ELSEVIER

Physics Letters A 305 (2002) 304–311

PHYSICS LETTERS A

www.elsevier.com/locate/pla

A note on estimating drift and diffusion parameters from timeseries

Philip Sura *, Joseph Barsugli

NOAA-CIRES Climate Diagnostics Center, Boulder, CO 80305-3328, USA

Received 9 September 2002; accepted 22 October 2002

Communicated by J. Flouquet

Abstract

Estimating the deterministic drift and stochastic diffusion parameters from discretely sampled data is fraught with the potential for error. We derive a simple way of estimating the error due to the finite sampling rate in these parameters for a univariate system using a straightforward application of the Itô–Taylor expansion. The error is calculated up to first order in the finite sampling time increment Δt. We then compare the approximate results with the analysis of numerically generated timeseries where the answer is known. Furthermore, a meteorological real world example is discussed.

PACS: 02.50. Fz; 02.50. Ey; 02.50. Ga

1. Introduction

In this Letter we consider a univariate Itô stochastic differential equation (SDE) of the form

$$dx = A(x)\,dt + B(x)\,dW, \tag{1}$$

where $A(x)$ and $B(x)$ are known functions, and W denotes a Wiener process. For sufficiently smooth and bounded $A(x)$ and $B(x)$ the probability density function $p(x,t)$ (PDF) of the Itô SDE (1) is governed by the corresponding Itô–Fokker–Planck equation [1–3], which reads

$$\frac{\partial p(x,t)}{\partial t} = -\frac{\partial}{\partial x} A(x) p(x,t) + \frac{1}{2}\frac{\partial^2}{\partial x^2} B(x)^2 p(x,t). \tag{2}$$

For a detailed discussion of stochastic integration and the differences between Itô and Stratonovich SDEs see, for example, [1,2]. To briefly summarize, the Stratonovich calculus best represents situations where rapidly fluctuating quantities with small but finite correlation times are parameterized as white noise. The Itô stochastic calculus is used when discrete uncorrelated fluctuations are approximated as continuous white noise. That means continuous physical systems are normally described by the Stratonovich calculus, whereas, for example, the financial market is best modeled by the Itô calculus [3]. Nevertheless, in the Itô interpretation the deterministic term $A(x)$ can simply be interpreted as the so-called "effective drift", which is the sum of the deterministic and the noise-induced drift in Stratonovich systems.

Suppose we wish to model an observed, univariate discrete timeseries $x(t_i)$ using the SDE (1). For parametric estimation of $A(x)$ and $B(x)$, that is if one specifies the functional form of $A(x)$ and $B(x)$ in advance,

* Corresponding author.
E-mail address: psura@cdc.noaa.gov (P. Sura).

0375-9601/02/$ – see front matter © 2002 Elsevier Science B.V. All rights reserved.
PII: S0375-9601(02)01474-3

Maximum Likelihood Estimate (MLE) methods are usually preferred [4]. However, we concern ourselves with non-parametric estimates of $A(x)$ and $B(x)$ obtained by binning the data in x. Then deterministic and stochastic parts can be determined directly from data by simply using their definition [5–8]:

$$A(x) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \langle X(t+\Delta t) - x \rangle \Big|_{X(t)=x}, \qquad (3)$$

$$B(x)^2 = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \big\langle (X(t+\Delta t) - x)^2 \big\rangle \Big|_{X(t)=x}, \qquad (4)$$

where $X(t+\Delta t)$ is a solution, that is, a single stochastic realization of the SDE (1), that starts at $X(t) = x$ at time t. $\langle \cdots \rangle$ denotes the averaging operator. At every point x in the state space spanned by the data whose neighborhood is visited often enough by the trajectory, deterministic and stochastic parts of the underlying dynamics can be estimated. These formulae are the embodiment of the property that the deterministic dynamics are proportional to Δt and the stochastic term to $\sqrt{\Delta t}$. Note that the definitions are only correct in the limit $\Delta t \to 0$. In order to verify the results, the estimated functions $A(x)$ and $B(x)^2$ can be inserted into the Fokker–Planck equation (2), and the resulting PDF predicted by (2) can be compared with the PDF obtained directly from the data. In the multivariate case the stochastic component is given by a matrix $\tilde{B}(\vec{x})$, and $\tilde{B}(\vec{x})\tilde{B}^T(\vec{x})$ is estimated from data. In general, it is impossible to find a unique expression for $\tilde{B}(\vec{x})$ in the multivariate case, because it is not guaranteed that $\tilde{B}(\vec{x})$ is invertible. However, in the univariate case $B(x) = \sqrt{B(x)^2}$. The sign of the square root is arbitrary because $B(x)$ is multiplied by Gaussian white noise with zero mean. Thus, in the univariate case the SDE (1) can be used to test the estimates of $A(x)$ and $B(x)$ by simply comparing the properties (e.g., moments, spectra, etc.) of the original time series with the properties of the time series obtained by integrating (1).

In the analysis of observed data, in particular in meteorology and other geophysical applications, one is often given a finite time increment Δt that is a bit too large for comfort; either through historical practice or economic necessity. This timestep may be of the order of 1/4 of the fastest timescale of the deterministic system. In this Letter we derive a simple way of estimating the error in the finite-difference approximations of $A(x)$ and $B(x)^2$ for a univariate system using a straightforward application of the Itô–Taylor expansion. In Section 2 the Itô–Taylor expansion is performed and discussed. In Section 3.1 we then compare the approximate results with the analysis of numerically generated timeseries where the answer is known. Furthermore, a meteorological real world example is discussed in Section 3.2. Finally, Section 4 provides a summary and a discussion.

2. Stochastic Itô–Taylor expansion

The definitions of $A(x)$ and $B(x)^2$ given by (3) and (4) are only correct in the limit $\Delta t \to 0$. For a given time increment Δt the finite-difference approximations $\tilde{A}(x)$ and $\tilde{B}(x)^2$ become

$$\tilde{A}(x) = \frac{1}{\Delta t} \langle X(t+\Delta t) - x \rangle \Big|_{X(t)=x}, \qquad (5)$$

$$\tilde{B}(x)^2 = \frac{1}{\Delta t} \big\langle (X(t+\Delta t) - x)^2 \big\rangle \Big|_{X(t)=x}. \qquad (6)$$

To estimate the error made by using a finite time increment Δt, $X(t+\Delta t)$ can be expanded in a stochastic Itô–Taylor series [4]. Because we want to keep only the terms in the expansion that lead to terms of the order Δt in $\tilde{A}(x)$ and $\tilde{B}(x)^2$, the weak (omitting triple stochastic integrals) Itô–Taylor approximation up to the order Δt^2 is sufficient:

$$\begin{aligned} X(t+\Delta t) = {} & X(t) + AI_{(0)} + BI_{(1)} \\ & + \left(AA' + \frac{1}{2}B^2A''\right) I_{(0,0)} \\ & + \left(AB' + \frac{1}{2}B^2B''\right) I_{(0,1)} \\ & + BA'I_{(1,0)} + BB'I_{(1,1)} \\ & + \text{residual}. \end{aligned} \qquad (7)$$

The Itô integrals $I_{(i,j)}$ are defined as in [4]:

$$I_{(0)} = \int_t^{t+\Delta t} dt', \qquad I_{(1)} = \int_t^{t+\Delta t} dW(t'),$$

$$I_{(0,0)} = \int_t^{t+\Delta t} \int_t^{s} dt'\, ds, \qquad I_{(0,1)} = \int_t^{t+\Delta t} \int_t^{s} dt'\, dW(s),$$

$$I_{(1,0)} = \int_t^{t+\Delta t} \int_t^s dW(t')\, ds,$$

$$I_{(1,1)} = \int_t^{t+\Delta t} \int_t^s dW(t')\, dW(s).$$

Inserting the expansion of $X(t+\Delta t)$ in (5) and (6), and keeping the terms up the order Δt yields the finite-difference estimates $\tilde{A}$ and $\tilde{B}^2$:

$$\tilde{A} = \frac{1}{\Delta t}\langle X(t+\Delta t) - x\rangle\Big|_{X(t)=x} = A + \left(\frac{AA'}{2} + \frac{B^2A''}{4}\right)\Delta t + O\big(\Delta t^2\big), \tag{8}$$

$$\tilde{B}^2 = \frac{1}{\Delta t}\big\langle (X(t+\Delta t) - x)^2\big\rangle\Big|_{X(t)=x} = B^2 + \bigg(A^2 + B^2A' + BAB' + \frac{1}{2}\big(B^2B'^2 + B^3B''\big)\bigg)\Delta t + O\big(\Delta t^2\big). \tag{9}$$

Note that the formulae (8) and (9) can also be derived from the Fokker–Planck equation as in [9]. From (8) and (9) one can calculate the expected error for a given finite time increment if $A(x)$ and $B(x)$ are known. Note that, of course, for $\Delta t \to 0$ the estimates $\tilde{A}(x)$ and $\tilde{B}(x)^2$ converge to $A(x)$ and $B(x)^2$. Other techniques to calculate the errors are proposed by [9,10]. The errors in $\tilde{A}(x)$ and $\tilde{B}(x)^2$ depend on nonlinear combinations of $A(x)$, $B(x)$ and the corresponding derivatives. Unfortunately, this implies that it is very hard to obtain general analytical expressions for the errors under consideration. Nevertheless, it can be seen immediately from (9) that it is problematic to detect the additive noise in an Ornstein–Uhlenbeck process with a finite time step. For example, if $dx = -ax\,dt + b\,dW$, where $a = b = 1$, and $\Delta t = 1/4$, a significant parabolic error emerges: $\tilde{B}^2 = 3/4 + 1/4x^2$. It should be noted that an error in the estimate of the linear term will induce a quadratic error in B^2 as well as a constant offset in B.

Because it is impossible to know $A(x)$ and $B(x)$ in advance, the most practical way to detect the error made by using a finite time step is to change Δt by subsampling the given timeseries and compare the results. Ref. [11] suggests a method based on Richardson extrapolation, whereby (5) and (6) are evaluated at time increments of Δt, $2\Delta t$, etc., and combined so as to cancel out successive terms in the stochastic Taylor series. Another, more accurate way to correct the error might be to solve the coupled second-order differential equations (8) and (9) for $A(x)$ and $B(x)$ for the given numerical estimates $\tilde{A}(x)$ and $\tilde{B}(x)$. Nevertheless, this imposes the problem to accurately specify $A(x)$, $A'(x)$, $B(x)$, and $B'(x)$ for an arbitrary $x = x_0$.

A pedagogical example

Often the following straightforward, but in general *wrong* calculation is made to account for the errors in (5) and (6). Thereby, the stochastic Euler scheme (the weak Itô–Taylor approximation up to the order Δt) $X(t+\Delta t) - x = A(x)\Delta t + B(x)\,dW$ is used to approximate (1), and is then inserted in (5) and (6):

$$\tilde{A} = \frac{1}{\Delta t}\langle X(t+\Delta t) - x\rangle\Big|_{X(t)=x} = \frac{1}{\Delta t}\langle A\Delta t + B\,dW\rangle = A, \tag{10}$$

$$\tilde{B}^2 = \frac{1}{\Delta t}\big\langle (X(t+\Delta t) - x)^2\big\rangle\Big|_{X(t)=x} = \frac{1}{\Delta t}\big\langle (A\Delta t + B\,dW)^2\big\rangle = B^2 + A^2\Delta t. \tag{11}$$

Because of the error term in (11), it could falsely be suggested that the finite-difference estimation of the diffusion term is given by the formula

$$B^2 = \tilde{B}^2 = \frac{1}{\Delta t}\Big\langle \big(X(t+\Delta t) - x - \tilde{A}\Delta t\big)^2\Big\rangle\Big|_{X(t)=x}, \tag{12}$$

in order to numerically obtain the correct diffusion term $B(x)^2$. Nevertheless, in light of the stochastic Taylor expansion performed previously, (12) omits several terms of order Δt. Even in the case of linear A and constant B (Ornstein–Uhlenbeck process) mentioned above, there is one term missing from the estimates of both A and B. The entire calculation is flawed by the fact that for finite time steps Δt the stochastic Euler approximation used to obtain (10) and (11) is in general *not* an accurate approximation of the original SDE (1). The Euler scheme obviously corresponds to the truncated Itô–Taylor series (7) containing only the single time and Wiener integrals $I_{(0)}$ and $I_{(1)}$. For finite time steps Δt the Euler scheme only gives good results when the drift and diffusion coefficients are nearly constant [4].

3. Examples

To qualitatively study the errors made by calculating the finite-difference estimates $\tilde{A}$ and $\tilde{B}^2$ from timeseries, known functions $A(x)$ and $B(x)$ are inserted into the error estimates (8) and (9) to calculate the theoretically expected errors $\tilde{A}(x) - A(x)$ and $\tilde{B}(x) - B(x)$. We then compare the theoretical results with the analysis of numerically generated timeseries. This is done by using the formulae (5) and (6) to calculate $\tilde{A}$ and $\tilde{B}^2$ from the data obtained by integrating the SDE (1) with the prescribed functions $A(x)$ and $B(x)$. The SDE (1) is numerically solved by the stochastic Milstein scheme [4], and is integrated for 250 000 time units Δt, whereby each time unit is divided into 40 time steps. Every 10th time step is saved to obtain an artificial timeseries with the increment $\Delta t = 0.25$. Thus, in the following the finite time step is set to $\Delta t = 0.25$. Finally, a relevant meteorological real world example is discussed.

3.1. Artificial functions

3.1.1. $A = -x$; $B = 1$, $B = |x| + 0.1$, $B = 0.1x^2 + 1$

Firstly, a linear deterministic damping term $A = -x$ is used in combination with three different stochastic terms: $B = 1$, $B = |x| + 0.1$, and $B = 0.1x^2 + 1$. The results are shown in Fig. 1. In general, the theoretical estimates (8) and (9) coincide very well with numerically obtained functions. Only for large values of x the first-order approximations are slightly different from the numerical results. Furthermore, the numerical estimates for large x are more noisy than the points near the origin, because these border points are visited rarely by the trajectory, and, therefore, the numerical estimates for a finite timeseries are more uncertain there than near the origin. From (8) it can be deduced that for a linear $A(x)$, $\tilde{A}(x)$ does not depend on $B(x)$. Thus, $\tilde{A}(x)$ is the same in all of the three examples. It can be seen that a linear damping term is captured relatively well by the finite-difference approximation (8). Nevertheless, it is rather problematic to detect pure additive noise ($B = 1$) using a finite step $\Delta t = 0.25$ in (9) because a significant parabolic error emerges (see Fig. 1(a)). The term $A^2 + B^2 A' = x^2 - 1$ is the only remaining error term in (9). Pure additive noise can only detected with very small time increments Δt. The method is much more successful in detecting a linear noise term $B = |x| + 0.1$ (Fig. 1(b)), as long as the additive part in B is not too large. Then, the leading error terms A^2 and $B^2 A'$ cancel each other. Nevertheless, for a much larger additive component the terms A^2 and $B^2 A'$ do not cancel each other any more, and even BAB' contributes to the error. In Fig. 1(c) it is shown that is even problematic do detect a weak parabolic multiplicative noise term ($B = 0.1x^2 + 1$).

3.1.2. $A = -0.1x^3$; $B = 1$, $B = |x| + 0.1$, $B = 0.1x^2 + 1$

Secondly, a nonlinear deterministic damping term $A = -0.1x^3$ is used in combination with the three different stochastic terms: $B = 1$, $B = |x| + 0.1$, and $B = 0.1x^2 + 1$. The results are shown in Fig. 2. It is important to note that in contrast to the previous examples with a linear deterministic damping term, $\tilde{A}(x)$ now depends on the structure of the deterministic term $A(x)$ and the stochastic term $B(x)$, because $B^2 A'' \neq 0$. Again, the theoretical estimates (8) and (9) coincide very well with the numerically obtained functions (with minor exceptions for large values of x, as already discussed). Fig. 2(a) shows that the deterministic and the constant noise term ($B = 1$) are relatively well captured in the case of the nonlinear damping. This behavior is due to the fact that A and A' are small for not too large values of x. The same holds for the other two examples presented in Figs. 2(b),(c). There, the deterministic and stochastic functions are relatively well captured by the finite-difference estimates, as long as x is not too large. This behavior highlights the fact that the errors in $\tilde{A}(x)$ and $\tilde{B}(x)$ depend on nonlinear combinations of both $A(x)$ and $B(x)$ (and its derivatives). In particular, the quality of the estimate $\tilde{B}$ depends on the structure of the deterministic term.

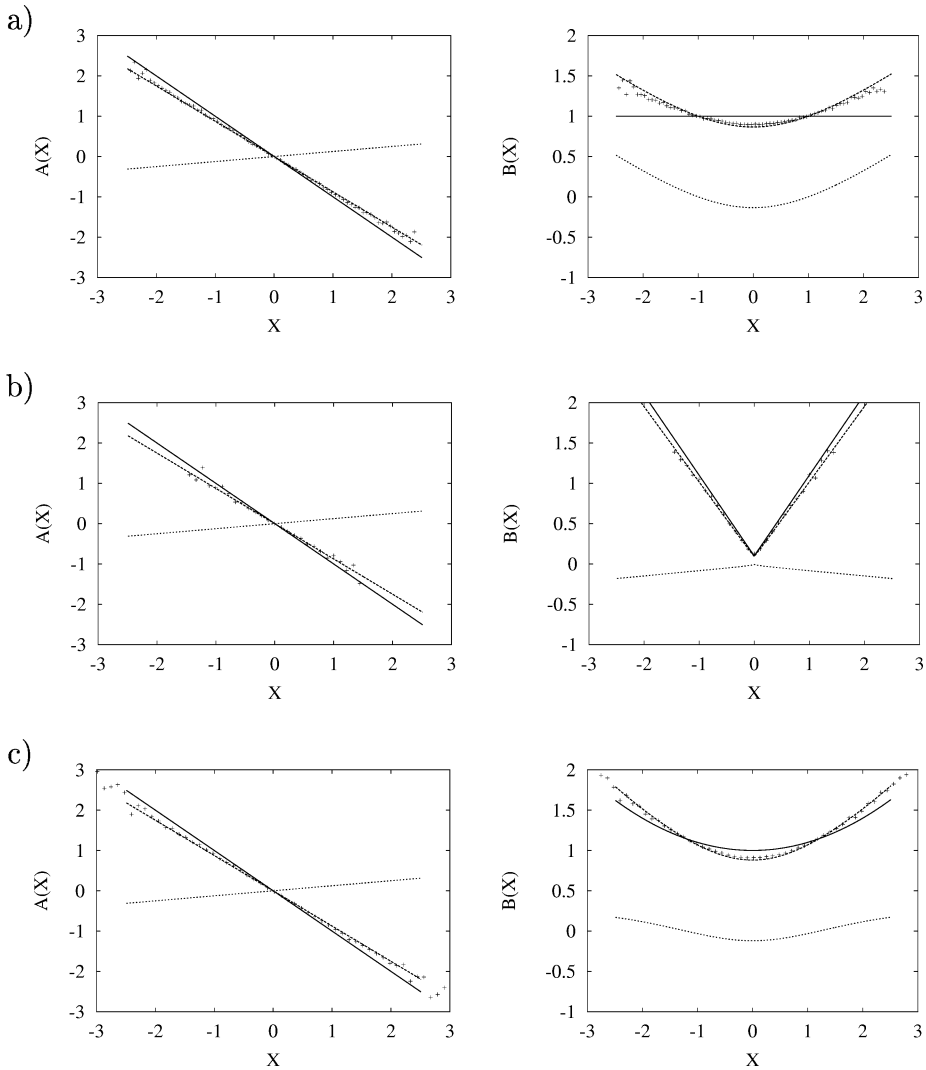

Fig. 1. Error estimates of the finite-difference ($\Delta t = 0.25$) approximations $\tilde{A}(x)$ (left) and $\tilde{B}(x)$ (right) in the case of $A = -x$ and (a) $B = 1$, (b) $B = |x| + 0.1$, (c) $B = 0.1x^2 + 1$. $A(x)$, $B(x)$: solid line; $\tilde{A}(x)$, $\tilde{B}(x)$: dashed line; $\tilde{A}(x) - A(x)$, $\tilde{B}(x) - B(x)$: dotted line. The corresponding numerical estimates are indicated by the '$+$' signs.

3.2. Real world data

The synoptic variability of midlatitude sea surface winds (obtained from 6 hourly scatterometter observations) can be well described by a univariate SDE [12]. As a representative result from [12] the numerically estimated functions $\tilde{A}(x)$ and $\tilde{B}(x)$ for the (normalized) zonally averaged zonal wind at 50 °S are shown in Fig. 3. The dimensional zonally averaged zonal wind speed is $\bar{u} = 6.6$ m s^{-1}. The corresponding zonally averaged standard deviation is $\bar{\sigma}_u = 5.7$ m s^{-1}. $\tilde{A}(x)$ and $\tilde{B}(x)$ are approximated by fourth-order polynomial fits:

$$\tilde{A}(x) = \sum_{i=0}^{4} a_i x^i, \qquad \tilde{B}(x) = \sum_{i=0}^{4} b_i x^i.$$

Near the origin the deterministic part consists of a nearly linear damping term with a damping time scale of about 1.5 days. For higher wind speeds the damping time scale is about 0.5 days. More impor-

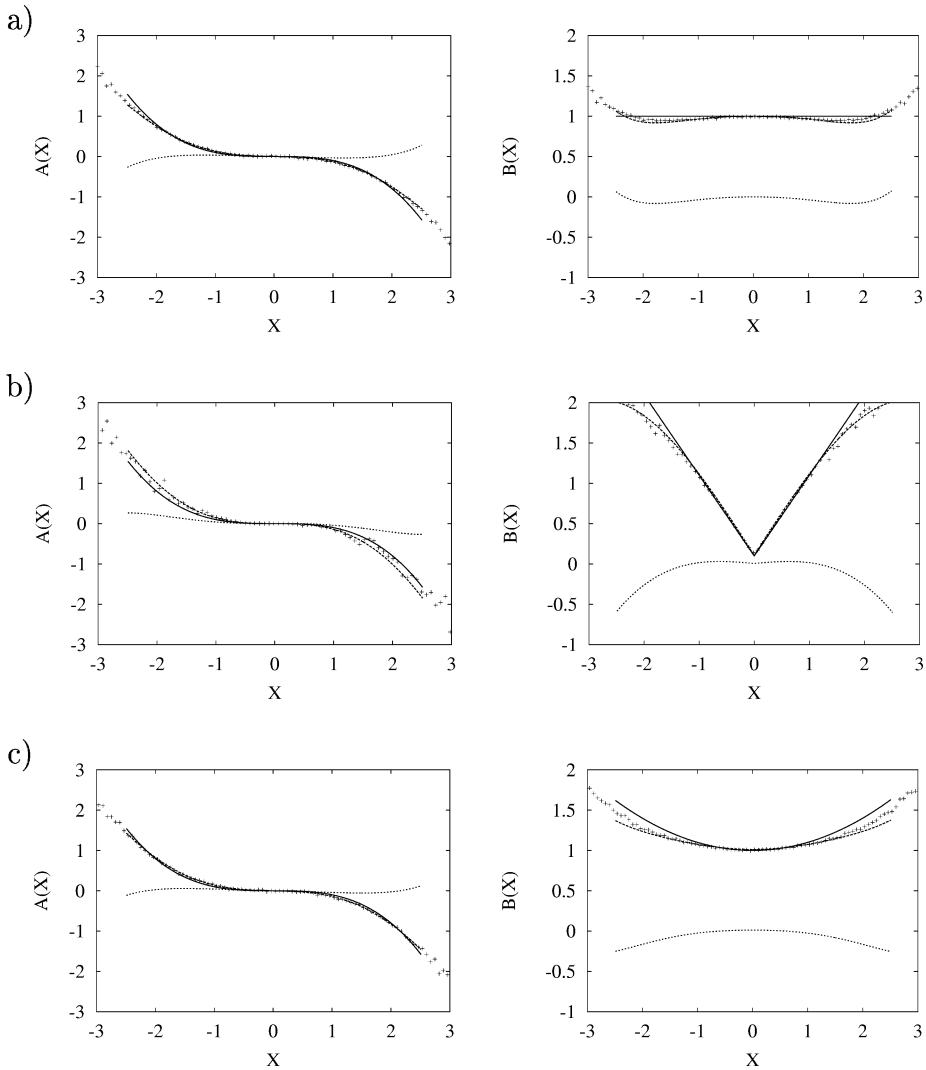

Fig. 2. Error estimates of the finite-difference ($\Delta t = 0.25$) approximations $\tilde{A}(x)$ (left) and $\tilde{B}(x)$ (right) in the case of $A = -0.1x^3$ and (a) $B = 1$, (b) $B = |x| + 0.1$, (c) $B = 0.1x^2 + 1$. $A(x)$, $B(x)$: solid line; $\tilde{A}(x)$, $\tilde{B}(x)$: dashed line; $\tilde{A}(x) - A(x)$, $\tilde{B}(x) - B(x)$: dotted line. The corresponding numerical estimates are indicated by the '$+$' signs.

tantly, a proper description of the winds requires a state-dependent white noise term, that is, multiplicative noise. The need for a parabolic multiplicative noise term to describe the variability of the midlatitude winds can be qualitatively interpreted by the fact that the variability (gustiness) of midlatitude winds increases with increasing wind speed. Moreover, the method used reveals another remarkable characteristic of the underlying timeseries: the variability of westward and eastward winds decreases for increasing wind speeds, until the winds exceed a certain threshold value. This behavior may be understood in terms of an instability mechanism in the presence of friction.

In the light of the discussion in Section 3.1, one might ask if the results from [12], in particular, the structure of the multiplicative noise, are only due to the error terms in (8) and (9). Because it is impossible to know the structure of the noise term $B(x)$ in advance, the most practical way to detect the error made by using a finite time step is to change Δt by subsampling

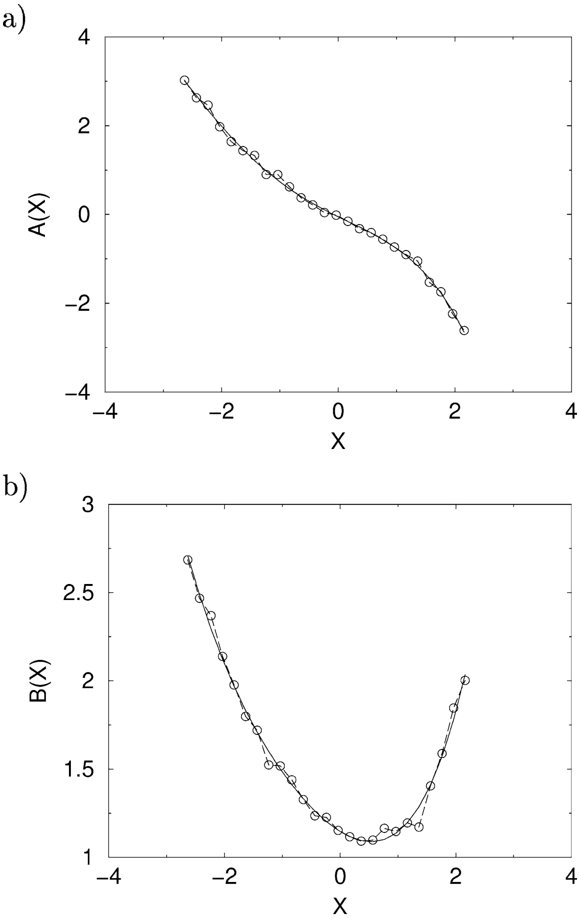

Fig. 3. (a) The estimated deterministic drift $\tilde{A}(x)$ and (b) the estimated noise $\tilde{B}(x)$ for the zonal wind at 50 °S (Southern Ocean). The dashed line with circles shows the actual estimated function, the solid line is a fourth-order polynomial fit.

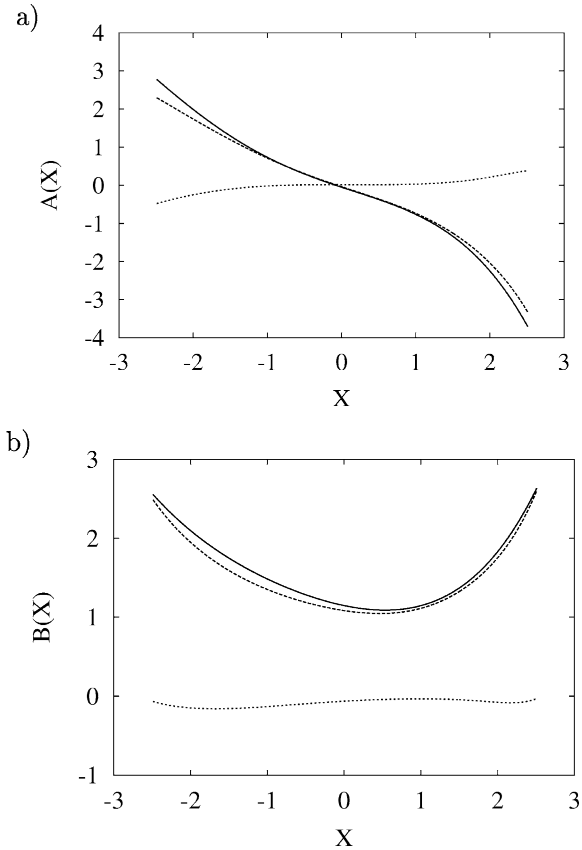

Fig. 4. Consistency check of the finite-difference ($\Delta t = 0.25$) approximations in the case of observed data: (a) $A(x)$ and (b) $B(x)$ (solid lines). The theoretically predicted functions (a) $\tilde{A}(x)$ and (b) $\tilde{B}(x)$ are indicated by the dashed lines. The errors (a) $\tilde{A}(x) - A(x)$ and (b) $\tilde{B}(x) - B(x)$ are indicated by the dotted lines.

the data and compare the results. This has been done, and it appears that the error is neglectable for $\Delta t = 6, 12$, and 18 h for midlatitude winds. The estimates of $B(x)$ begin to diverge for time steps equal to or larger than 24 h. Thus, the multiplicative noise found in the midlatitude wind data is not a spurious result. To test the numerically estimated functions $\tilde{A}$ and $\tilde{B}$ for consistency, we assume that the estimated functions are actually correct. Then, the "correct" estimates are inserted in (8) and (9). If the estimates are consistent with the analytical error estimation, the error terms in (8) and (9) should be small. This has been done with the numerical estimates, and the results are shown in Fig. 4. The error is indeed relatively small. That is, the estimates of $\tilde{A}$ and $\tilde{B}$ are consistent with the error formulae.

4. Summary and conclusions

In this Letter we derived a simple way of calculating the errors induced by a finite sampling rate in the numerically estimated drift and diffusion parameters of a univariate stochastic system. This has been done by a straightforward application of the Itô–Taylor expansion. The derived formulae show that the numerical estimates of these parameters from data is fraught with the potential for error. In particular, it has been shown that it is problematic to detect pure additive noise when the sampling period of the data is large compared to the deterministic timescale. The analytical results indicate that one should carefully test the numerically estimated drift and diffusion parameters. Because it is impossible to know the structure of the correct terms $A(x)$ and $B(x)$ in advance, the most practical way to detect the error made by using a finite time step is to change Δt by subsampling the data and compare the results. That is, the error term proportional to Δt has to be small and neglectable for the used time step.

To conclude, the discussed method is a very useful tool to analyze timeseries, if one has the potential for error in mind and, therefore, carefully checks the results.

Acknowledgements

We would like to thank Cécile Penland, Prashant Sardeshmukh, and Sarah Gille for helpful discussions and support. This work was funded by the Predictability DRI of the Office of Naval Research, Grant N00014-99-1-0021. Part of this work was done while P.S. was at the Scripps Institution of Oceanography, funded through the NASA Ocean Vector Wind Science Team, JPL Contract number 1222984.

References

[1] C.W. Gardiner, Handbook of Stochastic Methods for Physics, Chemistry and the Natural Science, 2nd Edition, Springer-Verlag, 1985.
[2] W. Horsthemke, R. Lefever, Noise-Induced Transitions: Theory and Applications in Physics, Chemistry, and Biology, Springer-Verlag, 1984.
[3] W. Paul, J. Baschnagel, Stochastic Processes: From Physics to Finance, Springer-Verlag, 1999.
[4] P. Kloeden, E. Platen, Numerical Solution of Stochastic Differential Equations, Springer-Verlag, 1992.
[5] S. Siegert, R. Friedrich, J. Peinke, Phys. Lett. A 243 (1998) 275.
[6] R. Friedrich, S. Siegert, J. Peinke, St. Lück, M. Siefert, M. Lindemann, J. Raethjen, G. Deusch, G. Pfister, Phys. Lett. A 271 (2000) 217.
[7] R. Friedrich, J. Peinke, Ch. Renner, Phys. Rev. Lett. 84 (2000) 5224.
[8] J. Gradišek, S. Siegert, R. Friedrich, I. Grabec, Phys. Rev. E 62 (2000) 3146.
[9] R. Friedrich, Ch. Renner, M. Siefert, J. Peinke, Phys. Rev. Lett. 89 (2002) 149401.
[10] M. Ragwitz, H. Kantz, Phys. Rev. Lett. 87 (2001) 254501.
[11] R. Stanton, J. Finance 52 (1997) 1973.
[12] P. Sura, J. Atmos. Sci. 60 (2003), in press.