Some Issues in Stochastic Weather/Climate Modeling or How do I use Stochastic Differential Equations to Model Something Real.

Cécile Penland

NOAA/Earth Science Research Laboratory/Physical Sciences Division
Boulder, CO

1. Introduction

Very often, a measured time series representing a geophysical system bears the signature of multiple forces operating at different timescales on that process and interacting with the system's own internal dynamics. Sometimes, the signature is blurred, as when the measuring process cannot resolve every consequence of every force acting on the measured system. If our goal is to find differential equations describing the system, i.e., a "model," consistent with physical reality, we should represent those unresolved forces in a way that accounts for their effect on the resolved scales of the system, while also estimating the uncertainty the unresolved scales render unavoidable. We emphasize that this estimation of uncertainty is part of the model.

Alternatively, we may already know the internal dynamics of a system. Say we need to know how that system is forced by a phenomenon for which we have representative time series, but that time series is unable to resolve the dynamics of the force on every scale. This is an easier problem than the one with which we introduced this article, but is still nontrivial. Further, since this serves as a useful starting point for considering the more difficult problem, it is where we start.

Let's say we wish to model a process x as

$$\frac{dx}{dt} = G(x,t) + F(t), \tag{1}$$

where $G(x)$ is a known function and where $F(t)$ is a time series as shown in Fig. *1*.

Now, it may be that the measuring device that has estimated $F(t)$ has smoothed the time series so that it looks very erratic, and a great deal of structure has been averaged out. Consider the interval between $t = 600$ and $t = 800$ in Fig. 1. The curve is reproduced as the black line in Fig. 2; the red line is what the measuring device would have seen had it been able to reproduce the forcing at much higher resolution.

Although the red line appears to have some deterministic structure, that structure has been lost in the black line because of the smoothing (a 200-point smoother, in this case), and it may indeed be possible to approximate the black line essentially as white noise on the time scales of interest. If the 200-point

smoother is appropriate to Eq. (1), it means that the "timescales of interest", i.e., those described by $G(x,t)$, are essentially 200 times as long as those of the unsmoothed $F(t)$.

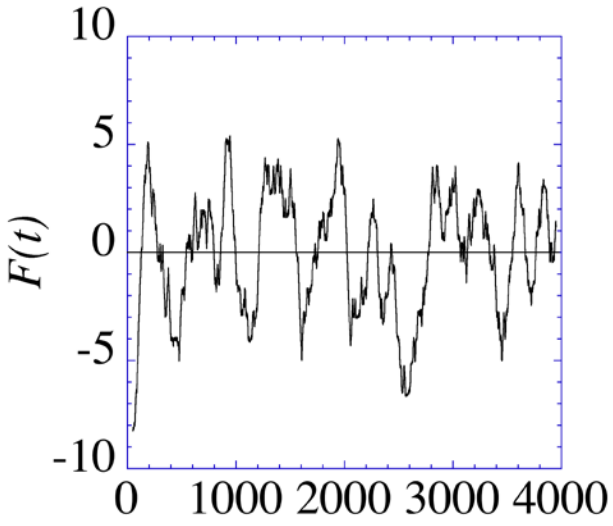

Figure 1: Time series representing the forcing of x.

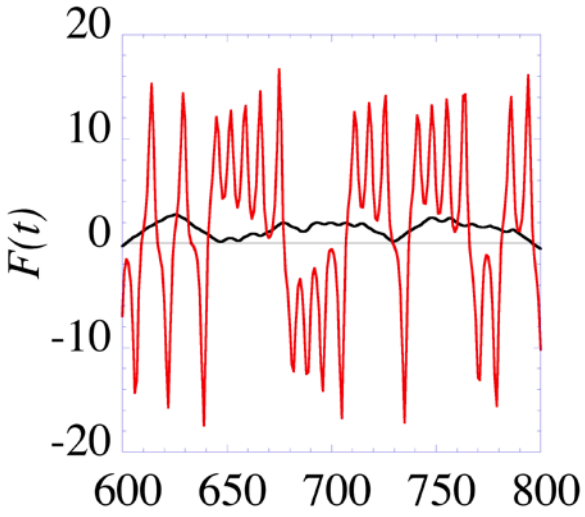

Figure 2: Black line: reproduction of interval shown in Fig. 1.
Red line: unsmoothed values of F(t).

It is the purpose of this article to discuss when this approximation is possible, to give references presenting more rigorous analysis than what can be given here, and to present practical advice for implementing the theory in numerical climate and weather models. It is obviously impossible to give an exhaustive treatment in this short communication, and no attempt to do so is made here. This article is not meant to be self-contained, but should rather be considered a guide to further information. Much more extensive explanations can be found in the references provided and in further citations therein.

2. Brownian motion, etc.

The following discussion follows Papanicolaou and Kohler (1974), although the theorem in that paper is more general than what is discussed here. Let's go back to Eq. (1), but this time we'll let $F(t)$ also depend on x. Retaining the properties that F and G have different timescales, we redefine the symbols F and G introducing a parameter ε as follows:

$$\frac{dx}{dt} = \varepsilon^2 G(x,t) + \varepsilon F(x,t). \tag{2}$$

In Eq. (2), ε is a smallness parameter. However, it is *not* an importance parameter. Although $\varepsilon^2 G$ is smaller than εF, we only consider systems where the nature of F is erratic enough to let the long timescale evolution of x depend primarily on G.

We choose a scaling $\Delta s = \varepsilon^2 \Delta t$, and rewrite Eq. (2) as

$$\frac{dx}{ds} = G(x, s/\varepsilon^2) + \frac{1}{\varepsilon} F(x, s/\varepsilon^2). \tag{3}$$

For illustrative purposes, we further restrict the classes of functions F to have the following form:

$$F(x, s/\varepsilon^2) = \sum_k F_k(x,s)\eta_k(s/\varepsilon^2). \tag{4}$$

That is, the erratic nature of F enters through a series of variables η_k that depend only on time, although they may be multiplied by smooth functions $F_k(x,s)$ that depend on both time and x. We now integrate the lagged covariance function of η_k over lag, and define a matrix $\boldsymbol{\phi}$ as follows:

$$\left(\boldsymbol{\phi}\boldsymbol{\phi}^{\mathrm{T}}\right)_{km} = \int_{-\infty}^{\infty} <\eta_k(t)\eta_m(t'+t)> dt'. \tag{5}$$

Note that this definition of $\boldsymbol{\phi}$ is not unique, but is defined only up to an arbitrary orthogonal matrix.

The theorem of Papanicolaou and Kohler (1974; PK74 hereafter) says that under specific, easily satisfied conditions (such as the existence of Eq. (5); for more detail, please see PK74), Eq. (3) converges weakly in the limit of $t \to \infty$ and $\varepsilon \to 0$ to a stochastic differential equation (in the scaled time coordinates) of the form

$$dx = G(x,s)ds + \sum_k F_k(x,s)\phi_{k\alpha} \circ dW_\alpha \tag{6}$$

In Eq. (6), W_α is a Wiener process, independent of other Wiener processes $\{W_\beta\}$, to be integrated in the sense of Stratonovich.

Let's clarify this last sentence. What is a Wiener process, and what does it mean to be integrated in the sense of Stratonovich? Well, a Wiener process, or Brownian motion, is a Gaussian random variable with the following properties:

- $< W(t) >= 0$ (7*a*)
- $< W(t)W(t') >= \min(t,t')$ (7*b*)
- $< dW(t)dW(t') >= dt\delta(t-t')$ (7*c*)

Here and below, angle brackets denote ensemble averages. The Wiener process is self-similar as illus trated in Fig. 3; a blow-up of the process between $t = 0$ and $t = 1$ has the same character as the process on a scale an order of magnitude larger or, indeed, at any other scale.

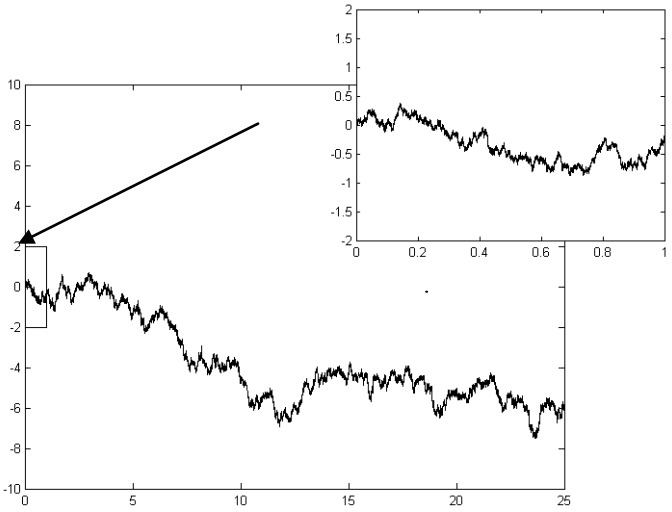

Figure 3: Legitimate simulation of a Wiener process, showing self-similarity.

Now that we've defined a Wiener process, what does it mean to be integrated in the sense of Stratonovich? Well, a Wiener process is not smooth. It is continuous, meaning that it exists at any pair of points in time, no matter how close those points are, but it is not differentiable. If we define an integral over a Wiener process as a Riemann sum in the usual way, we find that this integral is not unique. That is, say we want to evaluate the integral

$$I = \int_{t_1}^{t_2} W dW \,. \tag{8}$$

Dividing the interval $T = t_2\text{-}t_1$ into N intervals, each of width Δ, we define I_N as

$$I_N = \sum_{i=1}^{N} W(\tau_i)\left(W(t_i) - W(t_{i-1})\right), \tag{9}$$

where $T=N\Delta$ and where τ_i lies in the interval $[t_{i-1}, t_i)$. As usual, the integral I is achieved in the limit as $N\to\infty$, $\Delta\to 0$. If $W(t)$ behaved like a run-of-the mill, deterministic function, the integral I would not depend on where in the interval $[t_{i-1}, t_i)$ one chooses τ_i. For a Wiener process, though, it *does* depend. Let us choose $\tau_i = t_{i-1} + \alpha\Delta$ for all i, where $0 \le \alpha < 1$. Then, the solution to Eq.(8) is

$$I = \frac{1}{2}\left(W^2(t_2) - W^2(t_1)\right) + \left(\alpha - \frac{1}{2}\right)T \,. \tag{10}$$

If only one of these integrals (i.e., if only one value of α) were found in nature, then the non-uniqueness of the sum Eq. (9) would be only a mathematical curiosity. However, at least two are found experimentally, defining different sets of calculus rules. One of them, Stratonovich calculus, corresponds to the rules of calculus consistent with the usual Riemann rules of calculus. Stratonovich calculus is the calculus for which τ_i is placed in the center of each interval, and α = 1/2. Integrals obeying this calculus are indicated by a small symbol called "Ito's circle," and Eq. (8) is solved as follows:

$$I = \int_{t_1}^{t_2} W \circ dW = 1/2\left(W^2(t_2) - W^2(t_1)\right). \tag{11}$$

This is the integral appropriate to systems that are continuous but only approximately white. That is, as specified by PK74, this is the calculus one uses when a force has a correlation time that is not strictly zero but is so much smaller than the internal dynamics of the system to which it is applied that we may treat it as white noise.

In contrast is the integral applied to systems that occur at intrinsically discrete timesteps and are uncorrelated from one time to the next, but where the timesteps are close enough that it is essentially continuous to an observer measuring that system. In that case, the appropriate value of α is zero and the integral, called an "Ito" integral, are denoted without Ito's circle:

$$I = \int_{t_1}^{t_2} W dW = 1/2\left(W^2(t_2) - W^2(t_1)\right) - T/2. \tag{12}$$

The only issue from PK74 we have yet to explain is what was meant by a "weak" convergence. What this means is that all of the *moments* of the stochastic differential equations (i.e., Eq. 6) are good approximations to those of our original dynamical equation (i.e., Eq. 2), but we do not claim that the sample paths of Eq, 2 and Eq. 6 are necessarily the same. Nevertheless, since we have *all* of the moments, whether or not they are conditioned on some measured value, it is often useful for the scientist to treat Eq. (6) as the appropriate dynamical equation of a system even though, strictly speaking, one cannot know that for sure. That is why some researchers (e.g., Majda et al. 1999) start off by assuming that the Wiener process has a place in the dynamical equation. They are then strictly justified in considering sample paths of the system, but at the cost of assuming how the Wiener process enters the equation in the first place.

3. Stochastic integration schemes

The well-known property of computers that they do what we tell them to do rather than what we want them to do requires that we be very sure about what we tell them. Although we have identified two calculi found in nature, a computer will quite happily generate any one of an infinite number of self-consistent set of calculus rules if we unwittingly tell it to do so. Thus, it is imperative to understand the development of stochastic integration schemes in order to choose the appropriate scheme for the problem at hand. For an exhaustive treatment of this subject, see Kloden and Platen (1992).

To set notation, let us first review the deterministic Taylor expansion and chain rule. First, the chain rule:

$$\frac{dX}{dt} = f(X) \tag{13a}$$

$$\frac{dF(X)}{dt} = \frac{dF}{dX}\frac{dX}{dt} = f(X)\frac{dF}{dX}. \tag{13b}$$

Another way of writing this is as follows:

$$dF(X) = LF(X)dt, \tag{14a}$$

where the operator L is defined as below:

$$L \equiv f(X)\frac{d}{dX}. \tag{14b}$$

Through an iterative procedure, we find the classic Taylor expansion,

$$F\left(X\left(T\right)\right) = F\left(X\left(0\right)\right) + L\,F\left(X\left(0\right)\right)\,T + 1/2\,L^2\,F\left(X\left(0\right)\right)T^2 + \ldots \tag{15}$$

We can similarly derive the stochastic Taylor expansion. Consider the stochastic differential equation

$$dX = f(X,t)dt + g(X,t)(\circ)dW. \tag{16}$$

In Eq. (16), we have placed Ito's circle in parenthesis to indicate that the equation may be integrated either in the sense of Ito or of Stratonovich. If it is to be integrated in the sense of Stratonovich, we define the operator

$$L^0 = \frac{\partial}{\partial t} + f(X,t)\frac{\partial}{\partial X}, \tag{17a}$$

and if it is to be integrated in the sense of Ito, we define

$$L^0 = \frac{\partial}{\partial t} + f(X,t)\frac{\partial}{\partial X} + g^2(X,t)\frac{1}{2}\frac{\partial^2}{\partial X^2}. \tag{17b}$$

For either case, we define

$$L^1 = g(X,t)\frac{\partial}{\partial X}. \tag{18}$$

As in the deterministic case, we derive the stochastic Taylor equation by iterating. Considering a function $F(X)$ and using Eq. (16),

$$F(X(T),T) = F(X(0),0) + \int_0^T L^1F(X(t),t)(\circ)dW + \int_0^T L^0F(X(t),t)dt. \tag{19}$$

We develop integration schemes by identifying $F(X(t)) = X(t)$ and choosing T equal to the timestep Δ. The rules of order depend on whether the Wiener process is to be integrated in the sense of Ito or Stratonovich. For Ito systems, count 1 for every integral over time and 1/2 for every integral over dW, unless a term has only time (i.e., no dW) in it. In that case, subtract 1/2 from what it otherwise would be. Note the counter-intuitive notation that L^1 is of lower order than L^0.

For Stratonovich calculus, the rules are the same, but only integer orders are valid.

If g does not depend on X, then Ito and Stratonovich calculus are equivalent. Also, if g is independent of X, the Euler scheme (see below) is of order *1* rather than order 1/2.

We are now ready to consider some schemes for numerically integrating Eq. (16). As stated in the Introduction, we do not attempt to give an exhaustive treatment, or even an overview of all the stochastic schemes the reader may wish to implement. Rather, we choose a couple of schemes for illustrative purposes, drawing attention to some of the possible pitfalls, and refer to the literature when needed.

In what follows, we denote a Gaussian random number with mean zero and unit variance as $\mathscr{R}$. An increment of a Wiener process over a timestep Δ is $\Delta W = W(t+\Delta)-W(t)$. From Eq. (7*b*), ΔW is a Gaussian random variable with variance equal to Δ. Thus, in implementing stochastic integration schemes, we estimate ΔW as $\mathscr{R}\sqrt{\Delta}$.

3.1. Stochastic Euler scheme and explicit Mil'steyn Scheme

This is the lowest order stochastic integration scheme, generally of order 1/2, and converges to Ito calculus (Rümelin 1982; Kloden and Platen 1992):

$$X(t_{n+1}) = X(t_n) + f(X(t_n), t_n)\,\Delta + g(X(t_n), t_n)\,\Delta W. \tag{20}$$

The Mil'steyn scheme, which is of O(1), is obtained by adding a term to the stochastic Euler scheme:

$$\begin{aligned} X\left(t_{n+1}\right) &= X\left(t_n\right) + f\left(X\left(t_n\right), t_n\right)\Delta + g\left(X\left(t_n\right), t_n\right)\Delta W \\ &\quad + g\left(X, t_n\right)\frac{\partial g\left(X, t_n\right)}{\partial X} I_{(1,1)} \end{aligned} \tag{21}$$

Eq. (21) converges to Stratonovich calculus when the multiple stochastic integral $I_{(1,1)}$ is estimated as $(\Delta W)^2/2$, and to Ito calculus when $I_{(1,1)}$ is estimated as $[(\Delta W)^2-\Delta]/2$. Note that the last term in Eq. (21) is zero if $g(X, t)$ does not depend on $X(t)$, which is why the stochastic Euler scheme is of order unity for g independent of X. If the partial derivative of $g(X, t)$ cannot be evaluated analytically, one may either use a discrete version of Eq. (21) (Kloden and Platen 1992), or an alternative explicit scheme derived by Ewald and Témam (2005).

Generally speaking timesteps in stochastic integration are much smaller than those used in deterministic numerical integration. Kloden and Platen (1992) recommend a timestep smaller by a factor of 2^9 (i.e., about 500) than the smallest timescale in the system to be simulated; we have found we can get away with timesteps about 5 or 6 times larger than their recommendation. We suggest that the user experiment on systems for which the results are known to find an optimal timestep.

3.2. Cautionary tales

3.2.1. Does using the square root of the time step really matter?

We have discussed how the numerical generation of stochastic differential equations involves a Wiener process, which has a *variance* proportional to time and is therefore simulated by a term involving the square root of the timestep. What happens if we update the stochastic term using the

timestep rather than the square root of the timestep? Consider a simple linear equation (an "Ornstein-Uhlebeck process")

$$\frac{dX(t)}{dt} = -\gamma X(t) + \sigma\xi \,. \tag{22}$$

In Eq. (22), σ=1, γ=0.5, and ξdt is scientists' shorthand for dW. This can be solved analytically for $<X^2(t)>$ = 10. We numerically integrate Eq. (21) using the stochastic Euler scheme Eq. (20) and compare with the results obtained when naively integrated with the deterministic Euler scheme

$$X(t_{n+1}) = X(t_n) + (f(X(t_n), t_n) \quad + g(X(t_n), t_n)\ \sigma\mathscr{R}\ )\ \Delta. \tag{23}$$

An accurate estimation of the integral should not depend on the time step Δ, so we estimate this integral at varying Δ. The results are shown in Fig. 4. Clearly, naive use of the deterministic Euler scheme introduces a spurious dependence on the timestep Δ by the variance of $X(t)$.

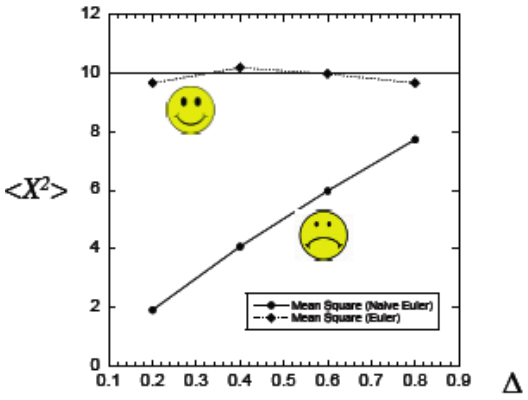

Figure 4: Estimation of $<X^2>$ *as a function of timestep* Δ*. Diamonds: stochastic Euler scheme (Eq. 20). Circles: naive use of deterministic Euler scheme (Eq. 23).*

3.2.2. Does it matter where I call the random number generator?

Although we have hitherto discussed only univariate systems, our next cautionary tale involves a vector process. Let's consider a multivariate version of Eq. (22):

$$d\boldsymbol{X}(t) \;=\; \mathbf{L}\,\boldsymbol{X}(t)\,dt + \;\mathbf{S}\,d\boldsymbol{W}\;, \tag{24}$$

where $\boldsymbol{X}(t)$ is an n-dimensional vector and where $\mathbf{L}$ and $\mathbf{S}$ are chosen so that the true covariance matrix should be

$$< \boldsymbol{X}\boldsymbol{X}^{\mathrm{T}} >= \begin{pmatrix} 2 & 0 & 1 & 0 \\ 0 & 2 & 0 & 1 \\ 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 2 \end{pmatrix}. \tag{25}$$

Numerical integration of this system involves looping through the components of the forcing in such a way that the covariance structure of $\boldsymbol{X}$ is preserved. This is done by defining the random numbers before entering that do-loop and not changing them until the next time step. For example, Fortran code implementing the Euler scheme might read as follows:

```
      do i = 1,n
         r(i) = gasdev(idm)*sqrt(dt)
      enddo
c
      do i = 1,n
         x(i) = x0(i)
         do j = 1,n
            x(i) = x(i) + L(i,j)*x0(i)*dt
     .         +S(i,j)*r(j)
         enddo
      enddo
c
```

In this code snippet, gasdev is a function that generates a centered Gaussian random deviate with unit standard deviation (Press et al. 1992) and idm is the seed that is updated every time the function is called. One thousand samples from the above code resulted in the following sample covariance matrix of $\boldsymbol{X}$:

$$< \boldsymbol{X}\boldsymbol{X}^{\mathrm{T}} >= \begin{pmatrix} 1.98 & -.01 & .99 & .00 \\ -.01 & 1.99 & .02 & 1.00 \\ .99 & .02 & 2.02 & -.01 \\ .00 & 1.00 & -.01 & 2.01 \end{pmatrix}. \tag{26}$$

This matrix shows obvious errors, but should be compared with results from the following snippet of code, where the random number is called within the do-loop:

```
c
      do i = 1,n
         x(i) = x0(i)
         do j = 1,n
            x(i) = x(i) + L(i,j)*x0(i)*dt
     .         +S(i,j)*gasdev(idm)*sqrt(dt)
         enddo
      enddo
c
```

This snippet results in the following covariance matrix for $\boldsymbol{X}$:

$$< \boldsymbol{X}\boldsymbol{X}^{\mathrm{T}} >= \begin{pmatrix} 2.00 & .01 & -.01 & .00 \\ .01 & 2.02 & .01 & -.01 \\ -.01 & .01 & 2.01 & .00 \\ .00 & -.01 & .00 & 2.02 \end{pmatrix}. \tag{27}$$

The use of n^2 random numbers rather than n random numbers introduces random phases that wipe out the covariance structure between different components of $\boldsymbol{X}(t)$.

3.2.3. Can I use deterministic implicit schemes?

No. So far, we have not considered any of the usual stochastic implicit schemes such as implicit Mil'shteyn or implicit Euler (see Kloden and Platen 1992 for the algorithms). In this section, we compare a naive use of a deterministic implicit scheme with the semi-implicit leapfrog scheme of Ewald-Témam (2005), which we introduce in this section. The bottom line is that one should never, ever, put a random number in a denominator. One may be lucky if the simulation is short enough, but in general a random number will eventually cause the system to diverge with probability one.

Consider a simple linear system with multiplicative noise

$$\frac{dx}{dt} = (k^2 - r)x + F \tag{28a}$$

with

$$r = r_o + \eta\xi \tag{28b}$$

and where F, k, and r_o are constant. This is a system for which the equation for the probability density $p(x)$ can be solved analytically. Its graph is shown by the heavy solid line in Fig. 5 (for details of parameter values, etc. see Ewald et al. 2004). Eq. (28) was generated with the semi-implicit leapfrog scheme of Ewald and Témam. (Ewald and Témam 2005; Ewald et al. 2004). That scheme is appropriate for a stochastic differential equation with the parts to be done explicitly and implicitly separated as follows:

$$dx = [a_1(x,t) + a_2(x,t)]dt + G(x,t)(\circ)dW. \tag{29}$$

The Ewald-Témam implicit leapfrog scheme with timestep Δ is as follows:

$$x_1(t_{n+2}) = x_1(t_n) + 2a_1\left(x(t_{n+1}),t_{n+1}\right)\Delta M_n\left(x(t_n),t_n\right) + M_{n+1}\left(x(t_{n+1}),t_{n+1}\right) \tag{30a}$$

$$x(t_{n+2}) = x_1(t_{n+2}) + 2a_2\left(x(t_{n+2}),t_{n+2}\right)\Delta, \tag{30b}$$

with

$$M(x,t) = G(x,t)\frac{\partial G(x,t)}{\partial x}I_{(1,1)} + G(x,t)\Delta W. \tag{30c}$$

The symbols $I_{(1,1)}$ and ΔW in Eq. (30) are discussed in above in section **3***a*. Unlike other schemes, the Ewald-Témam scheme has an unusual discrete estimation of the derivative:

$$\frac{\partial G(x,t)}{\partial x} \approx \frac{G(x+\alpha\sqrt{\Delta},t)-G(x,t)}{\alpha\sqrt{\Delta}}, \tag{30d}$$

where α is a parameter between 0 and 1 having units of $x/\sqrt{\Delta}$.

Getting back to our example, we choose $a_1 = F$, $a_2 = (k^2 - r_o)x$, and $G = \eta x$. The dots with crosses in Fig. 5 are the estimation of $p(x)$ estimated from a smoothed histogram of model output when Eq. (28) was generated with this scheme. For comparison, we consider the results from inserting the stochastic term into a deterministic implicit scheme:

$$x_1(t_{n+2}) = x(t_n) + 2F\Delta \tag{31a}$$

$$x(t_{n+2}) = x_1(t_{n+2}) + 2\Delta[k^2 - (r_o + \eta\xi)]\, x(t_{n+2}). \tag{31b}$$

The results of this integration are indicated by the light line in Fig. 5, which shows a tall, narrow peak centered on the deterministic value of the steady state. Thus, not only is the shape of $p(x)$ misrepresented, but the values most likely to be found in an observed time series of the real system are given very little probability; the peak of $p(x)$ is in the wrong place.

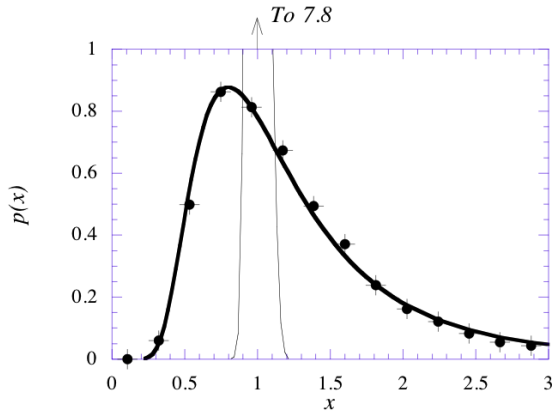

Figure 5: Estimations of the probability density p(x) *appropriate to Eq. 28. Heavy solid line: analytically calculated (i.e., true) curve of* p(x)*. Solid circles with crosses: semi-implicit leapfrog scheme of Ewald-Témam. Light line: naive use of a deterministic implicit scheme.*

4. Discussion and Conclusions

In this exposition, we have discussed a consistent method for using stochastic differential equations to describe a dynamical system containing scales that are rapidly varying and unresolved. We have seen that this method involves a generalized function, the Wiener process, for which the rules of calculus are not unique. We have clarified that Stratonovich calculus is appropriate for continuous systems that contain dynamical components that have finite but unresolved correlation times, and that Ito calculus is appropriate for systems that are multiscale systems non-differentiable at every timescale. The ramifications of multiple calculi to numerical integration were discussed, a few stochastic integration

schemes were introduced, and the reader was warned against several common pitfalls encountered by incorrect application of stochastic modeling.

It is probably unrealistic to expect numerical climate and weather modelers to rewrite their code completely in order to implement ideas that may seem unnecessarily complex. In fact, it is because we *are* scientists that such care is necessary. There is, in principle, a transformation between Ito calculus and Stratonovich calculus so that a mathematical proof for one calculus can, again in principle, be adapted for the other. As scientists, we often need a quantitative estimation of a physical effect, and the mathematical transformation from, say, Ito to Stratonovich calculus may be intractable. A thermometer is not going to do this "Ito transformation" for us and, in any case, we need to know which calculus is appropriate in the first place.

Does this mean we *always* need to rewrite the code to implement stochastic terms in an existing model? The answer is, "No, not always, but be *very* careful." First of all, the semi-implicit leapfrog scheme of Ewald and Témam (2005) was developed to add minimal changes to the usual schemes used, for example in barotropic vorticity models (e.g., Sardeshmukh and Hoskins 1988). Secondly, in explicit schemes such as Runge-Kutta, one can often get away with using deterministic code by replacing the amplitude σ of the noise at each timestep Δ with $\sigma/\sqrt{\Delta}$ (Hansen and Penland 2006), if the timestep Δ is so small that it is smaller than *any* timescale in the system, one may avoid having to insert the random number at every timestep and, thus, be able to use larger timesteps than strictly are required by stochastic integration schemes. Hansen and Penland (2007) show that they were able to inject the noise every N timesteps using deterministic code by replacing the amplitude of the σ of the noise by $\sigma/\sqrt{(N\Delta)}$. This is useful if the amplitude of the noise has been diagnosed using data assimilation using an assimilation window equal to $N\Delta$. If the modeler chooses to use any of the "beat the system" methods discussed in this paragraph, (s)he does so at his or her own risk.

Acknowledgements: The author wishes to acknowledge the years of discussion and collaboration with friends and colleagues at NOAA/ESRL and elsewhere, particularly Prashant Sardeshmukh, Roger Témam, Brian Ewald, and James A. Hansen.

REFERENCES

(See also references therein)

Ewald, B., C. Penland and R. Témam, 2004: Accurate integration of stochastic climate a models. *Mon. Wea. Rev.*, **132**, 134-164. (Beware of Fig. 1)

Ewald, B., and C. Penland, 2009: Numerical generation of stochastic differential equations in climate models, in *Handbook of Numerical Analysis XIV: Computational Methods for the Atmosphere and Oceans.* Témam and Tribbia, eds., North Holland, Oxford, pgs 279-306.

Hansen, J. A., and C. Penland, 2006: Efficient approximation techniques for integrating stochastic differential equations. *Mon. Wea. Rev.*, **134**, 3006-3014.

Hansen, J. A., and C. Penland, 2007: On stochastic parameter estimation using data assimilation. *Physica D*, **230**, 88-98.

Kloden, P.E., and E. Platen, 1992: *Numerical Solution of Stochastic Differential Equations.* Springer, Berlin, 632pp.

Majda, A. J., I. Timofeyev, and E. Vanden Eijnden, 1999: Models for stochastic climate prediction. *Proc. Natl. Acad. Sci. USA*, **96**, 14 687–14 691.

Papanicolaou, G., and W. Kohler, 1974: Asymptotic theory of mixing stochastic differential equations. *Comm. Pure Appl. Math.*, **27**, 641-668.

Press, W.H., B.P.Flannery, S.A. Teukolsky, and W.T.Vetterling, 1992: *Numerical Recipes in Fortran: the Art of Scientific Computing.* Cambridge, 933pp.

Rümelin, W,, 1982: Numerical Treatment of Stochastic Differential Equations. *SIAM J. Numer. Anal.*, **19**, 604-613.

Sardeshmukh, P. D., and B. Hoskins, 1988: The generation of global rotational flow by steady idealized tropical convergence. *J. Atmos. Sci.*, **45**, 1228–1251.