



ELSEVIER

Physics Letters A ●●● (●●●●) ●●●-●●●

PHYSICS LETTERS A 49

www.elsevier.com/locate/pla 52

# 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  $\Delta t$ . 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.

© 2002 Published by Elsevier Science B.V.

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

## 1. Introduction

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

$$dx = A(x) dt + B(x) dW, \quad (1)$$

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

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

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).

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

$$8 \quad A(x) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \langle X(t + \Delta t) - x \rangle \Big|_{X(t)=x}, \quad (3)$$

$$10 \quad B(x)^2 = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \langle (X(t + \Delta t) - x)^2 \rangle \Big|_{X(t)=x}, \quad (4)$$

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

41 In the analysis of observed data, in particular in  
 42 meteorology and other geophysical applications, one  
 43 is often given a finite time increment  $\Delta t$  that is a  
 44 bit too large for comfort; either through historical  
 45 practice or economic necessity. This timestep may be  
 46 of the order of 1/4 of the fastest timescale of the  
 47 deterministic system. In this Letter we derive a simple  
 48 way of estimating the error in the finite-difference

49 approximations of  $A(x)$  and  $B(x)^2$  for a univariate  
 50 system using a straightforward application of the  
 51 Itô–Taylor expansion. In Section 2 the Itô–Taylor  
 52 expansion is performed and discussed. In Section 3.1  
 53 we then compare the approximate results with the  
 54 analysis of numerically generated timeseries where the  
 55 answer is known. Furthermore, a meteorological real  
 56 world example is discussed in Section 3.2. Finally,  
 57 Section 4 provides a summary and a discussion.

## 2. Stochastic Itô–Taylor expansion

62 The definitions of  $A(x)$  and  $B(x)^2$  given by (3)  
 63 and (4) are only correct in the limit  $\Delta t \rightarrow 0$ . For a  
 64 given time increment  $\Delta t$  the finite-difference approxi-  
 65 mations  $\tilde{A}(x)$  and  $\tilde{B}(x)^2$  become

$$67 \quad \tilde{A}(x) = \frac{1}{\Delta t} \langle X(t + \Delta t) - x \rangle \Big|_{X(t)=x}, \quad (5)$$

$$69 \quad \tilde{B}(x)^2 = \frac{1}{\Delta t} \langle (X(t + \Delta t) - x)^2 \rangle \Big|_{X(t)=x}. \quad (6)$$

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

$$79 \quad X(t + \Delta t) = X(t) + AI_{(0)} + BI_{(1)} \\
 80 \quad + \left( AA' + \frac{1}{2} B^2 A'' \right) I_{(0,0)} \\
 81 \quad + \left( AB' + \frac{1}{2} B^2 B'' \right) I_{(0,1)} \\
 82 \quad + BA' I_{(1,0)} + BB' I_{(1,1)} \\
 83 \quad + \text{residual}. \quad (7)$$

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

$$85 \quad I_{(0)} = \int_t^{t+\Delta t} dt', \quad I_{(1)} = \int_t^{t+\Delta t} dW(t'), \\
 86 \quad I_{(0,0)} = \int_t^{t+\Delta t} \int_t^s dt' ds, \quad I_{(0,1)} = \int_t^{t+\Delta t} \int_t^s dt' dW(s),$$

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

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

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

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

$$\begin{aligned} \tilde{B}^2 &= \frac{1}{\Delta t} \langle (X(t + \Delta t) - x)^2 \rangle \Big|_{X(t)=x} \\ &= B^2 + \left( A^2 + B^2 A' + BAB' \right. \\ &\quad \left. + \frac{1}{2} (B^2 B'^2 + B^3 B'') \right) \Delta t + O(\Delta t^2). \end{aligned} \quad (9)$$

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  $\Delta t \rightarrow 0$  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  $\Delta t$  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  $\Delta t$ ,  $2\Delta t$ , 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  $\Delta t$ )  $X(t + \Delta t) - x = A(x)\Delta t + B(x) dW$  is used to approximate (1), and is then inserted in (5) and (6):

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

$$\begin{aligned} \tilde{B}^2 &= \frac{1}{\Delta t} \langle (X(t + \Delta t) - x)^2 \rangle \Big|_{X(t)=x} \\ &= \frac{1}{\Delta t} \langle (A\Delta t + B dW)^2 \rangle \\ &= B^2 + A^2 \Delta t. \end{aligned} \quad (11)$$

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

$$\begin{aligned} B^2 &= \tilde{B}^2 \\ &= \frac{1}{\Delta t} \langle (X(t + \Delta t) - x - \tilde{A}\Delta t)^2 \rangle \Big|_{X(t)=x}, \end{aligned} \quad (12)$$

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  $\Delta t$ . 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

1 flawed by the fact that for finite time steps  $\Delta t$  the sto- 49  
 2 chastic Euler approximation used to obtain (10) and 50  
 3 (11) is in general *not* an accurate approximation of the 51  
 4 original SDE (1). The Euler scheme obviously corre- 52  
 5 sponds to the truncated Itô–Taylor series (7) contain- 53  
 6 ing only the single time and Wiener integrals  $I_{(0)}$  and 54  
 7  $I_{(1)}$ . For finite time steps  $\Delta t$  the Euler scheme only 55  
 8 gives good results when the drift and diffusion coeffi- 56  
 9 cients are nearly constant [4]. 57  
 10  
 11

### 12 3. Examples 60

13  
 14 To qualitatively study the errors made by calculat- 61  
 15 ing the finite-difference estimates  $\tilde{A}$  and  $\tilde{B}^2$  from 62  
 16 timeseries, known functions  $A(x)$  and  $B(x)$  are in- 63  
 17 serted into the error estimates (8) and (9) to calculat- 64  
 18 e the theoretically expected errors  $\tilde{A}(x) - A(x)$  and 65  
 19  $\tilde{B}(x) - B(x)$ . We then compare the theoretical results 66  
 20 with the analysis of numerically generated timeseries. 67  
 21 This is done by using the formulae (5) and (6) to calculat- 68  
 22 e  $\tilde{A}$  and  $\tilde{B}^2$  from the data obtained by integrat- 69  
 23 ing the SDE (1) with the prescribed functions  $A(x)$  and 70  
 24  $B(x)$ . The SDE (1) is numerically solved by the 71  
 25 stochastic Milstein scheme [4], and is integrated for 72  
 26 250 000 time units  $\Delta t$ , whereby each time unit is di- 73  
 27 vided into 40 time steps. Every 10th time step is saved 74  
 28 to obtain an artificial timeseries with the increment 75  
 29  $\Delta t = 0.25$ . Thus, in the following the finite time step 76  
 30 is set to  $\Delta t = 0.25$ . Finally, a relevant meteorological 77  
 31 real world example is discussed. 78  
 32

#### 33 3.1. Artificial functions 81

##### 34 3.1.1. $A = -x; B = 1, B = |x| + 0.1, B = 0.1x^2 + 1$ 82

35 Firstly, a linear deterministic damping term  $A = -x$  83  
 36 is used in combination with three different stochastic 84  
 37 terms:  $B = 1$ ,  $B = |x| + 0.1$ , and  $B = 0.1x^2 + 1$ . The 85  
 38 results are shown in Fig. 1. In general, the theoretical 86  
 39 estimates (8) and (9) coincide very well with numeri- 87  
 40 cally obtained functions. Only for large values of  $x$  the 88  
 41 first-order approximations are slightly different from 89  
 42 the numerical results. Furthermore, the numerical esti- 90  
 43 mates for large  $x$  are more noisy than the points 91  
 44 near the origin, because these border points are vis- 92  
 45 ited rarely by the trajectory, and, therefore, the numeri- 93  
 46 cal estimates for a finite timeseries are more uncertain 94  
 47 there than near the origin. From (8) it can be deduced 95  
 48

that for a linear  $A(x)$ ,  $\tilde{A}(x)$  does not depend on  $B(x)$ . 49  
 Thus,  $\tilde{A}(x)$  is the same in all of the three examples. It 50  
 can be seen that a linear damping term is captured rela- 51  
 tively well by the finite-difference approximation (8). 52  
 Nevertheless, it is rather problematic to detect pure ad- 53  
 ditive noise ( $B = 1$ ) using a finite step  $\Delta t = 0.25$  in 54  
 (9) because a significant parabolic error emerges (see 55  
 Fig. 1(a)). The term  $A^2 + B^2 A' = x^2 - 1$  is the only 56  
 remaining error term in (9). Pure additive noise can 57  
 only be detected with very small time increments  $\Delta t$ . The 58  
 method is much more successful in detecting a linear 59  
 noise term  $B = |x| + 0.1$  (Fig. 1(b)), as long as the 60  
 additive part in  $B$  is not too large. Then, the leading 61  
 error terms  $A^2$  and  $B^2 A'$  cancel each other. Neverthe- 62  
 less, for a much larger additive component the terms 63  
 $A^2$  and  $B^2 A'$  do not cancel each other any more, and 64  
 even  $BAB'$  contributes to the error. In Fig. 1(c) it is 65  
 shown that it is even problematic to detect a weak par- 66  
 abolic multiplicative noise term ( $B = 0.1x^2 + 1$ ). 67  
 68  
 69

##### 70 3.1.2. $A = -0.1x^3; B = 1, B = |x| + 0.1,$ 71 $B = 0.1x^2 + 1$ 72

Secondly, a nonlinear deterministic damping term 73  
 $A = -0.1x^3$  is used in combination with the three 74  
 different stochastic terms:  $B = 1$ ,  $B = |x| + 0.1$ , and 75  
 $B = 0.1x^2 + 1$ . The results are shown in Fig. 2. It 76  
 is important to note that in contrast to the previous 77  
 examples with a linear deterministic damping term, 78  
 $\tilde{A}(x)$  now depends on the structure of the deterministic 79  
 term  $A(x)$  and the stochastic term  $B(x)$ , because 80  
 $B^2 A'' \neq 0$ . Again, the theoretical estimates (8) and 81  
 (9) coincide very well with the numerically obtained 82  
 functions (with minor exceptions for large values of 83  
 $x$ , as already discussed). Fig. 2(a) shows that the 84  
 deterministic and the constant noise term ( $B = 1$ ) are 85  
 relatively well captured in the case of the nonlinear 86  
 damping. This behavior is due to the fact that  $A$  and 87  
 $A'$  are small for not too large values of  $x$ . The 88  
 same holds for the other two examples presented in 89  
 Figs. 2(b), (c). There, the deterministic and stochastic 90  
 functions are relatively well captured by the finite- 91  
 difference estimates, as long as  $x$  is not too large. This 92  
 behavior highlights the fact that the errors in  $\tilde{A}(x)$  93  
 and  $\tilde{B}(x)$  depend on nonlinear combinations of both 94  
 $A(x)$  and  $B(x)$  (and its derivatives). In particular, the 95  
 quality of the estimate  $\tilde{B}$  depends on the structure of 96  
 the deterministic term.

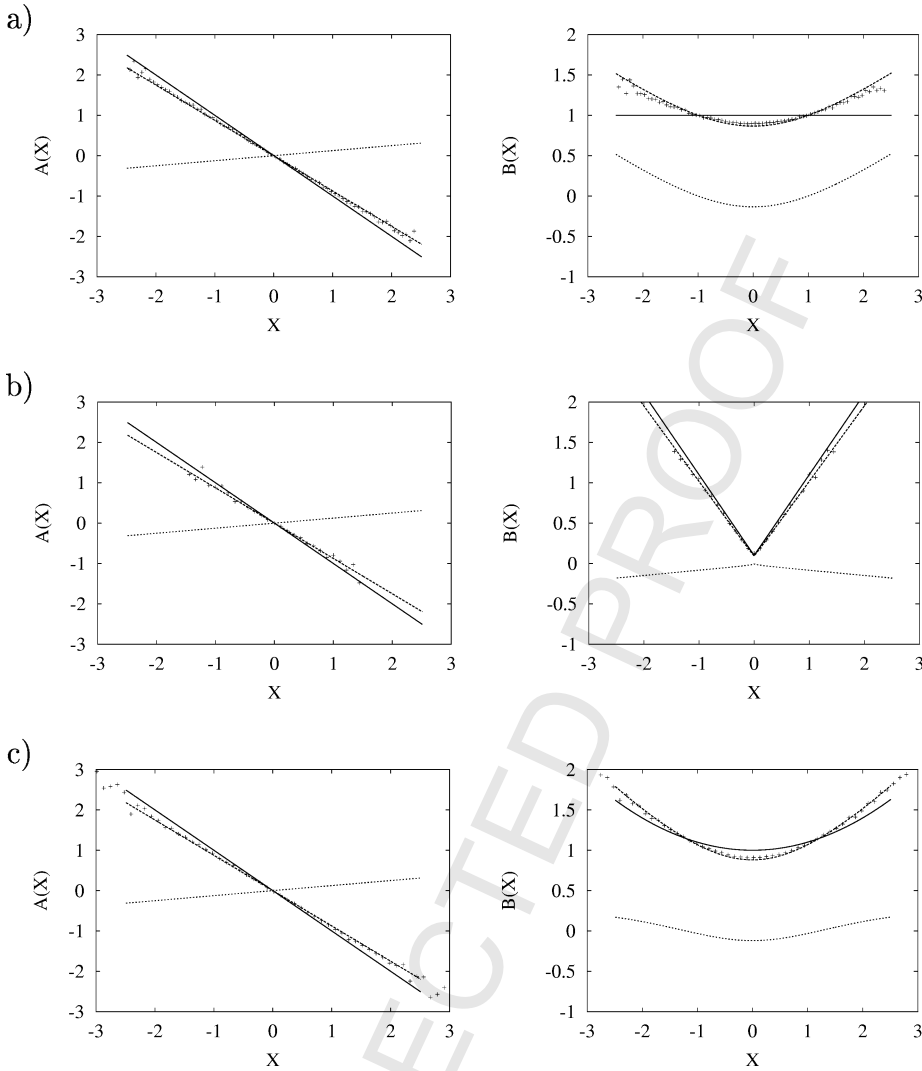


Fig. 1. Error estimates of the finite-difference ( $\Delta t = 0.25$ ) approximations  $\tilde{A}(x)$  (left) and  $\tilde{B}(x)$  (right) in the case of  $A = -x$  and (a)  $B = 1$ , (b)  $B = |x| + 0.1$ , (c)  $B = 0.1x^2 + 1$ .  $A(x)$ ,  $B(x)$ : solid line;  $\tilde{A}(x)$ ,  $\tilde{B}(x)$ : dashed 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 scatterometer 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  $\bar{u} = 6.6 \text{ m s}^{-1}$ . The corre-

sponding zonally averaged standard deviation is  $\bar{\sigma}_u = 5.7 \text{ m s}^{-1}$ .  $\tilde{A}(x)$  and  $\tilde{B}(x)$  are approximated by fourth-order polynomial fits:

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

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-

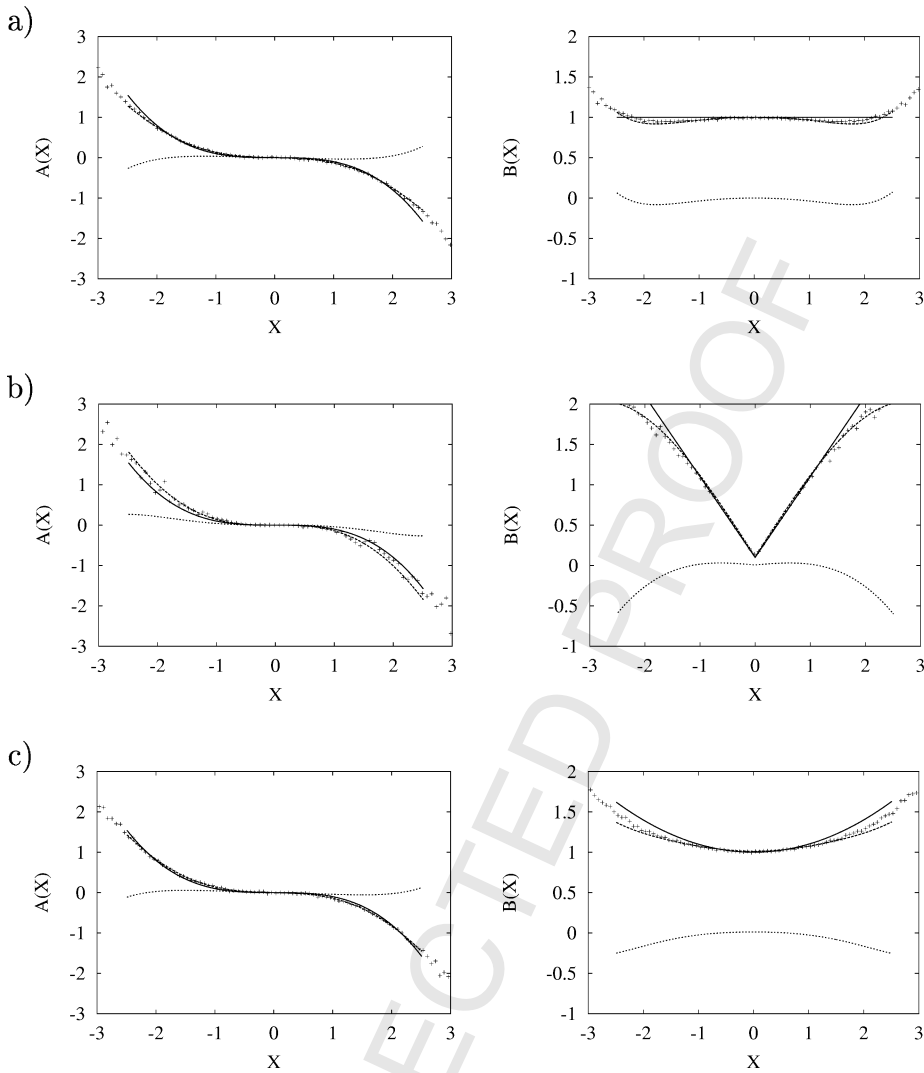


Fig. 2. Error estimates of the finite-difference ( $\Delta t = 0.25$ ) approximations  $\tilde{A}(x)$  (left) and  $\tilde{B}(x)$  (right) in the case of  $A = -0.1x^3$  and (a)  $B = 1$ , (b)  $B = |x| + 0.1$ , (c)  $B = 0.1x^2 + 1$ .  $A(x)$ ,  $B(x)$ : solid line;  $\tilde{A}(x)$ ,  $\tilde{B}(x)$ : 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  $\Delta t$  by subsampling

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48

49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96

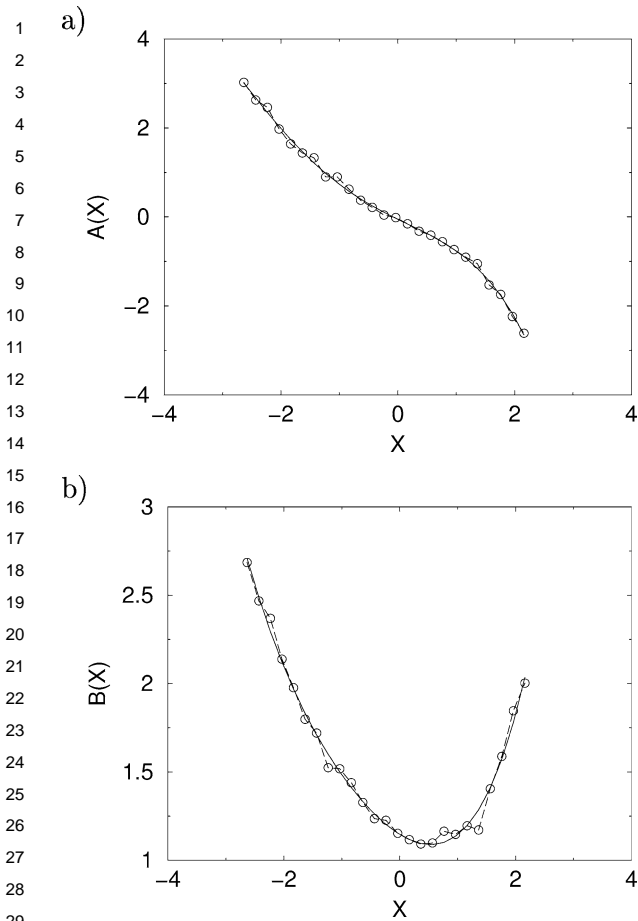


Fig. 3. (a) The estimated deterministic drift  $\tilde{A}(x)$  and (b) the estimated noise  $\tilde{B}(x)$  for the zonal wind at 50° S (Southern Ocean). The dashed line with circles shows the actual estimated function, the solid line is a fourth-order polynomial fit.

the date 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,

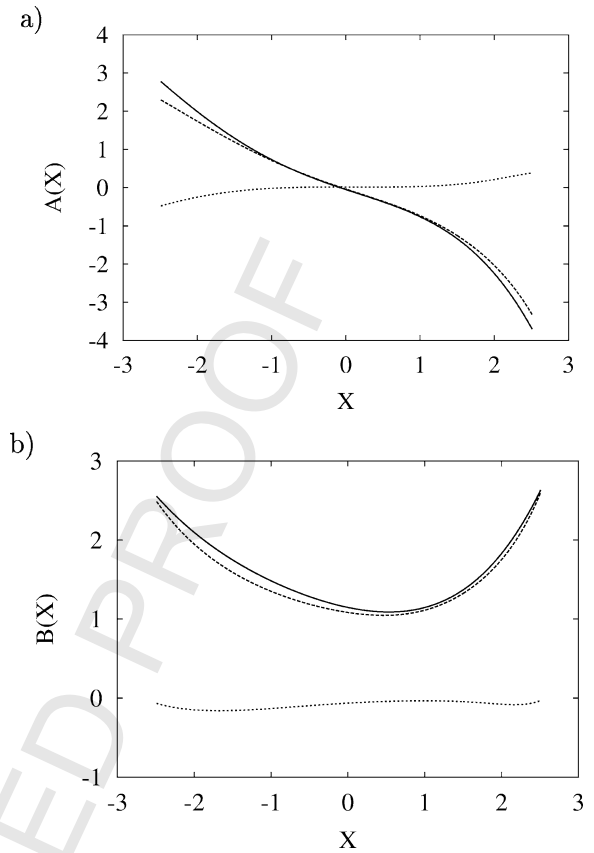


Fig. 4. Consistency check of the finite-difference ( $\Delta t = 0.25$ ) approximations in the case of observed data: (a)  $A(x)$  and (b)  $B(x)$  (solid lines). The theoretically predicted functions (a)  $\tilde{A}(x)$  and (b)  $\tilde{B}(x)$  are indicated by the dashed lines. The errors (a)  $\tilde{A}(x)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

the 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

1 noise when the sampling period of the data is large  
2 compared to the deterministic timescale. The analyt-  
3 ical results indicate that one should carefully test the  
4 numerically estimated drift and diffusion parameters.  
5 Because it is impossible to know the structure of the  
6 correct terms  $A(x)$  and  $B(x)$  in advance, the most  
7 practical way to detect the error made by using a fi-  
8 nite time step is to change  $\Delta t$  by subsampling the data  
9 and compare the results. That is, the error term pro-  
10 portional to  $\Delta t$  has to be small and neglectable for the  
11 used time step.

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

#### 19 Acknowledgements

22 We would like to thank Cécile Penland, Prashant  
23 Sardeshmukh, and Sarah Gille for helpful discus-  
24 sions and support. This work was funded by the Pre-  
25 dictability DRI of the Office of Naval Research, Grant  
26 N00014-99-1-0021. Part of this work was done while  
27 P.S. was at the Scripps Institution of Oceanography,

49 funded through the NASA Ocean Vector Wind Science  
50 Team, JPL Contract number 1222984.

#### 53 References

- 54
- [1] C.W. Gardiner, *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Science*, 2nd Edition, Springer-Verlag, 1985. 55
- [2] W. Horsthemke, R. Lefever, *Noise-Induced Transitions: Theory and Applications in Physics, Chemistry, and Biology*, Springer-Verlag, 1984. 56
- [3] W. Paul, J. Baschnagel, *Stochastic Processes: From Physics to Finance*, Springer-Verlag, 1999. 57
- [4] P. Kloeden, E. Platen, *Numerical Solution of Stochastic Differential Equations*, Springer-Verlag, 1992. 58
- [5] S. Siegert, R. Friedrich, J. Peinke, *Phys. Lett. A* 243 (1998) 275. 59
- [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. 60
- [7] R. Friedrich, J. Peinke, Ch. Renner, *Phys. Rev. Lett.* 84 (2000) 5224. 61
- [8] J. Gradišek, S. Siegert, R. Friedrich, I. Grabec, *Phys. Rev. E* 62 (2000) 3146. 62
- [9] R. Friedrich, Ch. Renner, M. Siefert, J. Peinke, *Phys. Rev. Lett.* 89 (2002) 149401. 63
- [10] M. Ragwitz, H. Kantz, *Phys. Rev. Lett.* 87 (2001) 254501. 64
- [11] R. Stanton, *J. Finance* 52 (1997) 1973. 65
- [12] P. Sura, *J. Atmos. Sci.* 60 (2003), in press. 66
- 67
- 68
- 69
- 70
- 71
- 72
- 73
- 74
- 75
- 76
- 77
- 78
- 79
- 80
- 81
- 82
- 83
- 84
- 85
- 86
- 87
- 88
- 89
- 90
- 91
- 92
- 93
- 94
- 95
- 96