as the

Table 1 $\{X_0, X_f, \mu, \sigma, \tau\} = \{0, 20, 3, 2, 5\} \Leftrightarrow \Theta = (-3.354, 1.118, 5)^T$

Δt	θ_1				θ_2				(θ_1, θ_2)		
	Interval			Normal	Interval			Normal	Ellipse		
	90	95	99	95	90	95	99	95	90	95	99
10^{-1}	66	75	92	No	6	8	25	No	6	12	31
10^{-2}	86	95	99	No	74	83	92	No	73	82	91
10^{-3}	90	98	100	Yes	84	94	98	No	83	90	99
10^{-4}	94	97	100	Yes	87	95	100	Yes	89	94	100

Number of times θ_j falls within α% confidence interval, the number of times (θ_1, θ_2) falls within α% confidence region/ellipse, and the asymptotic normality of (20) for 100 samples, 1000 first-passage times per sample generated with time step Δt.

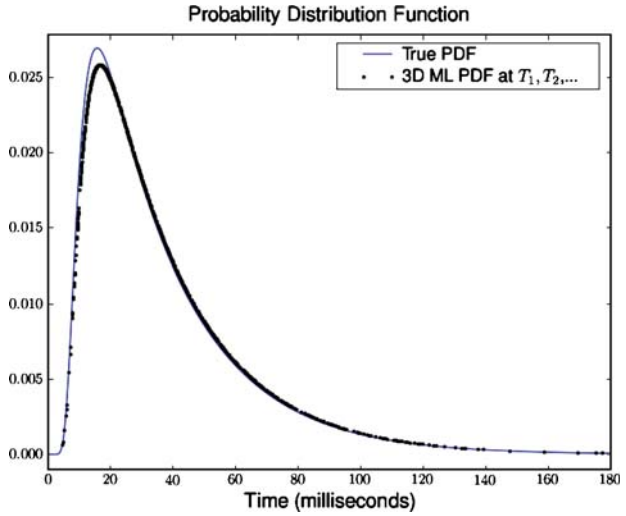

Fig. 5 $\{X_0, X_f, \mu, \sigma, \tau\} = \{0, 20, 3, 2, 5\} \Leftrightarrow \Theta = (-3.354, 1.118, 5)^T$. Simulation parameters: $n = 1000$ samples with time step $\Delta t = 10^{-4}$. ML pdf (*dots*) using $\hat{\Theta}_n = (-2.653, 1.035, 5.984)^T$ versus true pdf (*solid line*). Each distribution is generated from (16)

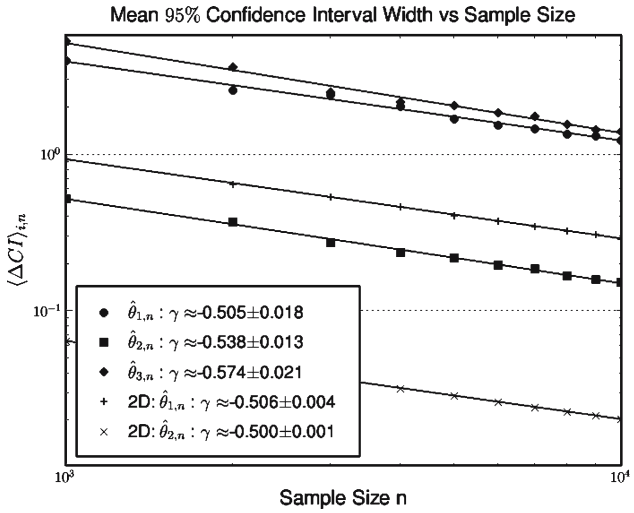

Fig. 6 $\{X_0, X_f, \mu, \sigma, \tau\} = \{0, 20, 3, 2, 5\} \Leftrightarrow \Theta = (-3.354, 1.118, 5)^T$. Simulation parameters: $\Delta t = 10^{-4}$. Average width of the 95% confidence interval, $\langle \Delta CI \rangle_{i,n}$, versus the number of samples for each parameter and each algorithm

trapezoidal rule integration error $O(10^{-4})(\Delta v = .015)$ in the inversions.

The previous computations measured the relative accuracy of the ML pdf for the two and three parameter algorithms. Now, we consider their respective degrees of variability, i.e. the confidence intervals for the estimates. To do this, we compute the 95% confidence intervals from (22–24) as a function of the sample size n. 20 unique simulations are performed for each n. For each sample size, we tabulate the number of times, N, that the true parameter value resides within the 95% interval as well as the average width (resolution) of the 95% confidence interval:

$$\langle \Delta CI \rangle_{i,n} = \frac{2z_{\frac{1-\alpha}{2}}}{20} \sum_{j=1}^{20} \sqrt{J(\hat{\Theta}_{n_j})^{-1}_{ii}} \,. \tag{26}$$

The results are given in Table 2 for both the two and three-parameter algorithms. From this data, it is obvious that the two-parameter algorithm provides a much higher resolution of the true parameter values for each sample size. This is because $\langle \Delta CI \rangle_{1,n}$ and $\langle \Delta CI \rangle_{2,n}$ are $4-5$ and $7-9$ times larger than their respective two-parameter algorithm counterparts. Even at 10000 samples, $\langle \Delta CI \rangle_{1,n} \approx 30\%$ larger than the same interval using the two-parameter algorithm and 1000 samples. This suggests that a much larger range of parameters is capable of representing the true distribution with reasonable accuracy. Thus, if θ_3 is not known a priori, then a significantly larger sample size is required to resolve the parameters with the same accuracy as the two-parameter algorithm.

In Fig. 6, $\langle \Delta CI \rangle_{i,n}$ is plotted against n for each parameter and each algorithm on a loglog scale. Each data set is fit with a function of form $\langle \Delta CI \rangle_{i,n} = \beta n^{\gamma}$. The results show that $\langle \Delta CI \rangle_{i,n} \sim n^{-1/2}$ as expected. Extrapolating these trends to larger n, we find that using the three-parameter algorithm with roughly 16000 samples gives the same resolution as the two-parameter algorithm with 1000 samples on $\hat{\theta}_{1,n}$. For $\hat{\theta}_{2,n}$, nearly 50000 samples are required.

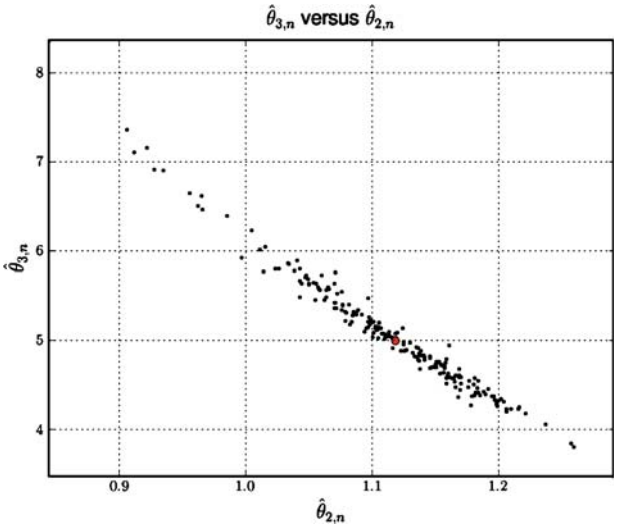

Fig. 7 $\{X_0, X_f, \mu, \sigma, \tau\} = \{0, 20, 3, 2, 5\} \Leftrightarrow \Theta = (-3.354, 1.118, 5)^T$. Simulation parameters: $\Delta t = 10^{-4}$. $\hat{\theta}_{3,n}$ versus $\hat{\theta}_{2,n}$ for all sample sizes and simulations in Table 2. The true value of Θ (i.e. the simulation parameters) is given by the *red dot*

Table 2 $\{X_0, X_f, \mu, \sigma, \tau\} = \{0, 20, 3, 2, 5\} \Leftrightarrow \Theta = (-3.354, 1.118, 5)^T$

n	$\hat{\theta}_{1,n}$				$\hat{\theta}_{2,n}$				$\hat{\theta}_{3,n}$	
	3 Parameter		2 Parameter		3 Parameter		2 Parameter		3 Parameter	
	N	$\langle \Delta CI \rangle_{1,n}$	N	$\langle \Delta CI \rangle_{1,n}$	N	$\langle \Delta CI \rangle_{2,n}$	N	$\langle \Delta CI \rangle_{2,n}$	N	$\langle \Delta CI \rangle_{3,n}$
1000	18	3.972	19	0.936	20	0.524	19	0.064	19	5.292
2000	19	2.588	19	0.646	20	0.370	17	0.045	20	3.612
3000	18	2.395	19	0.534	20	0.273	20	0.037	20	2.508
4000	20	2.029	19	0.466	20	0.237	19	0.032	20	2.180
5000	19	1.679	18	0.406	19	0.218	20	0.029	19	2.045
6000	20	1.536	18	0.374	20	0.198	18	0.026	20	1.846
7000	19	1.458	20	0.346	19	0.187	18	0.024	19	1.747
8000	20	1.350	17	0.324	20	0.169	20	0.023	19	1.561
9000	20	1.323	20	0.307	19	0.159	18	0.021	19	1.449
10000	19	1.230	18	0.290	19	0.153	20	0.020	19	1.407

Simulation parameters; $\Delta t = 10^{-4}$, 20 simulations with n samples in each simulation. Number of times, N, that the true parameter falls within the 95% confidence interval for $\hat{\theta}_{1,n}$, $\hat{\theta}_{2,n}$, and $\hat{\theta}_{3,n}$ for the two and three parameter algorithms. Also, the average width of the 95% confidence interval, $\langle \Delta CI \rangle_{i,n}$ versus n.

A close inspection of the data yielding the statistics in Table 2 also reveals a strong correlation between $\hat{\theta}_{2,n}$ and $\hat{\theta}_{3,n}$. Figure 7 gives a scatter plot of $\hat{\theta}_{3,n}$ versus $\hat{\theta}_{2,n}$ for every simulation in every sample size n. A marker denoting the true parameter value used for the simulation is given for reference. These results are not at all surprising considering the physics of the OU process, although the strength of the correlation is surprising. To see this, imagine two distinct OU processes labeled A and B that are subject to the same input drift μ and variance σ. Moreover, let them have the same resting and firing thresholds X_0 and X_f. Lastly, let A have a larger time constant, θ_3. Since A has a larger time constant, the state variable of A, $X_A(t)$, will decay more slowly than $X_B(t)$ and thus can be expected to have shorter first-passage times. Shorter first passage-times implies a smaller value of θ_2 due to mean-reversion. The same argument can be applied in reverse thus explaining the strong negative correlation between these parameters.

Although a variety of additional parameter studies could be performed at this juncture, we believe that the main differences between the two and three-parameter algorithms have been demonstrated. Moreover, since the results are typical for any set of parameters, Θ, we expect that our conclusions are quite general.

5 Discussion

We have given an algorithm for computing parameter estimates for the three identifiable parameters of the OU–LIF model using only ISI data. The computational approach is based on a numerical inversion of the Laplace transform of the first-passage time pdf. It then uses ML to find the optimal parameters. This technique differs significantly from other estimation approaches on LIF models in a variety of ways. Other recent approaches (Lánský et al. 2006; Paninski et al. 2004; Jolivet et al. 2004) make use of more data, like subthreshold voltage traces or stimulus current traces, to do an optimization over a (much) larger parameter space. Our approach is computationally efficient, robust in the parameter space, and can easily generate standard errors for the estimates. However, precise estimates of the parameters requires many ISI samples. In contrast, (Lánský et al. 2006) made estimates from a single ISI.

In Ricciardi and Sato (1988), Inoue et al. (1995), Ditlevsen and Ditlevsen (2006), estimations based on ISI data focused on the two-parameter case where the time constant, or leakage parameter τ is known a priori. However, τ is clearly an identifiable parameter from ISI data as shown here, thus the full three-parameter estimation is necessary. Our results show unequivocably that the addition of the τ parameter in the estimation procedure dramatically decreases the resolution of $\theta_{1,2}$ for a fixed sample size. Here, a decrease in resolution corresponds to an increase in the width of the parameters' 95% confidence intervals. In fact, our computations show that roughly 15 times more samples are required to resolve θ_1 to the same accuracy as the two-parameter algorithm. For θ_2, the situation is considerably worse since roughly 50 times more samples

are required for comparable resolution between the algorithms. However, since the typical neuron can yields between $500-50000$ samples in the spike train (Ricciardi and Sato 1988), this algorithm could resolve the identifiable parameters with good accuracy.

The difficulty in estimating τ is consistent with several recent studies in the literature. For instance, Lánský et al. (2006) noted that while their estimated firing threshold was consistent with that of Inoue et al. (1995) (with similar experimental conditions), their estimate of the mean input had a significantly smaller range which was closer to 0. This yielded an estimate of τ seven times larger than the value selected a priori in Inoue et al. (1995). As noted above, a priori assumptions on τ based only ISI data can be misleading. Our computations show that the size of 95% confidence for the normalized input (θ_1) and the normalized threshold (θ_2) are at least an order of magnitude larger when the full three parameter estimation is performed. Thus, algorithms which work on a restricted subset of the full parameter space cannot necessarilly be trusted. Furthermore, Lánský et al. (2006) computed τ with two algorithms from a single ISI under the OU approximation. For each method, the spread was rather large. In a time frame on the scale of a single spike (subseconds), it is not clear whether the input stimulus current can be written in such a simple form. This could give rise to the large fluctuations in their results. However, this approximation may be more suitable when many ISIs are recorded.

Similar problems in estimating τ were observed in Jolivet and Gerstner (2004). They compared the effective membrane time constant of a conductance-based IF model containing reversal potentials to a HH type model. They found that in regions where excitatory discharge dominates inhibitory, CIF agrees very well with HH models. When there is no clear dominant factor, the results are drastically different. Of course, our analysis of the OU model does not contain information on excitatory or inhibitory discharge and thus we cannot make a direct comparison between the studies. However, extensions of our work to the more general LIF models containing reversal potentials will enable more meaningful comparisons. For instance, models that include inhibitory reversal potentials while neglecting the excitatory are a good approximation to neuronal behavior (Lánský et al. 1995). An example of this is the Feller process;

$$\begin{aligned} \mathbf{d}X &= \mu(X)\mathbf{d}t + \sigma(X)\mathbf{d}W \\ \mu(X) &= \alpha - \beta X \\ \sigma^2(X) &= \sigma^2(X - X_I)\ . \end{aligned}$$

Here X_I is the inhibitory reversal potential and $1/\beta$ is the effective membrane time constant of the neuron. Initial investigations into this model reveal four identifiable parameters from ISI data. Moreover, the lack of time dependence in $\mu(X)$ and $\sigma(X)$ implies time-homogeneity in the evolution of the membrane potential. Thus, we need only find the Laplace transform of the transition density, from a corresponding Chapman–Kolmogorov equation, in order to compute the Laplace transform of the first-passage time density. Any attempt in this direction will benefit from the analysis in Capocelli and Ricciardi (1972). which gives the necessary conditions for a function to be a pdf of a first-passage time random variable for any time-homogeneous Markov diffusion process. Since similar results are given for the Laplace transform, it may be possible to analytically compute the Laplace transform of the first-passage time pdf for the Feller process. Then, an inversion technique similar to ours can be used for the parameter estimation from ISI data only. In this sense, we assert that our technique should generalize to any time-homogenous diffusion process provided the mathematical details, like identifiability and Laplace transform computation, are tractable.

Unfortunately, it is not clear whether this approach will apply to situations with time-dependent input currents. Although the problem can be recast as an OU LIF that defines the subthreshold membrane potential, the spike-generation threshold conditions become time-dependent. In this case, all the mathematical machinery based on time-homogoneity break down and a Laplace transform is not the right approach.

Our results also indicate that simulated first-passage time data can only be trusted when the time step of the simulation is small. This suggests the need for greater computational power when simulating and testing OU processes. This point is further reinforced by the fact that the data in Table 2 took roughly 30 h to generate on a dual-core multiprocessing cpu. Moreover, it is reasonable to conclude that this result would hold for any type of stochastic IF model with comparable complexity like the CIF or Feller process thus suggesting the need for greater computational power when simulating stochastic neural behavior. A natural extension of this work to coupled, stochastic IF models would impose an even greater computational burden thus demanding parallel implementations for the purpose of simulation. This demand will be unavoidable as the size of the coupled network grows, and complexity of the cellular models increases.

The availability of robust and efficient algorithms that relate the measurable data to the biophysical parameters would be an invaluable tool for any

neurobiologist. In particular, it would allow for rapid model evaluation and comparison based on rigorous estimates of the parameters and their standard errors. Moreover, changes in experimental conditions that impact the neuron's behavior could be evaluated with algorithms of this type enabling a more precise understanding of the biophysics.

References

Abramowitz, M., & Stegun, I. R. (1972). *Handbook of mathematical functions* (9th ed.). New York: Dover Publications Inc.

Arnold, L. (1974). *Stochastic differential equations: Theory and applications*. New York: John Wiley and Sons.

Bender, C. M., & Orzag, S. A. (1978). *Advanced mathematical methods for scientists and engineers*. New York: McGraw Hill.

Broyden, C. G. (1965). A class of methods for solving nonlinear simultaneous equations. *Mathematical Computing*, *19*, 577–593.

Burkitt, A. N., & Clark, G. M. (2000). Calculation of interspike intervals for integrate-and-fire neurons with poisson distribution of synaptic inputs. *Neural Computation*, *12*, 1789–1820.

Burkitt, A. N. (2001). Balanced neurons: Analysis of leaky integrate-and-fire neurons with reversal potentials. *Biological Cybernetics*, *85*, 247–255.

Burkitt, A. N. (2006a). A review of the integrate-and-fire neuron model: I. Homogenous synaptic input. *Biological Cybernetics*, *95*, 1–19.

Burkitt, A. N. (2006b). A review of the integrate-and-fire neuron model: II. Inhomogenous synaptic input and network properties. *Biological Cybernetics*, *95*, 97–112.

Capocelli, R. M., & Ricciardi, L. M. (1972). On the inverse of the first passage time probability problem. *Journal of Applied Probability*, *9*, 270–287.

Churchill, R. V. (1981). *Operational mathematics*. New York: McGraw Hill.

D'Amore, L., Laccetti, G., & Murli, A. (1999). An implementation of a fourier series method for the numerical inversion of the laplace transform. *ACM Transactions on Mathematical Software*, *25*, 279–305.

Darling, D., & Siegert, A. (1953). The first passage problem for a continuous markov process. *Annals of Mathematical Statistics*, *24*, 624–639.

De Hoog, F. R., Knight, J. H., & Stokes, A. N. (1982). An improved method for numerical inversion of laplace transforms. *SIAM Journal of Scientific and Statistical Computing*, *3*, 357–366.

Ditlevsen, S., & Ditlevsen, O. (2006). Parameter estimation from observations of first-passage times of the Ornstein–Uhlenbeck process and the feller process. *Presented at the fifth computational stochastic mechanics conference*.

Hodgkin, A. L., & Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerves. *Journal of Physiology*, *117*, 500–544.

Inoue, J., Sato, S., & Ricciardi, L. (1995). On the parameter estimation for diffusion models of single neuron's activities. *Biological Cybernetics*, *73*, 209–221.

Jolivet, R., & Gerstner, W. (2004). Predicting spike times of a detailed conductance-based neuron model driven by stochastic spike arrival. *Journal of Physiology* (*Paris*), *98*, 442–451.

Jolivet, R., Lewis, T. J., & Gerstner, W. (2004). Generalized integrate-and-fire models of neuronal activity approximate spike trains of a detailed model to a high degree of accuracy. *Journal of Neurophysiology*, *92*, 959–976.

Jolivet, R., Rauch, A., Lüscher, H. R., & Gerstner, W. (2006). Predicting spike timing of neocortical pyramidal neurons by simple threshold models. *Journal of Computational Neuroscience*, *21*, 35–49.

Kano, P. O., Moysey, B., & Moloney, J. V. (2005). Application of weeks method for the numerical inversion of the laplace transform of the matrix exponential. *Computational Mathematical Sciences*, *3*, 335–372.

Karlin, S., & Taylor, H. (1981). *A second course in stochastic processes*. New York: Academic Press.

Keat, J., Reinagel, P., Reid, R. K., & Meister, M. (2001). Predicting every spike: A model for the responses of visual neurons. *Neuron*, *30*, 803–817.

Lánský, P., Sacerdote, L., & Tomassetti, F. (1995). On the comparison of Feller and Ornstein–Uhlenbeck models for neural activity. *Biological Cybernetics*, *73*, 457–465.

Lánský, P., Sanda, P., & He, J. (2006). The parameters of the stochastic leaky integrate-and-fire neuronal model. *Journal of Computational Neuroscience*, *21*, 211–223.

Lebedev, N. N. (1972). *Special functions and their applications*. New York: Dover Publications.

Lehmann, E. L. (1983). *Theory of point estimation*. New York: John Wiley and Sons.

Miller, J. C. P. (1955). *National Physical Laboratory, tables of Weber parabolic cylinder functions*. London: Her Majesty's Stationary Office.

Morris, C., & Lecar, H. (1981). Voltage oscillations in the barnacle giant muscle fiber. *Biophysical Journal*, *35*, 193–213.

Paninski, L., Pillow, J. W., & Simoncelli, E. P. (2004). Maximum likelihood estimation of a stochastic integrate-and-fire neural encoding model. *Neural Computation*, *16*, 2533–2561.

Pillow, J. W., Paninski, L., Uzzel, V. J., Simoncelli, E. P., & Chichilnisky, E. J. (2005). Prediction and decoding of retinal ganglion cell responses with a probabilistic spiking model. *The Journal of Neuroscience*, *25*(47), 11003–11013.

Plesser, H. E., & Tanaka, S. (1997). Stochastic resonance in a model neuron with reset. *Physics Letters A*, *225*, 228–234.

Rabinovich, M. I., Varona, P., Selverston, A. I., & Abarbanel, H. D. I. (2006). Dynamical principles in neuroscience. *Reviews of Modern Physics*, *70*, 1213–1265.

Ricciardi, L., & Sacerdote, L. (1977). The Ornstein–Uhlenbeck process as a model for neuronal activity. *Biological Cybernetics*, *35*, 1–9.

Ricciardi, L., & Sato, S. (1988). First passage time density and moments of the Ornstein–Uhlenbeck process. *Journal of Applied Probability*, *25*, 43–57.

Sharp, A. A., O'Neil, M. B., Abbott, L. F., & Marder, E. (1993a). The dynamic clamp: Artificial conductances in biological neurons. *Trends in Neuroscience*, *16*, 389–394.

Sharp, A. A., O'Neil, M. B., Abbott, L. F., & Marder, E. (1993b). The dynamic clamp: Computer-generated conductances in real neurons. *Journal of Neurophysiology*, *69*, 992–995.

Shimokawa, T., Pakdaman, K., & Sato, S. (1999). Time-scale matching in the response of a leaky integrate-and-fire neuron model to periodic stimulus with additive noise. *Physical Review E*, *59*, 3427–3443.

Siegert, A. J. F. (1951). On the first passage time probablity functioin. *Physical Review*, *81*, 617–623.

Stein, R. B. (1965). A theoretical analysis of neuronal variability. *Biophysical Journal*, *5*, 173–194.

Stephens, M. A. (1974). Edf statistics for goodness of fit and some comparisons. *Journal of the American Statistical Association*, *69*, 730–737.

Tuckwell, H. (1988). *Introduction to theoretical neurobiology. Volume 2: Nonlinear and stochastic theories*. Cambridge: Cambridge University Press.

Uhlenbeck, G., & Ornstein, L. (1954). On the theory of brownian motion (1930). In: N. Wax (Ed.), *Selected papers in noise and stochastic processes*. New York: Dover Publications.

Weeks, W. T. (1966). Numerical inversion of the laplace transform using laguerre functions. *Journal of Association of Computational Mathematics*, *13*, 419–429.