Parameter estimation in nonlinear stochastic differential equations

J. Timmer

Fakultät für Physik, Hermann – Herder – Str. 3, 79104 Freiburg, Germany

and

Freiburger Zentrum für Datenanalyse und Modellbildung, Eckerstr. 1, 79104 Freiburg, Germany

e-mail: jeti@fdm.uni-freiburg.de

Phone: +49/761/203-5829

FAX: +49/761/203-5967

(January 8, 2000)

Abstract

We discuss the problem of parameter estimation in nonlinear stochastic differential equations based on sampled time series. A central message from the theory of integrating stochastic differential equations is that there exists in general two time scales, i.e. that of integrating these equations and that of sampling. We argue that therefore maximum likelihood estimation is computational extremely expensive. We discuss the relation between maximum likelihood and quasi maximum likelihood estimation. In a simulation study, we compare the quasi maximum likelihood method with an approach for parameter estimation in nonlinear stochastic differential equations that disregards the existence of the two time scales.

PACS: 02.50.Ey, 02.50.Ph, 05.45.-a, 05.45.Tp

Keywords: Parameter estimation, nonlinear stochastic differential equations

I. INTRODUCTION

The analysis of complex systems by nonlinear deterministic differential equations has attracted much attention in recent years. Given a parameterized differential equation, best-fit parameters can be obtained by a least-squares minimization, see e.g. [5,14,2,3,27].

Often, however, the dynamics does not follow a strict deterministic law. In dissipative systems, due to fluctuation-dissipation theorems, thermal noise might have to be taken into account [17]. Furthermore, in open nonequilibrium systems like in biology and economy, the dynamics is often also exposed to highdimensional, effectively random, influences, see [27] for an example. This calls for a description by nonlinear stochastic differential equations (SDE), respectively their corresponding Fokker-Planck equations [12,29,16]. While parameter estimation in linear stochastic and nonlinear deterministic differential equations even for data covered by additive observational noise is well known, see [27] for a review, the estimation of parameters in nonlinear SDEs is still under development. In this paper we discuss three approaches for parameter estimation in nonlinear SDEs. The results carry over to modeling by Fokker-Planck equations.

The organization of this paper is as follows: In the next section, we briefly summarize the theory of integrating SDEs and exemplify its practice for the van der Pol oscillator undergoing stochastic forcing. In Section III we discuss three approaches for parameter estimation in nonlinear stochastic differential equations. In Section IV we compare two of these approaches in a simulation study using the stochastic van der Pol oscillator.

II. INTEGRATING NONLINEAR STOCHASTIC DIFFERENTIAL EQUATIONS

A. Theory

A stochastic differential equation (SDE) with parameter vector $\vec{\theta}$ is given by:

$$\dot{\vec{x}} = \vec{a}(\vec{\theta}, \vec{x}) + \underline{b}(\vec{\theta}, \vec{x})\, \vec{\epsilon} \quad , \tag{1}$$

where $\vec{\epsilon}$ denotes uncorrelated Gaussian noise or, in mathematical terms, the increment of Brownian motion. This general form includes additive and multiplicative dynamical noise. $\vec{a}(\vec{\theta},\vec{x})$ is usually denoted as (deterministic) drift term, $\underline{b}(\vec{\theta},\vec{x})$ is called the diffusion term.

The integration of a SDE is not straightforward. This is due to the mathematical problems of evaluating integrals which involves the dynamical noise $\vec{\epsilon}$, see [19] for a brief introduction and [18] for a detailed discussion. Applying the same ideas underlying higher order integration schemes for deterministic differential equation like Runge-Kutta [23] to SDEs leads to hardly treatable stochastic integrals given in [16]. Thus, only low order integration schemes can be used. The lowest order so-called Euler-scheme for Eq. (1) is given by :

$$\vec{x}(t+\delta t) = \vec{x}(t) + \delta t\, \vec{a}(\vec{\theta},\vec{x}(t)) + \sqrt{\delta t}\, \underline{b}(\vec{\theta},\vec{x}(t))\vec{\epsilon}(t) \quad , \tag{2}$$

which is of order 1/2 for multiplicative noise and of order 1 for additive noise. The characteristics of integration schemes for SDEs is the $\sqrt{\delta t}$ term which results from the integration rules for white noise [18].

For the task of parameter estimation, we assume that the system under observation can be adequately described by Eq. (1) with unknown parameter vector $\vec{\theta}$. The process is observed at discrete time points given by the sampling interval Δt. For nonlinear deterministic differential equations it is usually possible to chose identical time steps for the integration and the sampling. The necessity of using a low order scheme for SDEs means that the integration step size δt is usually much smaller than the sampling interval Δt by which the time series is recorded. Thus, while Eq. (2) can be written as

$$\vec{x}(t+\delta t) = \vec{h}(\vec{x}(t)) + \nu(t) \quad , \tag{3}$$

with $\vec{h}(\vec{x}(t))$ a function that can be related back to the parameter vector $\vec{\theta}$ of $\vec{a}(\vec{\theta},\vec{x}(t))$ and $\nu(t)$ uncorrelated Gaussian noise, on the time scale Δt of sampling the relationships are more intricate. While it is still possible to formulate the dependence between present and future states as

$$\vec{x}(t+\Delta t) = \vec{g}(\vec{x}(t)) + \eta(t) \quad , \tag{4}$$

the parameter vector $\vec{\theta}$ of $\vec{a}(\vec{\theta}, \vec{x}(t))$ can, in general, not be inferred from $\vec{g}(\vec{x}(t))$. Furthermore, $\eta(t)$, while still uncorrelated with zero mean, does in general not represent Gaussian noise. In other words, in nonlinear SDEs the relation between the mean of the conditional density $p(\vec{x}(t+\Delta t)|\vec{x}(t))$ and the drift-term $\vec{a}(\vec{\theta}, \vec{x}(t))$ is not explicitly known (The analogous problem is given for modeling such systems by their corresponding Fokker-Planck equation). Furthermore, the conditional density $p(\vec{x}(t+\Delta t)|\vec{x}(t))$ is, in general, not Gaussian although $\vec{\epsilon}(t)$ in Eqs. (1,2) is Gaussian. For these reasons parameter estimation in SDEs is a non-trivial task.

B. An example

To exemplify the practical issues of integrating SDEs, we choose the van der Pol oscillator [?]:

$$\ddot{x} = \mu\,(1 - x^2)\,\dot{x} - x, \qquad \mu > 0. \tag{5}$$

This system exhibits a limit cycle due to the amplitude dependent change of the sign of the damping term. We expose it to a random force of unit variance, leading to

$$\dot{x}_1 = x_2 \tag{6}$$

$$\dot{x}_2 = \mu\,(1 - x_1^2)\,x_2 - x_1 + \epsilon \quad , \tag{7}$$

where x_1 denotes the location and x_2 the velocity. The Euler integration scheme for Eqs. (6,7) is given by

$$x_1(t + \delta t) = x_1(t) + \delta t\, x_2(t) \tag{8}$$

$$x_2(t + \delta t) = x_2(t) + \delta t\,(\mu\,(1 - x_1^2(t))\,x_2(t) - x_1(t)) + \sqrt{\delta t}\,\epsilon(t) \quad . \tag{9}$$

In the following we chose $\mu = 3$. To obtain a sampled time series of the system, one has to decide on δt and Δt. Δt should be chosen so that the process is reasonably sampled. Characteristics of the stochastic van der Pol oscillator and related perturbed limit cycles have been investigated in [20,22]. These authors have shown that the mean period of this system is slightly smaller than the period of the corresponding deterministic system which is approximately 9 s. By choosing $\Delta t = 0.5$, we obtain approximately 18 data points per mean period.

The choice of the integration step size δt is more difficult. For deterministic systems adaptive algorithms are well established that guarantee an upper bound for the deviation from the true trajectory [23]. No corresponding straightforward procedure is available for SDEs. For SDEs the characteristic quantities are the conditional densities $p(\vec{x}(t+\Delta t)|\vec{x}(t))$. To obtain a sampled trajectory that can be regarded as a realization of the SDE, δt has to be chosen such small that the conditional densities become independent of δt for smaller values. Figs. 1 and 2 shows this procedure for the first component of $\vec{x}(t+\Delta t)$ of the stochastic van der Pol oscillator denoted by $x_1(t+\Delta t)$ for the two state vectors $\vec{x}(t) = (-0.0935, -4.284)$ and $\vec{x}(t) = (1.021, -0.9375)$. The conditional densities $p(x_1(t+\Delta t)|\vec{x}_i(t))$ were estimated by triangular kernel estimators [26] based on 5000 integrations of the SDE. The quality of these estimated densities with respect to bias and variance depends on the width of the kernel. By visual inspection, the width of the kernel was chosen equal to 0.02 in Fig. 1 and equal to 0.1 in Fig. 2. In Fig. 1, the estimated conditional density changes drastically between $\delta t = 0.1s$ and $\delta t = 0.01s$. It becomes independent from δt for $\delta t = 0.001s$. In Fig. 2, this already happens for $\delta t = 0.01s$. The reason for the different behavior is that the time evolution in the first case experiences more of the nonlinearity of the system. By investigating numerous analogous simulations, we regarded $\delta t = 0.001s$ as an appropriate integration time step. The two state space vectors $\vec{x}(t)$ used for the above simulation were realizations of the trajectory obtained with $\delta t = 0.001s$. Thus, the procedure for determining δt is an interactive one that has to be selfconsistent.

Please locate Fig. 1 and Fig. 2 around here

Fig. 3 shows a realization of the stochastic van der Pol oscillator with $\mu = 3$, $\delta t = 0.001$ and $\Delta t = 0.5$. The two data points marked by arrows at time 28s, respectively 28.5s were used for the estimation of the conditional densities shown in Figs. 1 and 2.

Please locate Fig. 3 around here

C. A consequence

The above considerations also have consequences for mathematical models used in simulation studies and proposed for analyzing time series in the frame of nonlinear stochastic systems by difference equation of the type :

$$x(t+1) = f(\vec{p}, x(t), x(t-1), \ldots, x(t-m)) + \epsilon(t) \quad . \tag{10}$$

Here the sampling interval is set to unity, $\epsilon(t)$ denotes uncorrelated Gaussian noise, $\vec{p}$ the parameters and m the order of the model, see e.g. [28]. For these models the conditional distribution of $x(t+1)$ given the history is Gaussian. Due to the Gaussianity, least squares optimization leads to a maximum likelihood estimation of the parameters which is asymptotically unbiased and efficient [21,1]. Figs. 1 and 2 show that for difference equations that are thought to be discretized versions of SDEs, the dynamical noise should be non-Gaussian, see the skewed distribution in Fig. 2, and state dependent heteroscadistic. It might even be multimodal.

III. PARAMETER ESTIMATION

In this section we discuss three methods to estimate parameters in SDEs. Due to its superior statistical properties, the most desirable method would be a maximum likelihood estimation [21,1]. We argue that this approach is not feasible. Then we discuss recently suggested quasi-maximum likelihood approaches [4]. The third approach applies the integration scheme, Eq. (2), for the whole sampling interval by using $\Delta t = \delta t$. With respect to the identification of the two time scales this last approach is similar to a procedure to estimate parameters in SDEs proposed in [6,7,25]. In the simulation studies presented in these publications the time series were sampled at the time step of integration.

To simplify the exposition, we consider a scalar dynamics and a single parameter θ in the following.

A. Maximum likelihood estimation

Denoting the stationary distribution of the SDE for a given parameter θ by $\pi(x|\theta)$, the likelihood for a sampled time series of length N reads :

$$\mathcal{L}(x(t_1), x(t_2), \ldots, x(t_N); \theta) = \pi(x(t_1)|\theta) \prod_{i=1}^{N-1} p(x(t_{i+1})|x(t_i), \theta) \quad . \tag{11}$$

Maximizing $\mathcal{L}(.; \theta)$ leads to an estimator $\hat{\theta}$ that is asymptotically unbiased and has least conservative confidence regions. Note that a biased estimator can lead to erroneous interpretations of results, while suboptimal confidence regions 'only' lowers the power to reject simpler models in favor of more complex ones, but does not lead to statistically false positive results.

Usually, the logarithm of the likelihood is considered and the term $\pi(x(t_1)|\theta)$ whose influence vanishes asymptotically is neglected:

$$L(x(t_2), x(t_3), \ldots, x(t_N)|\theta) = \sum_{i=1}^{N-1} \ln\, p(x(t_{i+1})|x(t_i), \theta) \quad . \tag{12}$$

The maximum likelihood estimator $\hat{\theta}_{mle}$ is defined by:

$$\frac{d}{d\,\theta} L(x(t_2), x(t_3), \ldots, x(t_N); \theta) = 0 \quad . \tag{13}$$

To obtain $\hat{\theta}_{mle}$ an iterative optimization strategy has to be applied. Starting from an initial guess for the parameter, the conditional densities $p(x(t_{i+1})|x(t_i), \theta)$ in Eq. (12) have to be estimated and evaluated. This can be achieved by solving the corresponding Fokker-Planck equation or by applying the integration scheme, Eq. (2), with the actual guess of the parameters several times on the intervals $[t_i, t_{i+1}]$ starting from the observed $x(t_i)$ to estimate the conditional density, e.g. by kernel estimation [26]. Then, the log-likelihood can be calculated. Based on this procedure, the parameter is changed until the log-likelihood is extremal by applying some optimization algorithm [23]. The performance of this procedure depends heavily on the quality of the density estimation. Thus, for the procedure based on integrating the SDE thousands of trajectories in each interval $[t_{i_1}, t_i]$ have to be realized.

Note that the densities shown in Figs. 1 and 2 that were based on 5000 realizations are not smooth enough to enable a numerically stable estimation of the parameters. Furthermore, this procedure involves the choice of a parameter determining the bandwidth of the density estimator. This parameter has to be chosen data driven and state dependent, e.g. by a computer intensive cross-validation procedure [26,13,9]. Therefore, the desirable method of maximum likelihood is hardly applicable.

B. Quasi maximum likelihood estimation

Bibby and Sorensen [4] suggested to use a quasi maximum likelihood estimator instead of the infeasible maximum likelihood estimator, Eq. (13). The key idea of this procedure is that Eq. (13) can formally be read as a search for a root :

$$G(\theta) = 0 \quad , \tag{14}$$

defining an estimator $\hat{\theta}$. By virtue of choosing $G(\theta)$, the resulting computationally feasible estimator can be forced to be unbiased by paying the price that the resulting confidence region are no longer optimal. Experience shows that the loss in optimality is often rather small [15]. Eq. (14) is called estimation equation, and the resulting estimator is called quasi maximum likelihood estimator.

Bibby and Sorensen [4] have shown that in the case of SDEs one possible choice for $G(\theta)$ is given by:

$$G(\theta) = \sum_{i=1}^{N-1} g_i(.) \left(x(t_{i+1}) - E(x(t_{i+1})|x(t_i), \theta)\right) \quad , \tag{15}$$

where $g_i(.)$ is a function that is derived from the SDE and $E(x(t_{i+1})|x(t_i), \theta)$ denotes the expected value of $x(t_{i+1})$ conditioned on the observed $x(t_i)$ for a given θ. Different possible procedures for deriving $g_i(.)$ given the SDE are discussed in [4].

In opposite to the conditional density necessary for the maximum likelihood case, the conditional expectation in Eq. (15) can the estimated reliably with a comparable small numbers of integrations in the intervals $[t_i, t_{i+1}]$ starting from the observed $x(t_i)$.

C. "$\Delta t = \delta t$" approach

In the third approach we use the discretization scheme, Eq. (2), on the interval of the whole sampling time step Δt by setting $\Delta t = \delta t$ in Eq. (2). This yields an estimate $\hat{x}((t_{i+1})|x(t_i), \theta)$. The parameters are adjusted until the mean square error

$$\sum_{i=1}^{N-1} (x(t_{i+1}) - \hat{x}((t_{i+1})|x(t_i), \theta))^2 \tag{16}$$

is minimized. This method neglects the existence of the two time scales for integrating and sampling SDEs. Furthermore, it implicitly assumes that the variance of the conditional density is state space independent and Gaussian.

IV. SIMULATION STUDY

We investigate the behavior of the quasi maximum likelihood and the "$\Delta t = \delta t$" approach in a simulation study using the stochastic van der Pol oscillator introduced in Section II B. We integrated the process with $\delta t = 0.001s$. The smaller the sampling time, the smaller the differences between $\vec{h}(.)$ in Eq. (3) and $\vec{g}(.)$ in Eq. (4) will be. Thus, the less severe the effects of ignoring the two time scales by the "$\Delta t = \delta t$" approach will be [10,11,24]. Therefore, we investigate the behavior of the two approaches described above for different sampling times between $\Delta t = 0.005s$ and $\Delta t = 0.5s$. As in the simulation study presented in [25], we assume that the whole state space vector is observed.

For the quasi maximum likelihood approach, Section III B, we chose the term $g_i(.)$ in Eq. (15) to be:

$$g_i(\vec{x}(t_i)) = (1 - x_1^2)\, x_2 \quad , \tag{17}$$

generalizing the line of argument of [4]. The term "$(x(t_{i+1}) - E(x(t_{i+1})|x(t_i), \theta)$" in Eq. (15) is read as $(x_1(t_{i+1}) - E(x_1(t_{i+1})|\vec{x}(t_i), \theta)$. Thus, we predict the observed location $x_1(t_{i+1})$ based on the whole state space vector at the present time step t_i. The expected value of

$x_1(t_{i+1})$ is estimated based on 50 integrations. Finding the zero of Eq. (15) is performed by routines from [23].

For the "$\Delta t = \delta t$" approach, Section III C, we use the information from the whole state space from the history and from the present state. The minimization is perform by routines from [23].

Fig. 4 displays the results of the simulation study for sampling time steps Δt ranging from 0.005 to 0.5. The 2σ confidence intervals were calculated from 50 repetitions. The length of the time series for different sampling times were chosen such that 1000 data points enters the estimation. The quasi maximum likelihood approach yields unbiased results independent from the sampling interval. For realistic sampling intervals the "$\Delta t = \delta t$" approach gives strongly biased results. Only for a sampling time of $\Delta t \leq 0.025$ its estimate is consistent with the true parameter. This would require to sample the process with approximately 360 points per period. Note that classical time series like the sunspot data and the Canadian lynx data are sampled with 11, respectively 10 points per period, see e.g. [8].

Please locate Fig. 4 around here

Fig. 5 shows the results analogous to Fig. 4 for nonlinearity parameter $\mu = 5$. The relative bias of the "$\Delta t = \delta t$" approach for sampling intervals $\delta t = 0.05s, 0.1s$ and $0.5s$ is larger than for $\mu = 3$ due to the higher degree of nonlinearity. Thus, in general, a sufficient sampling for the linearly approximating "$\Delta t = \delta t$" approach to work depends on the degree of nonlinearity.

Please locate Fig. 5 around here

V. DISCUSSION

Modeling time series of open nonequilibrium systems by nonlinear stochastic differential equations (SDE) allows to take into account the effects of the huge number of degrees of freedom that might be active in such systems. A fundamental problem in estimating parameters in SDEs is taught by the theory of integrating nonlinear stochastic differential equations. A central message from this theory is that there are in general two time scales: that of integration and that of sampling. As a consequence, the mean and the distributional characteristics of conditional densities on the time scale of sampling can, in general, not be related back to the parameters of the stochastic differential equation (SDE).

We discussed three approaches for parameter estimation in SDEs based on sampled time series. Unfortunately, the desirable maximum likelihood approach is not feasible in the present context due to its extreme computational burden to estimate the conditional distributions. It might become tractable with the advent of more powerful massive parallel computers. In the presented simulation study, the quasi maximum likelihood approach yielded unbiased estimates for the parameter. Disregarding the existence of the two time scales of integration and of sampling in SDEs and using the discretization scheme of the SDE itself on the time scale of the sampling led to results that are unbiased only for the case of heavy oversampling, but biased for conventional sampling rates.

Thus, methods that require the sampling time interval to be an admissable integration time step, like suggested in [6,7], are applicable to measured time series only for sufficiently sampled data. The same holds for the desirable approach to nonparametrically estimate the complete functional form of the deterministic drift term of a stochastic differential equation based on the discretization of the corresponding Fokker-Planck equation proposed in [25] where formally even the limit "sampling interval $\to 0$" is required.

The approaches discussed in this paper require to observe the complete state space vector which is usually not possible. Our future work will concentrate on the generalization to the case of parameter estimation in nonlinear stochastic differential equations based on scalar observations as it is already possible in linear stochastic and nonlinear deterministic systems [27].

REFERENCES

[1] A. Azzalini. *Statistical inference - based on the likelihood.* Chapman and Hall, London, 1996.

[2] E. Baake, M. Baake, H.G. Bock, and K.M. Briggs. Fitting ordinary differential equations to chaotic data. *Phys. Rev. A*, 45:5524–5529, 1992.

[3] E. Baake and J.P. Schlöder. Modeling the fast flourescence rise of photosynthesis. *Bull. Math. Biol.*, 54:999–1021, 1992.

[4] B.M. Bibby and M. Sørensen. Martingale estimation functions for discretely observed diffusion processes. *Bernoulli*, 1:17–39, 1995.

[5] H.G. Bock. Recent advances in parameter identification for ordinary differential equations. In P. Deuflhard and E. Hairer, editors, *Progress in Scientific Computing*, volume 2, pages 95–121, Boston, 1983. Birkhäuser.

[6] L. Borland and H. Haken. Unbiased determination of forces causing observed processes. *Z. Phys. B - Cond. Matt.*, 88:95–103, 1992.

[7] L. Borland and H. Haken. Unbiased estimate of forces from measured correlation functions, including the case of strong multiplicative noise. *Ann. Phys.*, 1:452–459, 1992.

[8] P.J. Brockwell and R.A. Davis. *Time Series: Theory and Methods.* Springer, New York, 1991.

[9] P. Craven and G. Wahba. Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by the method of generalized cross-validation. *Numerische Mathematik*, 31:317–403, 1979.

[10] D. Dacunha-Castelle and D. Florens-Zmirou. Estimation of the coefficients of a diffusion from discrete observations. *Stochastics*, 19:263–284, 1986.

[11] D. Florens-Zmirou. Approximate discrete-time schemes for statistics of diffusion

processes. *Statistics*, 4:547–557, 1989.

[12] C.W. Gardiner. *Handbook of Stochastic Methods*. Springer, Berlin, 1997.

[13] T. Gasser and M. Rosenblatt. *Smoothing techniques for curve estimation*. Number 757 in Lecture Notes in Mathematics. Springer, New York, 1979.

[14] G. Gouesbet. Reconstruction of the vector fields of continuous dynamical systems from scalar time series. *Phys. Rev. A*, 43:5321–5331, 1991.

[15] C.C. Heyde. *Quasi-likelihood and its Application*. Springer, New York, 1997.

[16] J. Honerkamp. *Stochastic Dynamical Systems*. VCH, New York, 1993.

[17] J. Keizer, R.F. Fox, and J. Wagner. On the amplification of molecular fluctuations for nonstationary systems: hydrodynamical fluctuations for the Lorenz model. *Phys. Lett. A*, 175:17–22, 1993.

[18] P.E. Kloeden and E. Platen. *Numerical solution of stochastic differential equations*, volume 23 of *Applications of Mathematics*. Springer, New York, 1992.

[19] P.E. Kloeden, E. Platen, and H. Schurz. The numerical solution of nonlinear stochastic dynamical systems: a brief introduction. *Int. J. Bif. Chaos*, 1:277–286, 1991.

[20] C. Kurrer and K. Schulten. Effect of noise and pertubations on limit cylce systems. *Physica D*, 50:311–320, 1991.

[21] E.L. Lehmann. *Theory of point estimation, Chapt. 6.2*. Wadsworth Inc., New York, 1991.

[22] H.K. Leung. Stochastic transient of a noisy van der Pol oscillator. *Physica A*, 221:340–347, 1995.

[23] W.H. Press, B.P. Flannery, S.A. Saul, and W.T. Vetterling. *Numerical Recipes*. Cambridge University Press, Cambridge, 1992.

[24] I. Shoji and T. Ozaki. Comparative study of estimation methods for continuous time stochastic processes. *J. Time Series Analysis*, 18:487–506, 1997.

[25] S. Siegert, R. Friedrich, and J. Peinke. Analysis of data sets of stochastic systems. *Phys. Lett. A*, 243:275–280, 1998.

[26] B.W. Silverman. *Density Estimation.* Chapman and Hall, London, 1986.

[27] J. Timmer. Modeling noisy time series: Physiological tremor. *Int. J. Bif. Chaos*, 8:1505–1516, 1998.

[28] H. Tong. *Non-linear Time Series : A Dynamical System Approach*, volume 6 of *Oxford statistical science series.* Clarendon Press, Oxford, 1996.

[29] N.G. van Kampen. *Stochastic Processes in Physics and Chemistry.* Springer, Berlin, 1997.

FIGURE CAPTIONS

Fig. 1 Estimated conditional densities $p(x_1(t + \Delta t)|\vec{x}(t))$ for the stochastic van der Pol oscillator for $\mu = 3$, $\vec{x}(t) = (-0.0935, -4.284)$, integration time steps $\delta t = 0.1s, 0.01s, 0.001s, 0.0001s$ and sampling interval $\Delta t = 0.5s$.

Fig. 2 Estimated conditional densities $p(x_1(t + \Delta t)|\vec{x}(t))$ for the stochastic van der Pol oscillator for $\mu = 3$, $\vec{x}(t) = (1.021, -0.9375)$, integration time steps $\delta t = 0.1s, 0.01s, 0.001s, 0.0001s$ and sampling interval $\Delta t = 0.5s$.

Fig. 3 Realization of the stochastic van der Pol oscillator for $\mu = 3$, integration time step $\delta t = 0.001s$ and sampling interval $\Delta t = 0.5s$. The two data points marked by arrows at time 28s, respectively 28.5s were used for the estimation of the conditional densities shown in Figs. 1 and 2.

Fig. 4 Dependence of the estimated parameter $\hat{\mu}$ of the stochastic van der Pol oscillator on the sampling interval. The integration time step δt was $0.001s$. The 95% confidence intervals were calculated based on 50 repetitions. +: quasi maximum likelihood approach, $*$: "$\Delta t = \delta t$" approach. For sake of clearity the results are slightly scattered around the applied sampling time steps $\Delta t = 0.005s, 0.01s, 0.025s, 0.05s, 0.1s$ and $0.5s$. The true value of the parameter $\mu = 3$ is marked by the solid line.

Fig. 5 Analogous to Fig. 4 for $\mu = 5$.

FIGURES

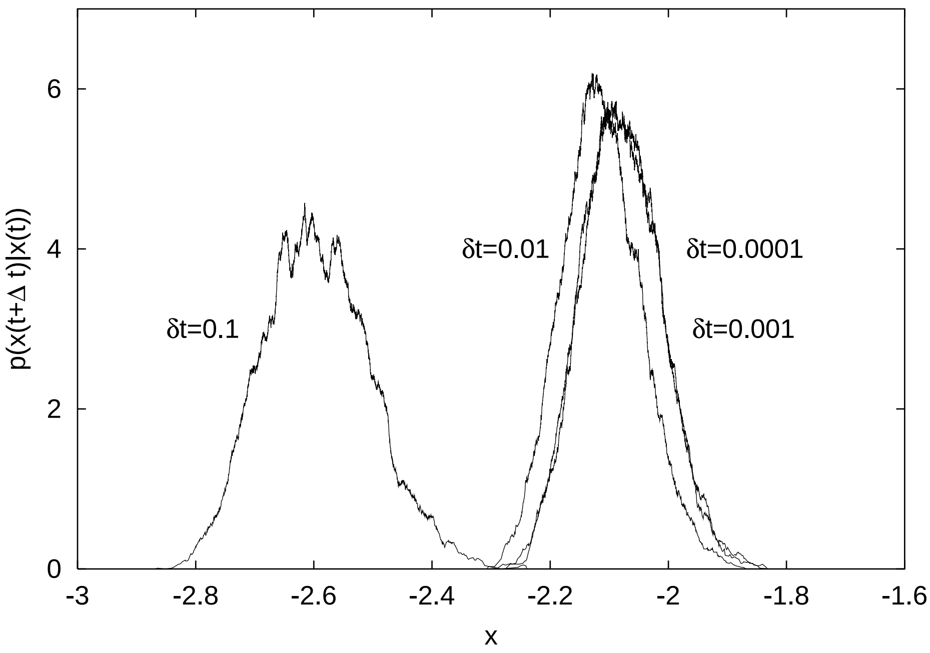

FIG. 1.

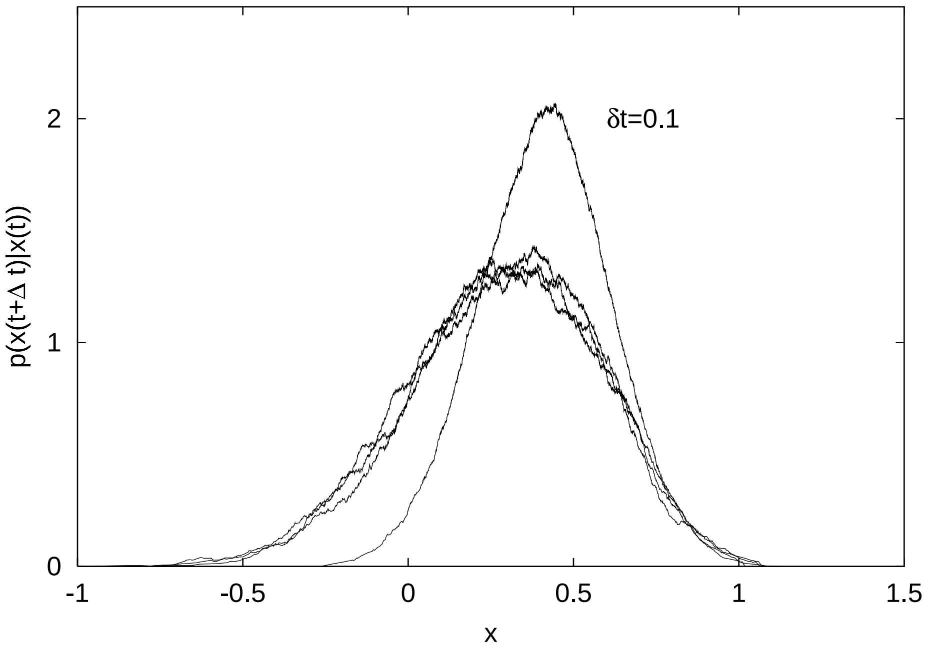

FIG. 2.

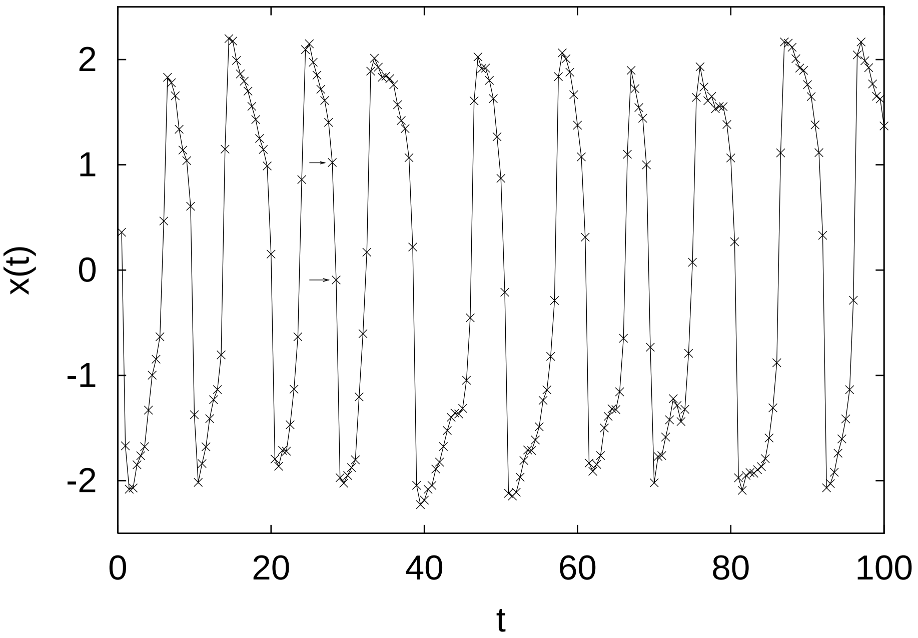

FIG. 3.

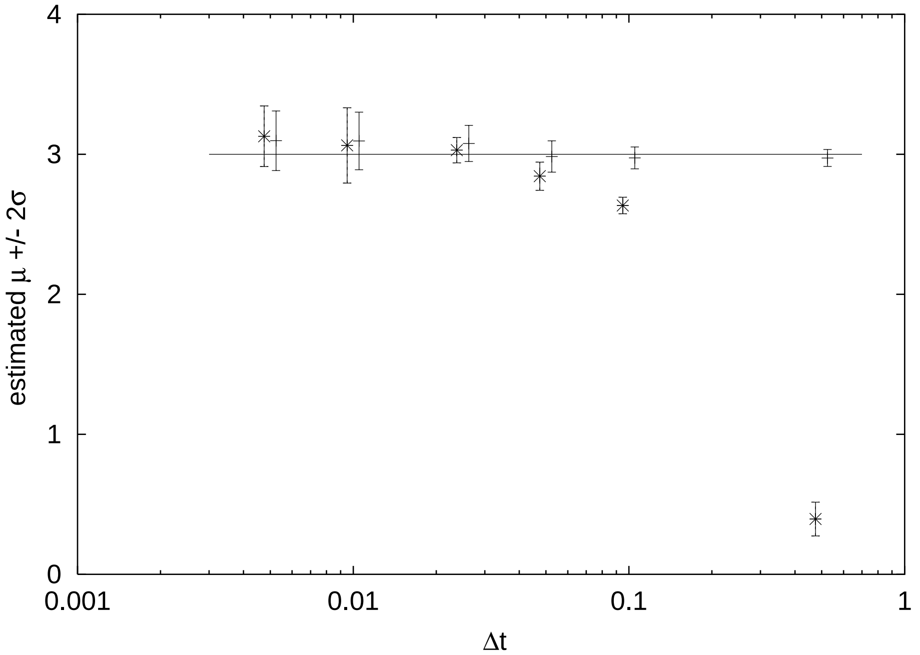

FIG. 4.

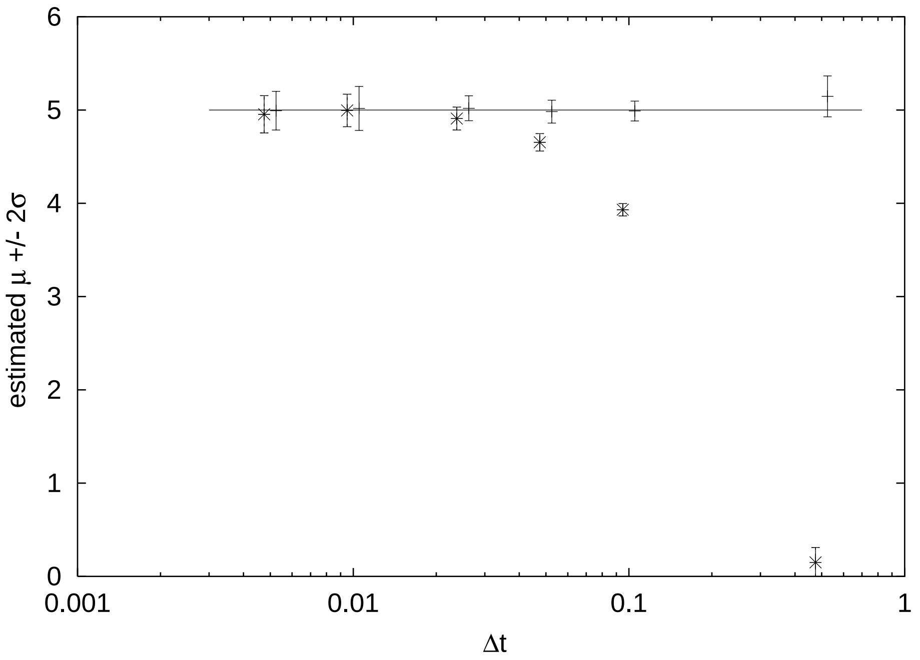

FIG. 5.