

On modelling physical systems with stochastic models: diffusion versus Lévy processes

Cécile Penland and Brian D Ewald

Phil. Trans. R. Soc. A 2008 **366**, 2455-2474 doi: 10.1098/rsta.2008.0051

| References | This article cites 34 articles http://rsta.royalsocietypublishing.org/content/366/1875/2455.ful l.html#ref-list-1 |
|------------------------|--|
| Rapid response | Respond to this article http://rsta.royalsocietypublishing.org/letters/submit/roypta;366/ 1875/2455 |
| Subject collections | Articles on similar topics can be found in the following collections |
| | atmospheric science (46 articles) |
| Email alerting service | Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click here |

To subscribe to *Phil. Trans. R. Soc. A* go to: http://rsta.royalsocietypublishing.org/subscriptions



On modelling physical systems with stochastic models: diffusion versus Lévy processes

By Cécile $\operatorname{Penland}^{1,*}$ and Brian D. Ewald^2

¹NOAA/ESRL/Physical Sciences Division, 325 Broadway, Boulder, CO 80305, USA ²Department of Mathematics, Florida State University, Room 205C, 1017 Academic Way, Tallahassee, FL 32306-4510, USA

Stochastic descriptions of multiscale interactions are more and more frequently found in numerical models of weather and climate. These descriptions are often made in terms of differential equations with random forcing components. In this article, we review the basic properties of stochastic differential equations driven by classical Gaussian white noise and compare with systems described by stable Lévy processes. We also discuss aspects of numerically generating these processes.

Keywords: stochastic differential equations; Lévy processes; Gaussian processes; numerical methods

1. Introduction

Subgrid-scale processes must be treated somehow in numerical weather and climate models, whatever these models' spatial and temporal resolutions. First of all, one could ignore them. Although this is often done, the procedure has not proven particularly successful. A more common, if not *the* most common, approach is to parametrize them deterministically in terms of resolved processes. Some authors even define 'parametrization' that way, reserving the term 'unparametrizable' to describe what cannot be represented in terms of what a numerical model can resolve.

Modern numerical models (e.g. Buizza *et al.* 1999) employ stochastic parametrizations to account for both mean effects and variability due to dynamical interactions between processes that cannot be explicitly represented in a numerical model and the large-scale systems the model attempts to predict. At the European Centre for Medium-Range Weather Forecasting, for example, efforts are made to parametrize turbulent energy backscatter (e.g. Fjørtoft 1953) from the unresolved scales to the resolved scales in the ensemble prediction system by means of cellular automata (Shutts 2005) or using red noise processes (Berner *et al.* submitted).

There is both empirical and modelling evidence that such an approach is both practical and physically meaningful. The almost embarrassing similarity of

*Author for correspondence (cecile.penland@noaa.gov).

One contribution of 12 to a Theme Issue 'Stochastic physics and climate modelling'.

prediction skill in statistical and numerical models of El Niño (e.g. Penland & Sardeshmukh 1995, PS95 hereafter; Saha *et al.* 2006) and medium-range weather (Winkler *et al.* 2001) lends credibility to the idea that at least some aspects of dynamical forcing can be treated stochastically (see also Hasselmann 1976).

The fact that forcing may be treated stochastically does not mean that details of the stochastic treatment are arbitrary. Failure to identify the physical origins of the stochastic forcing, which usually results in a mistaken mathematical description of that forcing, can easily translate into a misinterpretation of the dynamical response. For example, PS95 present evidence that El Niño dynamics, as represented by tropical sea-surface temperatures (SSTs), is very well explained in terms of a stable linear process driven by Gaussian stochastic forcing. In fact, the 'tau test' (see their article for details) that is passed by the statistics of tropical SSTs can *only* be passed by such a system. The problem with PS95 is that the authors placed too strong an emphasis on tests for Gaussianity of the SSTs themselves; while a linear dynamical system driven by Gaussian forcing may be Gaussian, that need not be the case if the stochastic forcing is modulated by a linear function of the SSTs (Müller 1987; Sardeshmukh & Sura submitted; Sura & Sardeshmukh 2008). That is, if the amplitudes of rapidly varying wind stress and heat flux depend on SST, which they do, then the distribution of SST cannot be expected to be Gaussian even when the underlying SST dynamics are stable and linear. A result of PS95's misplaced emphasis on Gaussianity was the publication of articles (e.g. An & Jin 2004), which found a small but significant non-Gaussian behaviour in tropical east Pacific SST, and concluded that the dynamics of El Niño must be primarily nonlinear. In other words, scientists were arguing about the resolved dynamics of an extremely important phenomenon without giving due consideration to the nature of the stochastic forcing. There were indeed studies that considered the effect of stochastic forcing in the El Niño system (e.g. Flügel *et al.* 2004), but these did not consider how the basic mathematical description of the stochastic forcing could affect the marginal distribution of SST.

It turns out that the marginal distribution of a linear Gaussian-driven process may not only be non-Gaussian but may also exhibit skew, fat tails and other properties usually associated with more exotic types of stochastic phenomena. such as non-Gaussian α -stable Lévy processes. Even when the dynamical system can be shown to be linear and Gaussian driven, the distribution of that system depends on whether the dynamics describing it respond to the stochastic forcing through 'Itô' or 'Stratonovich' calculus, a differentiation that we explain below. Nevertheless, although these characteristics permit linear dynamics, they differ from the tau test in which they do not require it. Further, the distribution of these different kinds of stochastic dynamics may have many characteristics in common, but the classes of physical causes giving rise to them are quite different. Thus, when using stochastic numerical models of finite resolution to diagnose the physical system, it is not only necessary to choose accurately the stochastic parametrizations appropriate to the class of subgrid-scale processes one wishes to represent but also to ensure that the numerical algorithms built into the models are able to reproduce the correct stochastic response.

The stochastic parametrizations for which the following remarks are appropriate comprise a somewhat limited class of stochastic models. There are certainly other valuable methods, such as random cascades and particle-interaction

techniques, which have a valid place in the literature. However, in this article, we shall confine ourselves to the rather restricted field of stochastic differential equations (SDEs) driven by either Gaussian white noise, i.e. diffusion processes. or, more generally, by α -stable Lévy processes. The special case of diffusion processes has already been extensively treated in the climate literature (e.g. Hasselmann 1976; Penland 1989; Penland 1996; Majda et al. 1999; Sardeshmukh & Sura submitted), but this literature has yet to be widely applied in climate science. At the same time, the ability of Lévy processes to describe properties such as intermittency has attracted the attention of climate scientists, some of whom (e.g. Ditlevsen 1999) have begun to employ a wider class of SDEs (LSDEs) driven by stable Lévy processes to describe observational records in palaeoclimatology. Many scientific studies using LSDEs do not allow the random forcing to depend on the state of the system, i.e. they employ additive rather than multiplicative Lévy noise. We follow this approach as it greatly simplifies both the theoretical and the numerical descriptions of such systems, deferring the more complex issues to a later paper.

The purpose of this article is to present a fairly basic synopsis of properties of diffusion processes and stable Lévy processes. We emphasize what we believe would be useful to scientists who wish to use SDEs in stochastic parametrizations. This exposition is certainly not exhaustive and is not meant to be so. Rather, we hope we present a starting point from which scientists can develop an intuitive feel for the reasoning behind SDEs and an appreciation for the necessity of rigor. Rather than repeat lengthy derivations, we have tried to give a qualitative understanding of the physical processes for which a class of stochastic models is appropriate. When possible, we have tried to provide sufficient quantitative guidance for the numerical generation of that class.

The article is organized as follows: §2 reviews classical SDEs based on Langevin systems with Brownian motion, i.e. Markov diffusion processes. Section 3 discusses α -stable Lévy processes and simple additive LSDEs. Section 4 provides examples of the stochastic models discussed in §§2 and 3, including a numerical comparison of a simple system driven by Gaussian white noise with that same system driven by a symmetric α -stable Lévy process with identical scale parameter. Section 5 concludes the article.

2. Stochastic systems with Brownian motion

(a) Preliminary discussion

Classical SDEs are valid when there is a clear time-scale separation between 'fast' and 'slow' terms in a dynamical system, although many researchers (Dozier & Tappert 1978*a,b*; Penland 1985; Majda *et al.* 1999) have found that this restriction can be quite forgiving, and that useful results can be made even when a clear time-scale separation is not obtained. Hasselmann (1976) assigns meteorological meaning to the fast and slow systems, which he calls 'weather' and 'climate', respectively. We sometimes use this terminology, although the reader should in no way take this usage as strict definitions of weather and climate.

As an example of a system that might be governed by an SDE, consider a state vector $\boldsymbol{x}(t)$ representing some surface phenomenon (e.g. SST, etc.) in the tropical

C. Penland and B. D. Ewald

ocean. Let us also imagine that the evolution of $\boldsymbol{x}(t)$ might be written as a differential equation as follows:

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = \boldsymbol{c}(\boldsymbol{x}, t) + \boldsymbol{w}(\boldsymbol{x}, t).$$
(2.1)

In equation (2.1), the climate variable c(x, t) could be modelled easily using a time step of, say, 10 days. This was the time step used in the El Niño model of Zebiak & Cane (1986). By contrast, the weather term w(x, t) may represent tropical convection, so that an accurate model of this term would require a time step on the order of minutes or seconds; the mesoscale model MM5 having a spatial resolution of 1 km uses a time step of 3 s (G. Bryan & T. Hamill 2007, personal communication). It is often assumed that w(x, t) can either be parametrized by processes represented by c(x, t) or even neglected completely on large time scales if it varies rapidly enough. One then assumes that the long time-scale evolution of x(t) in equation (2.1) might be well described using only c(x, t) in a model integrated with a 10-day time step. Unfortunately, this is very inaccurate for highly nonlinear systems. On the other hand, time steps small enough to resolve the interactions between c(x, t) and w(x, t) may be impractical. If the time-scale separation between c(x, t) and w(x, t) is sufficiently large, we can still model an approximate version of equation (2.1) using the longer time step, with w(x, t) replaced by a deterministic function of x and t (which may or may not be a constant) multiplying a Gaussian stochastic quantity. That is, w(x, t) varies rapidly enough so that its autocorrelation has decayed to insignificance over the course of the long time step, but the size of w(x, t), though finite, is big enough that its effects on x cannot be neglected. Thus, the effects of nearly independent values of w(x, t) are combined in such a way that central limit theorem (CLT) behaviour obtains, and Gaussian statistics are introduced into the evolution equation for \boldsymbol{x} on long enough time scales.

(b) An approximation using standard Brownian motion

The Gaussian stochastic quantity introduced in the previous paragraph is not arbitrary, but rather has a variance dependent on the time scales of the system. Quantifying these ideas requires consideration of a 'Wiener process', also called 'Brownian motion', which we denote by W(t). The Brownian motion is also sometimes called a 'continuous random walk'. We shall use these terms interchangeably. Using angle brackets to denote expectation values, we state the following properties of the vector Wiener process W.

W(t) is a vector of Gaussian random variables, and

$$\boldsymbol{W}(0) = \boldsymbol{0}, \tag{2.2a}$$

$$\langle \boldsymbol{W}(t) \rangle = \boldsymbol{0}, \qquad (2.2b)$$

$$\langle \boldsymbol{W}(s) \, \boldsymbol{W}^{\mathsf{T}}(t) \rangle = \mathsf{I} \min(s, t)$$
 (2.2*c*)

and

$$\langle \mathrm{d}\boldsymbol{W}(s)\mathrm{d}\boldsymbol{W}^{\mathsf{T}}(t)\rangle = \mathsf{I}\,\delta(s-t).$$
 (2.2*d*)

In equation (2.2), I denotes the identity matrix, and the δ function in equation (2.2d) approaches dt as s goes to t. The Wiener process is continuous but is only differentiable in a generalized sense,

$$\mathrm{d}W_k = \xi_k \,\mathrm{d}t. \tag{2.3}$$

2461

Equation (2.3) defines the kth component ξ_k of 'white noise'. Note that the units of $W_k(t)$ are the square root of time; numerical stochastic models typically involve terms based on deterministic dynamics, which are updated with increments equal to the time step, and other terms involving stochastic terms, which are updated using increments equal to the square root of the time step. More detailed descriptions of Wiener processes and white noise may be found in Arnold (1974, ch. 3).

The Fourier spectrum of $W_k(t)$ varies everywhere as the inverse square of the frequency. Also note that the inverse square spectrum implies that every finite sample of Brownian motion is dominated by an oscillation having a period near to the length of the time series.

Before continuing on to how the Wiener process is used in modelling a physical system with an SDE, it is necessary to mention a crucially important property: unlike the deterministic Riemann integral, which has a single 'fundamental theorem of integral calculus' (e.g. Purcell 1972), it is possible to define infinitely many different integration rules involving Brownian motion. Two of these calculi are found in nature. The one that primarily interests us corresponds to the continuous multiscale interaction problem with which we began this section. It is called 'Stratonovich calculus' and has the same integration rules as standard Riemannian calculus. The other physically meaningful calculus, called 'Itô calculus', obtains when a system consists of discrete jumps, but the time between jumps is vanishingly small compared with the time scales of interest. That is, there are two limits here that do not commute, the white noise and the continuous limits, and we want to take them both. A computer, of course, requires discretization with a small enough time step that the system is approximately continuous, but this is yet another approximation and is separate from the two limits that obtain even before we begin to think about the computer program. Now, even though a hydrodynamic system is made up of molecules, we usually base the models of the ocean and the atmosphere on a continuous set of deterministic equations, such as rotating Navier–Stokes. Then, in order to make progress, we make some assumptions about the time scales of the system. This is equation (2.1). Finally, we use the dynamic CLT to approximate the fast variables as stochastic terms. These are the conditions leading to Stratonovich calculus, and appropriate numerical schemes are required to approximate numerically the solution to the SDE. We revisit this issue below.

Let us consider a case that is less often used, but may be geophysically relevant in some cases, where a fluid is at an intermediate level of concentration where it is dense enough that continuity equations are still valid on average, but rarified enough that momentum transfer through frequent individual molecular collisions presents a non-negligible stochastic effect. In this case, the physical system might well be described as an Itô SDE, and numerical models of it require different schemes from those modelling Stratonovich systems. Fortunately, although numerical schemes describing the different calculi are not the same, they are often rather similar. Exhaustive discussions may be found in, for example, Kloeden & Platen (1992).

As might be expected, Itô and Stratonovich numerical schemes converge to the same answer in the deterministic limit. This most emphatically does *not* mean that a 'stochastification' of a deterministic numerical scheme is obvious. Computers do not care whether a numerically generated solution to an equation

C. Penland and B. D. Ewald

is physical or not, and they can quite happily spit out a perfectly accurate solution to an SDE corresponding to a calculus that does not exist in nature if the numerical scheme is not appropriate to the physical system at hand. We will revisit this issue below, but first we need to set up the problem correctly. This is the subject of the 2c.

(c) The central limit theorem

Informally, the traditional CLT (e.g. Doob 1953; Wilks 1995) usually employed by geoscientists to justify use of Gaussian distributions states that the sum of independently sampled quantities having finite variance is approximately Gaussian. As discussed above, we consider here dynamical systems described by a slow time scale and faster time scales. The equations are averaged over a temporal interval large enough that the fast time scales collectively act as Gaussian stochastic forcing of the slow, coarse-grained system. In the mathematical literature (e.g. Feller 1966), the fact that fine details of how the fast processes are distributed do not strongly affect the coarse-grained behaviour of the slower dynamics is often called 'the invariance principle'. As the proof of this theorem is outside the scope of this paper, we state a commonly used form of it and refer the interested reader to the literature for details. Gardiner (1985) gives a heuristic description; we prefer Papanicolaou & Kohler (1974) for a technical statement of the theorem.

A dynamical system consisting of separated time scales may be written in the following manner:

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = \varepsilon \boldsymbol{G}(\boldsymbol{x}, t) + \varepsilon^2 \boldsymbol{F}(\boldsymbol{x}, t), \qquad (2.4)$$

where \boldsymbol{x} is an *N*-dimensional vector. In equation (2.4), the smallness parameter ε should not be taken as a measure of importance. It does measure the rapidity with which the terms on the r.h.s. of (2.4) vary relative to each other; one can think of ε^2 as the ratio of the characteristic time scale of the first term to the characteristic time scale of the second. If we now cast (2.4) in terms of a scaled time coordinate

$$\Delta s = \varepsilon^2 \Delta t, \tag{2.5}$$

it becomes

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}\boldsymbol{s}} = \frac{1}{\varepsilon} \boldsymbol{G}\left(\boldsymbol{x}, \frac{s}{\varepsilon^2}\right) + \boldsymbol{F}\left(\boldsymbol{x}, \frac{s}{\varepsilon^2}\right). \tag{2.6}$$

We further assume that the first term in (2.4) decays quickly 'enough' in the time interval Δt . In the limit $\varepsilon \to 0$, $\Delta t \to \infty$ with $\varepsilon^2 \Delta t$ remaining finite, the CLT states that (2.6) converges weakly to a Stratonovich SDE in the scaled coordinates,

$$d\boldsymbol{x} = \boldsymbol{F}'(\boldsymbol{x}, s)ds + \boldsymbol{\mathsf{G}}'(\boldsymbol{x}, s) \circ d\boldsymbol{W}(s). \tag{2.7}$$

The primes in (2.7) denote only that F and G may be somewhat different than the terms in (2.4). W in (2.7) is a K-dimensional vector, each component of which is an independent Wiener process, or Brownian motion, and the ' \circ ' indicates that it is to be interpreted in the sense of Stratonovich. G'(x, s) is a matrix, the first index of which corresponds to a component of x and the second index of which corresponds to a component of dW.

For details and proof of the CLT, we recommend, for example, the articles by Wong & Zakai (1965), Khasminskii (1966) and by Papanicolaou & Kohler (1974). Examples of geophysical applications may be found in Kohler & Papanicolaou (1977), Penland (1985), Majda *et al.* (1999) and Sardeshmukh *et al.* (2001).

Remark 8 of Papanicolaou & Kohler (1974) is particularly applicable to many problems. In that case, the *i*th component of the rapidly varying term in equation (2.6) can be given as

$$G_i\left(\boldsymbol{x}, \frac{s}{\varepsilon^2}\right) = \sum_{k=1}^K G_{ik}(\boldsymbol{x}, s) \eta_k\left(\frac{s}{\varepsilon^2}\right), \qquad (2.8)$$

where $\eta_k(s/\epsilon^2)$ is a stationary, centred and bounded random function. The integrated lagged covariance matrix of η is defined to be

$$C_{km} \equiv \int_0^\infty \left\langle \eta_k(t)\eta_m(t+t') \right\rangle \,\mathrm{d}t', \quad k,m = 1,2,\dots,K.$$
(2.9)

With these restrictions, the CLT states that in the limit of long times $(t \rightarrow \infty)$ and small ε ($\varepsilon \rightarrow 0$), taken so that $s = \varepsilon^2 t$ remains fixed, the conditional probability density function (pdf) for \boldsymbol{x} at time s given an initial condition $\boldsymbol{x}_0(s_0)$ satisfies the backward Kolmogorov equation (e.g. Horsthemke & Lefever 1984; Bhattacharya & Waymire 1990),

$$\frac{\partial p(\boldsymbol{x}, s | \boldsymbol{x}_0, s_0)}{\partial s_0} = \mathcal{L} p(\boldsymbol{x}, s | \boldsymbol{x}_0, s_0), \qquad (2.10)$$

where

$$\mathcal{L} = \sum_{i,j=1}^{N} a_{ij}(\boldsymbol{x}_0, s_0) \frac{\partial^2}{\partial x_{0i} \partial x_{0j}} + \sum_{i=1}^{N} b_i(\boldsymbol{x}_0, s_0) \frac{\partial}{\partial x_{0i}}$$
(2.11)

and

$$a_{ij}(\boldsymbol{x}, s) = \sum_{k,m=1}^{K} C_{km} G_{ik}(\boldsymbol{x}, s) G_{jm}(\boldsymbol{x}, s), \qquad (2.12a)$$

$$b_i(\boldsymbol{x},s) = \sum_{k,m=1}^{K} C_{km} \sum_{j=1}^{N} G_{jk}(\boldsymbol{x},s) \frac{\partial G_{im}(\boldsymbol{x},s)}{\partial x_j} + F_i(\boldsymbol{x},s).$$
(2.12b)

In this limit, the conditional pdf also satisfies a forward Kolmogorov equation (called a 'Fokker–Planck equation' in the scientific literature) in the scaled coordinates

$$\frac{\partial p(\boldsymbol{x}, s | \boldsymbol{x}_0, s_0)}{\partial s} = \mathcal{L}^{\dagger} p(\boldsymbol{x}, s | \boldsymbol{x}_0, s_0), \qquad (2.13)$$

where

$$\mathcal{L}^{\dagger}p = \sum_{i,j=1}^{N} \frac{\partial^2}{\partial x_i \partial x_j} [a_{ij}(\boldsymbol{x}, s)p] - \sum_{i=1}^{N} \frac{\partial}{\partial x_i} [b_i(\boldsymbol{x}, s)p].$$
(2.14)

In short, the *moments* of \boldsymbol{x} can be approximated by the moments of the solution to the Stratonovich SDE

$$d\boldsymbol{x} = \boldsymbol{F}(\boldsymbol{x}, s) \, ds + \boldsymbol{\mathsf{G}}(\boldsymbol{x}, s) \boldsymbol{\mathsf{S}} \circ d\boldsymbol{W}. \tag{2.15}$$

C. Penland and B. D. Ewald

In equation (2.15), $\mathbf{G}(\mathbf{x}, s)$ is the matrix whose (i, k)th element is $G_{ik}(\mathbf{x}, s)$ and \mathbf{S} is a matrix where the (k, m)th element of \mathbf{SS}^{T} is C_{km} . Note that \mathbf{S} is only unique up to its product with an arbitrary orthogonal matrix. Also note that the usual factor of one-half found in most formulations of the Fokker–Planck equation has been absorbed into the definition of C_{km} .

From the first temporal derivative in (2.13) and the second derivative with respect to the state variable in (2.14), it can be seen that the Fokker–Planck equation is a type of diffusion equation. For this reason, b(x, s) in (2.14) is often called the 'drift', $\mathbf{a}(x, s)$ is called the 'diffusion', and the process x is called a 'Markov diffusion process'. This term is used to make the distinction between systems driven by Gaussian white noise and those driven by non-Gaussian stable Lévy processes, as discussed below.

There do exist Kolmogorov equations (equations (2.10) and (2.13)) for Itô systems, with the modification that in (2.12b), $b_i(\boldsymbol{x}, s)$ is simply equal to $F_i(\boldsymbol{x}, s)$. The difference between Itô and Stratonovich calculi is important for scientists to understand because most of the physical phenomena they deal with are Stratonovich, while most mathematical references on stochastic numerical techniques are primarily interested in Itô schemes. There is a formal equivalence between Itô & Stratonovich descriptions of reality, so a theorem about an Itô process can generally be carried over to the corresponding Stratonovich process. More precisely, the Stratonovich SDE (absorbing the matrix **S** into **G**)

$$dx_i = F_i(\boldsymbol{x}, t)dt + \sum_{\alpha} G_{i\alpha}(\boldsymbol{x}, t) \circ dW_{\alpha}$$
(2.16a)

is equivalent to the Itô SDE

$$dx_{i} = \left[F_{i}(\boldsymbol{x},t) + \frac{1}{2}\sum_{\alpha,j}G_{j\alpha}(\boldsymbol{x},t)\frac{\partial G_{i\alpha}(\boldsymbol{x},t)}{\partial x_{j}}\right]dt + \sum_{\alpha}G_{i\alpha}(\boldsymbol{x},t)dW_{\alpha}.$$
 (2.16b)

By 'equivalent' it is meant that solving equation (2.16a) using Stratonovich calculus gives the same solution as solving equation (2.16b) using Itô calculus. That is, each equation evaluated for \boldsymbol{x} using its appropriate calculus would describe an experimental outcome equally well as the other; the statistics of \boldsymbol{x} in each case are the same. Note that if $\mathbf{G}(\boldsymbol{x}, \boldsymbol{s})$ is independent of \boldsymbol{x} . Itô and Stratonovich calculi converge to each other. In equation (2.16b), we have omitted 'o' in keeping with standard mathematical notation of an Itô SDE. Unfortunately, the transformation from one description to the other can be prohibitively difficult in physical applications.

For longer discussions on the difference between Itô and Stratonovich systems, we refer the reader to Arnold (1974), Horsthemke & Lefever (1984), Kloeden & Platen (1992) and Ewald & Penland (2008). The important message here is for when one wishes to approximate a continuous real system with short but finite correlation time in a general circulation model (GCM) with a Gaussian stochastic variable. As noted in equation (2.9), the correlation structure of the fast variable comes into play. Simply replacing the fast term with a Gaussian random deviate with standard deviation equal to that of the variable to be approximated, and then using deterministic numerical integration schemes, is a recipe for disaster (Sura & Penland 2002; Ewald *et al.* 2004).

(d) Notes on numerical techniques involving SDEs

One can write, and some have written (e.g. Kloeden & Platen 1992), enormous tomes on the theory and practice of numerically integrating SDEs. A review of some of these techniques, including longer discussions of the schemes we present here, may be found in Ewald & Penland (2008). In this subsection, we confine ourselves to discussing the procedure and results of using an explicit stochastic Euler scheme (Rümelin 1982), an explicit stochastic Heun scheme (Rümelin 1982) and an implicit stochastic integration scheme developed by Ewald & Temam (2005; see also Ewald *et al.* 2004) especially for spectral versions of GCMs. To set notation, we denote the time step by Δ , while the $\{z_{\mu i}\}$ denote centred Gaussian random deviates, each with *variance* Δ , sampled at time t_i . If μ varies from 1 to M, then $z_{\mu i}$ is the μ th component of the M-dimensional vector z_i . x_m represents the *m*th component of an N-dimensional vector x obeying the following SDE:

$$d\boldsymbol{x} = \boldsymbol{F}(\boldsymbol{x}, s)ds + \boldsymbol{\mathsf{G}}(\boldsymbol{x}, s)(\circ)d\boldsymbol{W}.$$
(2.17)

In equation (2.17), $G(x, s)(\circ) dW$ equals G(x, s) dW if the system is Itô, and $G(x, s) \circ dW$ if it is Stratonovich.

The stochastic Euler scheme

$$x_m(t_{i+1}) = x_m(t_i) + F_m(\boldsymbol{x}, t_i)\Delta + \sum_{\mu} G_{m\mu}(\boldsymbol{x}, t_i) z_{\mu i}$$
(2.18)

converges to Itô calculus, unless $\mathbf{G}(\mathbf{x}, s)$ is not really a function of \mathbf{x} . In that case, we repeat, the Itô and Stratonovich calculi give the same moments of \mathbf{x} . It is important that the vector \mathbf{z}_i be generated outside the loop performing the summation in (2.18); otherwise, random phases are generated, which erroneously eliminate the contribution to $\langle \mathbf{x}\mathbf{x}^{\mathsf{T}} \rangle$ by any off-diagonal elements of \mathbf{GG}^{T} . Again, if $\mathbf{G}(\mathbf{x}, s)$ is a function of \mathbf{x} , equation (2.18) generates Itô calculus.

For reasons explained elsewhere (e.g. Rümelin 1982; Ewald & Penland 2008), explicit schemes generating Stratonovich calculus are usually predictor–corrector methods. The simplest version is a second-order Runge–Kutta scheme, called the Heun scheme. Here, one generates an intermediate variable using an Euler estimate

$$x'_{m}(t_{i+1}) = x_{m}(t_{i}) + F_{m}(\boldsymbol{x}, t_{i})\Delta + \sum_{\mu} G_{m\mu}(\boldsymbol{x}, t_{i})z_{\mu i}, \qquad (2.19a)$$

which is then updated as follows:

$$x_{m}(t_{i+1}) = x_{m}(t_{i}) + \frac{1}{2} \{F_{m}(\boldsymbol{x}, t_{i}) + F_{m}(\boldsymbol{x}', t_{i+1})\} \Delta + \frac{1}{2} \sum_{\mu} \{G_{m\mu}(\boldsymbol{x}, t_{i}) + G_{m\mu}(\boldsymbol{x}', t_{i+1})\} z_{\mu i}.$$
(2.19b)

Note that the same random vector z_i is used in both (2.19a) and (2.19b).

Our final example of a numerical scheme designed for SDEs is the implicit scheme of Ewald & Temam (2003, 2005). This algorithm was devised to accommodate the architecture of extant climate models, including barotropic vorticity models (e.g. Sardeshmukh & Hoskins 1988) and full GCMs

(e.g. Saha *et al.* 2006). The deterministic climate models usually integrate the state vector first using a leapfrog step, followed by an implicit step. To implement the stochastic analogue of this procedure, we rewrite equation (2.17) as

$$d\boldsymbol{x} = \boldsymbol{F}_1(\boldsymbol{x}, t)dt + \boldsymbol{F}_2(\boldsymbol{x}, t)dt + \boldsymbol{\mathsf{G}}(\boldsymbol{x}, t)(\circ)d\boldsymbol{W}. \tag{2.20}$$

In (2.20), $F_1(\mathbf{x}, t)$ and $F_2(\mathbf{x}, t)$ are the explicit and the implicit parts of the model, respectively. The implicit leapfrog scheme of Ewald & Temam (2003, 2005) is as follows:

$$\boldsymbol{x}'(t_{i+2}) = \boldsymbol{x}(t_i) + 2\boldsymbol{F}_1(\boldsymbol{x}(t_{i+1}), t_{i+1})\boldsymbol{\Delta} + \boldsymbol{M}(\boldsymbol{x}(t_i), t_i) + \boldsymbol{M}(\boldsymbol{x}(t_{i+1}), t_{i+1}), \quad (2.21a)$$

and

$$\boldsymbol{x}(t_{i+2}) = \boldsymbol{x}'(t_{i+2}) + 2\boldsymbol{F}_2(\boldsymbol{x}(t_{i+2}), t_{i+2})\boldsymbol{\Delta}.$$
 (2.21b)

In the updating expressions, the *m*th component of the vector $\boldsymbol{M}(\boldsymbol{x}(t_i), t_i)$ is

$$M_m(\boldsymbol{x}, t_i) = \sum_{n, \mu, \nu} G_{n\mu}(\boldsymbol{x}, t_i) \frac{\partial G_{m\nu}(\boldsymbol{x}, t_i)}{\partial x_n} I_{(\mu, \nu)} + \sum_{\mu} G_{m\mu}(\boldsymbol{x}, t_i) z_{\mu}.$$
 (2.21c)

The derivative in (2.21c) can be approximated as

$$\frac{\partial G_{m\nu}(\boldsymbol{x}, t_i)}{\partial x_n} = \frac{G_{m\nu}(\boldsymbol{x} + \varepsilon_n \sqrt{\Delta} \hat{\boldsymbol{e}}_n, t_i) - G_{m\nu}(\boldsymbol{x}, t_i)}{\varepsilon_n \sqrt{\Delta}}.$$
 (2.21*d*)

In (2.21d), $\hat{\boldsymbol{e}}_n$ is a unit vector corresponding to the component x_n . The vector $\boldsymbol{\varepsilon}$ has components less than or equal to unity, in units of \boldsymbol{x}/\sqrt{t} , and allows the modeller to adjust the discretized derivatives to the problem at hand. The difference between Itô and Stratonovich calculi is effected through estimates of the multiple stochastic integral in (2.21c)

$$I_{(\mu,\nu)} = \frac{1}{2} (z_{\mu} z_{\nu} - \delta_{\mu\nu} \Delta) \quad (\text{Itô}), \qquad (2.22a)$$

$$I_{(\mu,\nu)} = \frac{1}{2} z_{\mu} z_{\nu} \quad \text{(Stratonovich)}. \tag{2.22b}$$

In this and in all implicit stochastic algorithms, the stochastic terms enter the problem during the explicit step. When random numbers occur in the denominator of the implicit step, the numerical solution eventually explodes.

3. Aspects of modelling with Lévy processes

(a) Preliminary discussion

The fact that a process has infinite variance, or even infinite mean, does not preclude its existence. Consider the game proposed in 1713 by Nicolaus Bernoulli in a letter to Pierre Raymond de Montmort. A player flips a coin repeatedly until he gets 'tails', at which point the game ends. If the tails appears on the first flip, he wins \$1. If it appears on the second flip, he wins double that, or \$2. If he achieves two heads before a tails on the third flip, he wins double *that*, or \$4. If the tails shows up on the (k+1)th flip, he wins 2^k . Thus, his expected winnings are a sum of terms, each consisting of the probability that a certain event occurs

2466

times the return, or $1/2 + (1/4) \cdot 2 + (1/8) \cdot 4 + \cdots$, which sums to infinity. Of course, it would take an infinite amount of time to get this infinite return, but that does not mean the game cannot be played to its end. This paradox is called 'the St Petersburg paradox' since it was published in 1738 by Nicolaus' brother Daniel Bernoulli (presumably with Nicolaus' permission) in the *Commentaries of the Imperial Academy of Science of St Petersburg*.

With these ideas in mind, we return to the classic CLT, where independent, identically distributed normalized variables are added together. However, the condition of finite variance is relaxed. The result is a class of distributions called ' α -stable Lévy distributions', of which the Gaussian distribution is a special case (Lévy 1937; Gnendenko & Kolmogorov 1954). These distributions are characterized by a parameter α , $0 < \alpha \leq 2$, for which moments of order μ diverge for $\mu \geq \alpha$. The exception to this rule is when $\alpha = 2$, which corresponds to the Gaussian distribution. The reason these distributions are called 'stable' is because sums of appropriately centred and normalized (by $n^{1/\alpha}$, with n being the number of terms in the sum) variables sampled from such a distribution belong to the same distribution. For example, the sum of Gaussian variables is also a Gaussian variable.

The class of α -stable Lévy processes are themselves a special case of systems called simply Lévy processes (Appelbaum 2004; Protter 2005). All that is required of a Lévy process X(t) is that (i) it has independent increments, i.e. X(t) - X(s) is independent of X(r) - X(q) at times q, r, s and t for increments that do not overlap, (ii) the increments are stationary, i.e. the pdf of X(t) - X(s) is the same as that of X(t-s), (iii) X(t) is continuous in probability, i.e. for any $\delta > 0$, no matter how small δ is, the probability that $|X(t) - X(s)| > \delta$ is zero in the limit that $t \rightarrow s$, and (iv) X(0) = 0. Theorems proved for Lévy processes are obviously true for α -stable Lévy processes.

Why do we care about non-Gaussian α -stable Lévy processes? Well, as we saw in 2, the distance a Brownian particle travels from the origin varies as

$$|W(t)|^2 \sim t. \tag{3.1}$$

More generally, the diffusion away from its origin travelled by an α -stable Lévy process $L_{\alpha}(t)$ is given by

$$|L_{\alpha}(t)|^2 \sim t^{2/\alpha}.$$
(3.2)

Thus, for $\alpha < 2$, large excursions from the mean can occur in a much shorter amount of time than for Gaussian-driven processes. This property has its effect on the probability of finding large excursions $P\{L_{\alpha}(t) > r\}$, which has 'heavy tails', i.e.

$$P\{L_{\alpha}(t) > r\} \sim r^{-\alpha}.$$
(3.3)

A formalism allowing the existence of heavy tails in the distribution of stochastic forcing allows a much broader class of observed phenomena to be modelled as SDEs of the form

$$d\boldsymbol{x} = \boldsymbol{F}(\boldsymbol{x}, s)ds + \boldsymbol{\mathsf{G}}(\boldsymbol{x}, s)(\circ)d\boldsymbol{L}.$$
(3.4)

(N.B. the Itô–Stratonovich quandary persists for $\alpha < 2$.) Indeed, non-Gaussian Lévy models have been applied to atmospheric turbulence (e.g. Viecelli 1998; Ditlevsen 2004) and palaeoclimate (e.g. Ditlevsen 1999), and have been used to explain anomalous diffusion observed in hydrological records (Hurst 1951).

C. Penland and B. D. Ewald

Just as the pdfs of Markov diffusion processes obey a Fokker–Planck equation, so do those of systems involving non-Gaussian stable Lévy processes. However, these Fokker–Planck equations involve fractional derivatives of order α rather than second derivatives with respect to the state variable in the diffusion term (e.g. Chechkin *et al.* 2003; Ditlevsen 2004). It is often more convenient to consider a spectral form of the Fokker–Planck equation so that the fractional derivatives are replaced by conjugate variables raised to the power α . Not surprisingly, much of the theory of Lévy α -stable processes also involves Fourier and Laplace transforms, as we see in the next sections. For clarity, we shall confine our discussion to univariate variables unless the multivariate generalization is clear.

(b) Poisson processes and compound Poisson processes

This subsection is intended to give the reader an intuitive feel for Lévy processes. To do so, it is extremely useful to employ the fact that Lévy processes can be written in terms of a 'Lévy–Itô decomposition', which is a combination of a drift, a Brownian motion and compound Poisson processes (Appelbaum 2004). We discussed Brownian motions in detail in §2; we now turn our attention to Poisson and compound Poisson processes.

Consider a rabbit standing (or sitting) at a place we shall designate the origin. After some time t, the bunny may or may not have taken one or more jumps. We shall denote the number of jumps taken by our lapine friend in time t as N(t). Using Protter's (2005) formalism, let us say the rabbit jumps for the nth time at time T_n , with $T_0=0$ and $T_{n+1} > T_n$. The counting process N(t) can be written in terms of the indicator function $1_{\{t \ge T_n\}}$, which is one for $t \ge T_n$ and zero for $t < T_n$, i.e.

$$N(t) = \sum_{n \ge 1} \mathbb{1}_{\{t \ge T_n\}}.$$
(3.5)

If N(t) obeys a Poisson process, then the probability that N(t) equals some non-negative integer n is

$$P\{N(t) = n\} = \frac{(\lambda t)^n \exp(-\lambda t)}{n!},$$
(3.6)

for some $\lambda > 0$. That is, λt is the parameter of the Poisson process and λ is called the *arrival intensity*. The Poisson process has some well-known properties. For example, its mean and variance are time dependent and each is equal to λt . Note that this probability is not only continuous for t>0 but also differentiable with respect to time,

$$\frac{\mathrm{d}P\{N(t) = n\}}{\mathrm{d}t} = \lambda P\{N(t) = n - 1\} - \lambda P\{N(t) = n\}.$$
(3.7)

A Poisson process from which the mean λt is subtracted is called a *compensated Poisson process*, and has mean zero. Note further that we confine ourselves to finite λ .

So far, we have not said anything about how big these jumps are. Let us say the *k*th jump has length Y_k . Then, the distribution of the sum $Y_1 + Y_2 + \cdots + Y_{N(t)}$, with N(t) a Poisson process, is a *compound Poisson process* (Feller 1966; Protter 2005). The compound Poisson processes as just described are Lévy processes.

It is possible to combine the compound Poisson process and Brownian motion into yet another Lévy process: let the combination of a Brownian motion W(t)and a linear drift bt be denoted as C(t), and a compound Poisson process as just described be denoted as Y(t). Then, we may define yet another Lévy process X(t)(Appelbaum 2004) as

$$X(t) = C(t) + Y(t).$$
 (3.8)

This means that X(t) is simply the Brownian motion with drift until the first jump, say, at time T_1 and height Y_1 . After that, X(t) evolves again as a Brownian motion with drift until the next jump, and so on. That is,

$$X(t) = \begin{cases} C(t) & \text{for } 0 \le t < T_1, \\ C(T_1) + Y_1 & \text{for } t = T_1, \\ X(T_1) + C(t) - C(T_1) & \text{for } T_1 < t < T_2, \\ X(T_1) + C(T_2) - C(T_1) + Y_2 & \text{for } t = T_2, \text{ etc.} \end{cases}$$
(3.9)

Equation (3.9) is an example of a Lévy process represented as the sum of a drift, a Brownian motion and superpositions of compound Poisson processes (Bhattacharya & Waymire 1990; Appelbaum 2004; Eliazar & Klafter 2003). This is what is meant by the qualitative description of Lévy forcing as white noise with jumps, particularly since T_1 and T_2 may be as close together as we like.

(c) Characteristics of Lévy α -stable processes

We do not use the term 'characteristic' lightly. In fact, Lévy α -stable processes are generally defined in terms of their characteristic functions $\phi(u) = \langle \exp(iuL_{\alpha}) \rangle$ (e.g. Weron & Weron 1995; Eliazar & Klafter 2003; Dybiec *et al.* 2006; Nolan 2007),

$$\ln \phi(u) = -\sigma^{\alpha} |u|^{\alpha} \left\{ 1 - \mathrm{i}\beta \operatorname{sgn}(u) \tan\left(\frac{\pi\alpha}{2}\right) \right\} + \mathrm{i}\mu u, \quad \alpha \neq 1,$$
(3.10*a*)

$$\ln \phi(u) = -\sigma^{\alpha} |u|^{\alpha} \left\{ 1 + \mathrm{i}\beta \operatorname{sgn}(u) \frac{2}{\pi} \ln |u| \right\} + \mathrm{i}\mu u, \quad \alpha = 1.$$
(3.10b)

In (3.10*a*) and (3.10*b*), the location parameter μ shifts the distribution to the l.h.s. or r.h.s. For $\alpha=2$, μ represents the Gaussian mean. The scale parameter σ , which is greater than zero, represents the width of the distribution about μ . In this notation, the variance of a Gaussian distribution is $2\sigma^2$. This convention is somewhat inconvenient for those of us whose main experience is with Gaussian distributions, but it is the convention used in most of the Lévy process literature, so we may as well get used to it. The skewness parameter β lies in the interval [-1,1], and the distribution is symmetric when $\beta=0$. In fact, when both β and μ are zero, the characteristic function has the form

$$\phi(u) = \exp(-\sigma^{\alpha}|u|^{\alpha}). \tag{3.11}$$

For the Lévy process corresponding to a Brownian motion W(t), $\sigma^{\alpha} = \sigma^2 = t/2$. There are two other α -stable Lévy processes for which the pdf is known to have a closed-form solution. One is the Cauchy distribution, where $\alpha = 1$ and $\beta = 0$,

$$p(x) = \frac{1}{\pi} \frac{1}{\sigma^2 + (x - \mu)^2}.$$
(3.12)

The other is the Lévy distribution, where $\alpha = 0.5$ and $\beta = 1$,

$$p(x) = \sqrt{\frac{\sigma}{2\pi}} \frac{\exp\left(-\frac{\sigma}{2x}\right)}{x^{3/2}}.$$
(3.13)

For more general cases, one may evaluate the Fourier transform of (3.10a) and (3.10b) numerically.

(d) Additive α -stable Lévy-driven SDEs

Consider differential equations of the form

$$d\boldsymbol{x} = \boldsymbol{F}(\boldsymbol{x}, t)dt + \boldsymbol{\mathsf{G}}d\boldsymbol{L}_{\alpha}, \qquad (3.14)$$

where the coefficient **G** of the Lévy noise vector increment dL_{α} is simply a constant matrix. Although (3.14) is far from general, there are certainly plenty of applications for which it is applicable. In these cases, the standard numerical approach (e.g. Dybiec *et al.* 2006) is to use a Euler scheme

$$x_m(t_{i+1}) = x_m(t_i) + F_m(\boldsymbol{x}(t_i), t_i)\Delta + \sum_{\mu} G_{m\mu} \Delta^{1/\alpha} z_{\mu i}, \qquad (3.15)$$

where Δ is the time step and $\{z_{\mu i}\}$ are random variables sampled at time t_i from a centred α -stable Lévy distribution. Note that the time step in (3.15) is raised to the power $1/\alpha$. For $\alpha = 2$, (3.15) is equivalent to (2.18) with $\Delta^{1/2} z_{\mu i}$ representing the Gaussian random deviate of variance Δ .

Techniques for approximating random variables $\{z_{\mu i}\}\$ from α -stable Lévy distributions may be found in Dybiec *et al.* (2006) as well as Weron & Weron (1995). As of this writing, mathematician John Nolan, a professor at American University, has a website: http://academic2.american.edu/~jpnolan/stable/stable.html from which one may download Lévy-variable generators for use with proper acknowledgement. An example of an additive LSDE is given in §4.

4. Diffusion and α -stable Lévy processes: some examples

(a) Ornstein–Uhlenbeck processes

To compare the qualitative properties of a system driven by Gaussian white noise with that driven by an α -stable Lévy noise, we consider a simple Ornstein– Uhlenbeck (OU) process as follows:

$$\mathrm{d}x = -\gamma x \mathrm{d}t + \mathrm{d}\zeta. \tag{4.1}$$

In (4.1), ζ is either a symmetric α -stable Lévy variable with $\alpha = 1.5$ or Brownian motion ($\alpha = 2$). The decay parameter γ is equal to 0.1/tu and the scale parameter σ for each noise process is unity. Recall that $\sigma = 1$ for a Gaussian process is equivalent to a variance of 2. Equation (4.1) is integrated numerically using an Euler integration scheme ((3.15); see also Protter & Talay 1997) with a time step of $\Delta = 0.1$ tu. Random numbers in the interval (0,1] are generated using a Mersenne Twister (Matsumoto & Nishimura 1998), which are then used to generate either Gaussian random variables using a Box-Mueller scheme (Press et al. 1992) or Lévy variables (Weron & Weron 1995).

The time series of random variables used to integrate (4.1) are shown in figure 1. Note that the ordinate of the graph showing the Lévy process is an order of magnitude larger than that showing the Gaussian process, though both



Figure 1. Time series of random numbers used to numerically integrate (4.1). (a) Gaussian random variables and (b) Lévy-distributed variables.



Figure 2. First 10 tu of time series as shown in figure 1. Solid line, Gaussian random noise; dashed line, Lévy-distributed noise.

processes are characterized by the same scale parameter σ . The first 10 tu, comprising 100 time steps, of the time series are shown in figure 2. There it can be seen that the two time series are qualitatively similar, except in the region of the occasional jump.

The Gaussian-driven and the Lévy-driven OU processes are shown in figure 3a, b, respectively. Also shown (figure 4) is a close-up of the two OU processes; note the effect of the jump that occurred around 6 tu persists for approximately 15 time steps (1.5 tu) in the Lévy-driven system. The global effect of the jumps is seen in the cumulative distributions of the two OU processes (figure 5), where the fat tails of the α -stable Lévy distributed forcing are translated into fat tails of the Lévy-driven OU process.

2471



Figure 3. Time series of OU process forced by (a) Gaussian random noise and (b) Lévy-distributed noise.



Figure 4. First 10 tu of time series as shown in figure 3. Solid line, Gaussian random noise; dashed line, Lévy-distributed noise.

(b) Diffusion and Lévy processes with some similarities

Sura & Sardeshmukh (2008) have shown that daily departures of SSTs from the annual cycle are well described by the Stratonovich SDE

$$dT = \left(AT - \frac{1}{2}Eg\right)dt + b \circ dW_1 + (ET + g) \circ dW_2, \qquad (4.2)$$

where W_1 and W_2 are independent Wiener processes and A, E, g and b are constants. Sardeshmukh & Sura (submitted) similarly showed that daily



Figure 5. Cumulative probability distribution of OU process forced by Gaussian random noise (solid line) and Lévy-distributed noise (dashed line). Nonlinear abscissa yields a straight line for a Gaussian distribution.

departures of 300 mb vorticity from the annual cycle obey an equation of the same form. The marginal pdf corresponding to this diffusion process is

$$p(T) = \frac{1}{\Re} \left[(ET+g)^2 + b^2 \right]^{-(\gamma+1)} \exp\left[\frac{2\gamma g}{b} \arctan\left(\frac{ET+g}{b}\right) \right], \tag{4.3}$$

where \mathfrak{N} is the normalization constant. The quantity γ is defined as $-(A + 0.5E^2)/E^2$ and is positive for physically allowable systems, i.e. A is strictly negative and |A| is larger than $0.5E^2$. Thus, $0 < \gamma < \infty$ and, for large T_0 , $P\{T > T_0\} \sim T_0^{-(2\gamma+1)}$. Thus, as pointed out by Sardeshmukh & Sura (submitted), moments of order higher than $2\gamma + 1$ do not exist. The distribution of this diffusion process allows both skew and heavy tails, depending on the level of stochastic forcing and the level of correlation between the additive and multiplicative noises. The question arises, then, whether (4.3) is the pdf of a diffusion that is also an α -stable Lévy process. In general, the answer is 'No'. While (4.3) does converge to a Cauchy distribution for $\gamma \to 0$ and to a Gaussian for $\gamma \to \infty$, it is straightforward to show that the characteristic function of T(t) cannot be written in the form of (3.10*a*) or (3.10*b*) for finite $\gamma > 0$.

5. Conclusions

With the increasing popularity of stochastic parametrizations in numerical weather and climate models, it is necessary for us to remind ourselves of how the physical and mathematical bases for such approximations are related. In this article, we have tried to summarize some of this theory without pretending to be exhaustive. In particular, we have not discussed the newly developed theory of integrating SDEs driven by state-dependent Lévy processes. The procedures for doing so involve what are called 'Marcus integrals', and are beyond the scope of this article. The most accessible reference we have found on the subject is in the book by Appelbaum (2004).

Even the numerical simulation of Gaussian-driven processes is not trivial, but the theory for these systems is quite well developed. Obviously, it is not sufficient for the mere existence of power law tails or skew in a probability distribution to preclude a valid approximation of a physical system as forced by Gaussian white noise (see also Sardeshmukh & Sura submitted). Thus, unless there are physical reasons for doubting the validity of such an approximation, we recommend using the theoretical arsenal developed around classical SDEs if at all possible.

We conclude with a reminder that the difference between Itô and Stratonovich calculi is important. For mathematicians, it may suffice to know that there is a transformation between them; scientists have to know what that transformation is and how to evaluate it. A thermometer gives only one reading at a time, and knowing that an isomorphism exists between the output of a numerical weather prediction model and the temperature that will eventually be observed does not help the farmer or fisherman unless scientists can apply that isomorphism in advance. Of course, if the stochastic forcing in a numerical model can be argued on physical grounds to exist independently of the system being modelled, the issue of multiple calculi becomes moot and one may use the Euler scheme to model an SDE driven by an α -stable Lévy random variable, whether or not $\alpha = 2$. The key phrase here involves argument on physical grounds; concern for the fidelity of any type of parametrization to the physical system being modelled is always our first priority.

The authors are pleased to acknowledge valuable conversations with P. D. Sardeshmukh, P. Imkeller, T. Hamill, M. Charnotskii, P. Sura and I. Pavlyukevich. Particular gratitude is expressed to an anonymous reviewer for the extremely valuable advice.

References

An, S.-I. & Jin, F.-F. 2004 Nonlinearity and asymmetry of ENSO. J. Clim. 17, 2399–2412. (doi:10. 1175/1520-0442(2004)017<2399:NAAOE>2.0.CO;2)

Appelbaum, D. 2004 Lévy processes and stochastic calculus, vol. 93. Cambridge studies in advanced mathematics. Cambridge, UK: Cambridge University Press.

Arnold, L. 1974 Stochastic differential equations: theory and applications. New York, NY: Wiley.

Berner, J., Shutts, G. J., Leutbecher, M. & Palmer, T. N. Submitted. A spectral kinetic energy backscatter scheme and its impact on flow-dependent predictability in the ECMWF ensemble prediction system.

Bhattacharya, R. N. & Waymire, E. C. 1990 Stochastic processes with applications. New York, NY: Wiley.

2474

- Buizza, R., Miller, M. & Palmer, T. N. 1999 Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. Q. J. R. Meteorol. Soc. 125, 2887–2908. (doi:10.1256/ smsqj.56005)
- Chechkin, A. V., Klafter, J., Gonchar, V. Yu., Metzler, R. & Tanatarov, L. V. 2003 Bifurcation, bimodality, and finite variance in confined Lévy flights. *Phys. Rev. E* 67, 010 102. (doi:10.1103/ PhysRevE.67.010102)
- Dybiec, B., Gudowska-Nowak, E. & Hanggi, R. 2006 Lévy-Brownian motion on finite intervals: mean first passage time analysis. *Phys. Rev. E* 73, 046 104. (doi:10.1103/PhysRevE.73.046104)
- Ditlevsen, P. D. 1999 Observation of α-stable noise induced millenial climate changes from an icecore record. *Geophys. Res. Lett.* **26**, 1441–1444. (doi:10.1029/1999GL900252)
- Ditlevsen, P. D. 2004 *Turbulence and climate dynamics*. Copenhagen, Denmark: J. & R. Frydenberg A/S.
- Doob, J. L. 1953 Stochastic processes. New York, NY: Wiley.
- Dozier, L. B. & Tappert, F. D. 1978a Statistics of normal mode amplitudes in a random ocean. 1. Theory. J. Acoust. Soc. Am. 63, 353–365. (doi:10.1121/1.381746)
- Dozier, L. B. & Tappert, F. D. 1978b Statistics of normal mode amplitudes in a random ocean. 2. Computations. J. Acoust. Soc. Am. 64, 533–547. (doi:10.1121/1.382005)
- Eliazar, I. & Klafter, J. 2003 Lévy-driven Langevin systems: targeted stochasticity. J. Stat. Phys. 111, 739–767. (doi:10.1023/A:1022894030773)
- Ewald, B. D. & Penland, C. 2008 Numerical generation of stochastic differential equations in climate models. In *Handbook of numerical analysis: computational methods for the atmosphere and the oceans* (eds R. Temam & J. Tribbia). Amsterdam, The Netherlands: Elsevier.
- Ewald, B. D. & Temam, R. 2003 Analysis of stochastic numerical schemes for the evolution equations of geophysics. Appl. Math. Lett. 16, 1223–1229. (doi:10.1016/S0893-9659(03)90121-2)
- Ewald, B. D. & Temam, R. 2005 Numerical analysis of stochastic schemes in geophysics. SIAM J. Numer. Anal. 42, 2257–2276. (doi:10.1137/S0036142902418333)
- Ewald, B. D., Penland, C. & Temam, R. 2004 Accurate integration of stochastic climate models. Mon. Weather Rev. 132, 154–164. (doi:10.1175/1520-0493(2004)132<0154:AIOSCM>2.0.CO;2)
- Feller, W. 1966 An introduction to probability theory and its applications, vol. II. New York, NY: Wiley.
- Fjørtoft, R. 1953 On the changes in the spectral distribution of kinetic energy for two-dimensional nondivergent flow. *Tellus* 5, 225–230.
- Flügel, M., Chang, P. & Penland, C. 2004 The role of stochastic forcing in modulating ENSO predictability. J. Clim. 17, 3125–3140. (doi:10.1175/1520-0442(2004)017<3125:TROSFI>2.0.CO;2)
- Gardiner, C. W. 1985 Handbook of stochastic methods. Berlin, Germany: Springer.
- Gnendenko, B. V. & Kolmogorov, A. N. 1954 Limit distributions for sums of independent random variables. Reading, UK: Addison-Wesley.
- Hasselmann, K. 1976 Stochastic climate models. 1. Theory. Tellus 28, 473-485.
- Horsthemke, W. & Lefever, R. 1984 Noise-induced transitions: theory and applications in physics, chemistry, and biology. Berlin, Germany: Springer.
- Hurst, H. 1951 Long-term storage capacity of reservoirs. Proc. Inst. Civil Eng. 5, 519–577.
- Khasminskii, R. Z. 1966 A limit theorem for solutions of differential equations with random righthand side. *Theory Prob. Appl.* 11, 390–406. (doi:10.1137/1111038)
- Kloeden, P. E. & Platen, E. 1992 Numerical solution of stochastic differential equations. Berlin, Germany: Springer.
- Kohler, W. & Papanicolaou, F. C. 1977 Wave propagation in a randomly inhomogeneous ocean. In Wave propagation and underwater acoustics, vol. 70 (eds J. B. Keller & J. S. Papadakis). Lecture Notes in Physics, ch. IV, pp. 153–223. Berlin, Germany: Springer.
- Lévy, P. 1937 Theorie de l'addition des variables aleatoires. Paris, France: Gauthier-Villars.
- Majda, A. J., Timofeyev, I. & Vanden Eijnden, E. 1999 Models for stochastic climate prediction. Proc. Natl Acad. Sci. USA 96, 14 687–14 691. (doi:10.1073/pnas.96.26.14687)
- Matsumoto, M. & Nishimura, T. 1998 Mersenne twister: a 623-dimensionally equidistributed uniform pseudorandom number generator. ACM Trans. Model. Comput. Simul. 8, 3–30. (doi:10.1145/272991.272995)

C. Penland and B. D. Ewald

- Müller, D. 1987 Bispectra of sea surface temperature anomalies. J. Phys. Oceanogr. 17, 26–36. (doi:10.1175/1520-0485(1987)017<0026:BOSSTA>2.0.CO;2)
- Nolan, J. P. 2007 Stable distributions—models for heavy tailed data. Boston, MA: Birkhauser.
- Papanicolaou, G. C. & Kohler, W. 1974 Asymptotic theory of mixing stochastic differential equations. Commun. Pure Appl. Math. 27, 641–668.
- Penland, C. 1985 Acoustic normal mode propagation through a three-dimensional internal wave field. J. Acoust. Soc. Am. 78, 1356-1365. (doi:10.1121/1.392906)
- Penland, C. 1989 Random forcing and forecasting using principal oscillation pattern analysis. Mon. Weather Rev. 117, 2165–2185. (doi:10.1175/1520-0493(1989)117<2165:RFAFUP>2.0.CO;2)
- Penland, C. 1996 A stochastic model of IndoPacific sea surface temperature anomalies. *Physica D* 98, 534–558. (doi:10.1016/0167-2789(96)00124-8)
- Penland, C. & Sardeshmukh, P. D. 1995 The optimal growth of tropical sea surface temperature anomalies. J. Clim. 8, 1999–2024. (doi:10.1175/1520-0442(1995)008<1999:TOGOTS>2.0.CO;2)
- Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. 1992 Numerical recipes in Fortran: the art of scientific computing. Cambridge, UK: Cambridge University Press.
- Protter, P. 2005 Stochastic integration and differential equations. Berlin, Germany: Springer.
- Protter, P. & Talay, D. 1997 The Euler scheme for Lévy-driven stochastic differential equations. Ann. Prob. 25, 393–423. (doi:10.1214/aop/1024404293)
- Purcell, E. J. 1972 Calculus with analytic geometry. New York, NY: Meredith Corporation.
- Rümelin, W. 1982 Numerical treatment of stochastic differential equations. SIAM J. Numer. Anal. 19, 605–613.
- Saha, S. et al. 2006 The NCEP climate forecast system. J. Clim. 19, 3483–3517. (doi:10.1175/ JCLI3812.1)
- Sardeshmukh, P. D. & Hoskins, B. J. 1988 The generation of global rotational flow by steady idealized tropical divergence. J. Atmos. Sci. 45, 1228–1251. (doi:10.1175/1520-0469(1988)045 < 1228:TGOGRF>2.0.CO;2)
- Sardeshmukh, P. D. & Sura, P. Submitted. Reconciling non-Gaussian climate statistics with linear dynamics.
- Sardeshmukh, P. D., Penland, C. & Newman, M. 2001 Rossby waves in a fluctuating medium. In Stochastic climate models, vol. 49 (eds P. Imkeller & J.-S. von Storch). Progress in probability. Basel, Switzerland: Birkhaueser.
- Shutts, G. 2005 Kinetic energy backscatter for NWP models and its calibration. In Proc. ECMWF workshop on representation of subscale processes using stochastic-dynamic models, Reading, UK, 6-8 June 2005.
- Sura, P. & Penland, C. 2002 Sensitivity of a double-gyre ocean model to details of stochastic forcing. Ocean Modelling 4, 327–345. (doi:10.1016/S1463-5003(02)00008-2)
- Sura, P. & Sardeshmukh, P. D. 2008 A global view of non-Gaussian SST variability. J. Phys. Oceanogr. 38, 639–647. (doi:10.1175/2007JPO3761.1)
- Viecelli, J. A. 1998 On the possibility of singular low-frequency spectra and Lévy law persistence in the planetary-scale turbulent circulation. J. Atmos. Sci. 55, 677–687. (doi:10.1175/1520-0469(1998)055<0677:OTPOSL>2.0.CO;2)
- Weron, A. & Weron, R. 1995 Computer simulation of Lévy α-stable variables and processes. Lecture Notes in Physics, no. 457, pp. 379–392. Berlin, Germany: Springer.
- Wilks, D. S. 1995 Statistical methods in the atmospheric sciences. San Diego, CA: Academic Press.
- Winkler, C. R., Newman, M. & Sardeshmukh, P. D. 2001 A linear model of wintertime lowfrequency variability. Part I: formulation and forecast skill. J. Clim. 14, 4474–4494. (doi:10. 1175/1520-0442(2001)014<4474:ALMOWL>2.0.CO;2)
- Wong, E. & Zakai, M. 1965 On the convergence of ordinary integrals to stochastic integrals. Ann. Math. Stat. 36, 1560–1564. (doi:10.1214/aoms/1177699916)
- Zebiak, S. & Cane, M. 1986 A model El Niño-Southern Oscillation. Mon. Weather Rev. 115, 2262–2278. (doi:10.1175/1520-0493(1987)115<2262:AMENO>2.0.CO;2)