Earth and Space Science

Mandelbrot's Stochastic Time Series Models

N. W. Watkins[1,2,3]

[1]Centre for the Analysis of Time Series, London School of Economics and Political Science, London, UK, [2]Faculty of Science Technology Engineering and Mathematics, Open University, Milton Keynes, UK, [3]Centre for Fusion, Space and Astrophysics, University of Warwick, Coventry, UK

Abstract I survey and illustrate the main time series models that Mandelbrot introduced into time series analysis in the 1960s and 1970s. I focus particularly on the members of the additive fractional stable family including Lévy flights and fractional Brownian motion (fBm), noting some of the less well-known aspects of this family, such as the cases when the self-similarity exponent H and the Hurst exponent J differ. I briefly discuss the role of multiplicative models in modeling the physics of cascades. I then recount the still little-known story of Mandelbrot's work on fractional renewal models in the late 1960s, explaining how these differ from their more familiar fBm counterpart and form a "missing link" between fBm and the problem of random change points. I conclude by highlighting the frontier problem of damped fractional models.

1. Mandelbrot and the Geosciences

Benoit Mandelbrot, the "father of fractals" (Giles, 2004) was the first person to deliver the AGU's Lorenz Lecture in 2001. His title, "Fractals as a measure of roughness in Earth sciences," captures the essential unifying role fractality has played in the geosciences; allowing us to describe nature's roughness realistically by expanding the notion of dimensionality to fractional cases (Feder, 1988; Mandelbrot, 1982). Fractals, which he did so much to communicate, develop, and apply, have influenced nonlinear geophysics and, more broadly, all the Earth and space sciences in many ways. Their diverse impact on nonlinear geophysics alone can be appreciated just by noting the breadth of the research topics of many of the subsequent Lorenz lectures in which fractals were featured, including climate (Lovejoy & Schertzer, 2013), hydrology (Koutsoyiannis, 2013; Kondrashov et al., 2005), meteorology and atmospheric science (Osso et al., 2010), solid earth geophysics (Mandal et al., 2005; Rundle & Klein, 1993; Sornette, 2003; Turcotte, 1997), and space plasma physics (Sharma, 1995).

My own short article in this special issue cannot span this breadth, but instead has a narrower aim: to extract the essence of Mandelbrot's cross-disciplinary contribution to the modeling of stochastic time series, drawing on several previous surveys of this aspect of the history of complexity science (Graves et al., 2017; Watkins, 2013, 2017a, 2017b; Watkins et al., 2016).

I will focus on several effects that Mandelbrot drew attention to and sought to model: heavy-tailed amplitude distributions (the Noah effect) in section 2.2; long-range temporal dependence (the Joseph effect), in section 2.3; their combination in section 2.4; multiplicative processes and multifractals in section 3; and heavy-tailed waiting times in section 4. I will conclude in section 5 by examining models which combine long-range dependence (LRD) with the mean reversion necessary to describe the dissipation of energy, thus uniting fractals with more familiar tools like the archetypal Markovian short-ranged memory model, the AutoRegressive process of order 1, $AR(1)$ (Brockwell & Davis, 2016).

To put the models in their motivating context I will now briefly summarize the types of roughness in geophysical phenomena, which inspired fractal and multifractal modeling. They can be organized into four overlapping threads:

1. *Roughness in real spatial patterns*: The fractal *morphology* of natural forms such as clouds, coastlines, and riverbeds (Feder, 1988; Mandelbrot, 1982, 2002; Turcotte, 1997). This research has not only given quantitative answers to Richardson's question "how long is the coast of Britain" but also led to energy optimization principles for evolving riverbeds (Rodríguez-Iturbe & Rinaldo, 1997). In most cases these

forms are slowly evolving compared to the sampling time scale, and so the focus, initially at least, has been on their static properties or on quasi-static snapshots of their evolution, even in the pioneering study of the fractal shapes of clouds by Lovejoy (1982).

2. *Roughness in abstract spatial patterns*: The objects traced out over time by low dimensional *dynamical systems* as they explore phase space (Mandelbrot, 2004). A famous example is the relevance to meteorological forecasting of the noninteger dimension of attractors (Baker & Gollub, 1990) in deterministic dissipative chaotic systems such as the model that (Lorenz, 1963) abstracted from atmospheric circulation. The complex structure of the attractor means that very closely separated initial conditions will diverge rapidly. Another example is the role of fractal structures in the phase space of nondissipative, Hamiltonian chaos (Zaslavsky, 2002). These create regions that can temporarily trap particles, giving rise to anomalously slow ("sub") diffusion when compared to ordinary Brownian motion, with applications including transport in oceanography and laboratory fluid experiments (Klafter et al., 1996). This research thread has given rise both to models of how the fractal dynamical structures arise and to models of how diffusion of test particles is affected by such structures, including the heavy-tailed waiting times seen in the modified random walks (Kutner & Masoliver, 2017; Watkins, 2017b) discussed in section 4.

3. *Roughness in noisy time series*: The use of additive *fractal stochastic processes* to describe time series (Mandelbrot, 2002) drawn from natural systems, such as the water level of the Nile, which exhibit "anomalous" properties, such as the Hurst effect, when compared to Brownian motion. Characteristically, Mandelbrot's eye for similarities in geometry enabled him to use the same fractional Gaussian and Brownian models to describe aspects of both quasi-static coastlines and statistically stationary time series. In section 2 I will describe the key properties of these models.

4. *Roughness in fluid turbulence*: The use of stochastic multifractal models (Mandelbrot, 2004; Schmitt & Huang, 2016), which have a continuous range of scaling exponents rather than just one, as *models of the intermittency* of time series (Mandelbrot, 1999) such as those arising from energy dissipation on the complex spatiotemporal structures observed in fluid turbulence. Multiplicative models are the subject of section 3.

It is important to recognize that Mandelbrot's research was intrinsically interdisciplinary and that he was continuously borrowing tools and observations from one area to illuminate aspects of another. Some of the models I describe thus owe more to his pioneering investigations in describing financial markets (Mandelbrot, 1997; Mandelbrot & Hudson, 1998) than to the geosciences.

As well as the intrinsic interest of fractals in mathematics education (e.g., Falconer, 2003), the static and dynamic models of fractal structures used in the first two threads were quite rapidly integrated into the curriculum of physicists and geoscientists, at least at the graduate level. They have thus long been served by dedicated surveys (e.g., Baker & Gollub, 1990; Rodríguez-Iturbe & Rinaldo, 1997; Turcotte, 1997) and reprint selections (e.g., Cvitanovic, 1984).

In contrast, it has been my experience that even 50 years on, the geoscience community is less clear about *why* stochastic fractals and multifractal time series models are central in the last two threads, and why heavy-tailed waiting times appear in the anomalous diffusion models in the second thread. This is despite the wealth of excellent geophysically relevant tutorial material, which has appeared since the late 1980s (Feder, 1988), and continues to develop in sophistication (Lovejoy & Schertzer, 2013; Turcotte, 1997; Schmitt & Huang, 2016; Sornette, 2003), supplemented by workshops, short courses at meetings such as The European Geosciences Union, and so on.

I think three factors may have contributed to this: (i) The underlying physics may seem less directly coupled to the measured fractal properties, and thus less easily pictured; (ii) it is hard to disentangle the proposed underlying physical origin of fractals from the fractal properties themselves and the measurement algorithms used to quantify them; and (iii) there is also the notoriety (Watkins et al., 2016) of the most ambitious proposed "universal" mechanism for spatiotemporal fractals, self-organized criticality (Bak & Chen, 1989; Watkins et al., 2016). I think these are all genuine problems and have struggled with them myself. They relate to active research frontiers and are far from fully settled.

I do not, unfortunately, believe that I can solve all these problems by means of this article. What I have tried to do instead, as my contribution to this AGU anniversary special issue, is seek to clarify how Mandelbrot's various theoretical models differ from each other. I thus give a bird's eye view of the key models he introduced
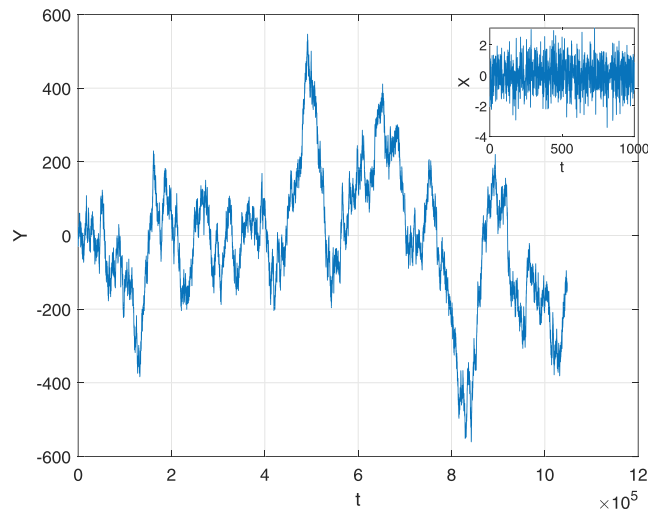
Figure 1. A time series of discrete Brownian motion Y_t formed by summing (inset) discrete white $N(0,1)$ Gaussian noise X_t.

or popularized and a very brief sketch of a new direction, fractal mean-reverting models, to which they have now led.

2. Noises With Random Phases and Additive Monofractal Walks

I first consider the simplest stochastic models, which are derived by modifying random phase white noise, and Brownian motion. They are called additive because randomness is added to the time series model at each point along its time evolution. For each case considered I will refer to the continuous stationary noise process $X(t)$, the continuous nonstationary motion $Y(t)$ formed by integrating X, and the corresponding discrete processes X_t, Y_t produced by sampling X and Y at discrete time intervals $t, t+\Delta t, t+2\Delta t$, etc.

2.1. The iid Gaussian Family

Stationary white (d=0) *Gaussian* (α=2) *noise*: Shown in the inset of Figure 1 is the simplest case, an independent and identically distributed (iid), sequence X_t of discrete white Gaussian noise. This, and all other simulations shown in Figures 1 to 8, was produced using the `fftlfsn` algorithm of Stoev and Taqqu (2004). In all simulations shown Δt is set to 1 and 1,047,976 points have been generated. Here the probability distribution function (pdf) of amplitudes is Gaussian,

$$p(X) = \frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/2\sigma^2} \tag{1}$$

and is thus specified entirely by its first two moments, the mean μ and the variance σ^2, which are constant in such a stationary time series. The absence of time correlations means that fluctuations are very short-lived, quantified by the autocorrelation function (ACF)

$$R(\tau) = <X(t)X(t+\tau)> = \sigma^2\delta(\tau) \tag{2}$$

which is a delta function. The theoretical power spectral density (PSD) is thus a constant

$$S(f) = \int d\tau R(\tau)\exp(2\pi i f \tau) = \sigma^2 \tag{3}$$

and flat, with no excess at high or low frequencies, giving the name "white" noise. Inspection of the inset to Figure 1 shows that its values nearly all lie within $\pm 3\sigma$ as expected for the $N(0,1)$ Gaussian distribution. The Fourier transform $P(k)$ of the Gaussian pdf is itself a Gaussian function of k, $P(k) \sim \exp(-k^2)$, and is known as the characteristic function.

Self-similar H=1/2 *Brownian motion*: Integrated white noise $X(t)$ is Brownian motion $Y(t)$, for which a discrete sequence Y_t is shown in the main panel of Figure 1. This is usually described as self-similar (although Mandelbrot preferred the more precise term self-affine, see p. 157 of Mandelbrot, 2002 and p. ix of Embrechts
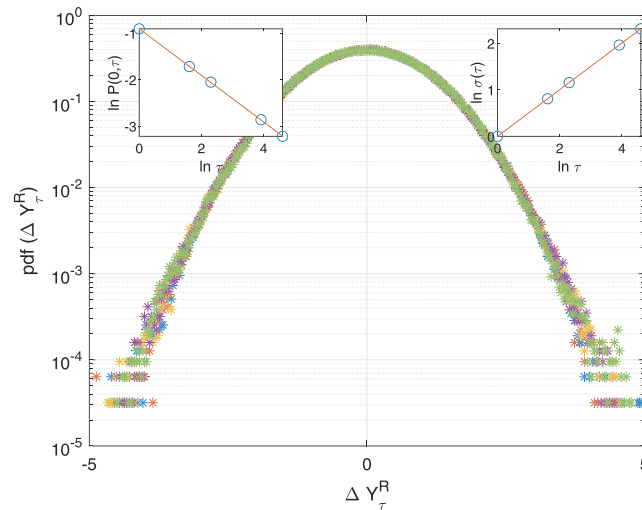
Figure 2. For the Brownian motion $Y(t)$ in Figure 1, this figure shows scaling collapse of the probability distribution functions (pdfs) $P(\Delta Y^R, \tau)$ of $\Delta Y^R(\tau) = (\Delta Y - \hat{\mu}(\Delta Y))/\hat{\sigma}(\Delta Y)$, where the differencing lags were $\tau=1,5,10,50$, and 100. The left inset shows $P(0,\tau)$ versus τ and the right inset shows sample standard deviation $\hat{\sigma}$ versus τ.

& Maejima, 2003), a property quantified by the self-similarity or Holder exponent H. This describes how such a nonstationary motion traces out a curve which is in a precise sense a statistically rescalable copy of itself as the length of the time series increases. Embrechts and Maejima (2003) give rigorous definitions of self-similarity and H: (i) An R^d-valued stochastic process is said to be "self-similar" with exponent H if for any $a>0$, there exists $b>0$ such that $\{X(at)\}=^d\{bX(t)\}$; (ii) if $\{X(t), t \geq 0\}$ is nontrivial, stochastically continuous at $t=0$ and self-similar, then there exists a unique $H \geq 0$ such that b can be expressed as a^H. In the case of Brownian motion H is 1/2.

To illustrate the self-similarity of the discrete sample of Brownian motion from Figure 1 one first calculates the pdf $P(\Delta Y, \tau)$ of $\Delta Y(\tau)$ differenced at lags τ of 1,5, 10, 50, and 100. The left inset of Figure 2 shows that the peak of this pdf, $P(0,\tau)$ is a decaying function of lag τ. The scaling exponent of this power law decay is $-H$ and found to be -0.5017, again very close to the theoretical value of $-1/2$ (Chapman et al., 2005). If one now calculates a rescaled $\Delta^R(\tau)$ by first subtracting the sample means $\hat{\mu}$ of the set of $\Delta Y(\tau)$ for a given lag τ and then dividing by the corresponding sample standard deviation $\sigma(\Delta Y(\tau))$, one can plot the pdfs of $\Delta Y(\tau)$, which collapse onto each other, as seen in the main panel of Figure 2.
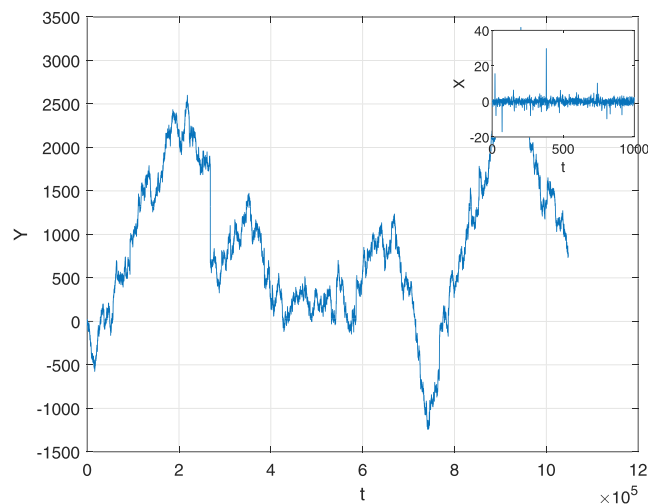

Figure 3. Stable motion Y_t with $\alpha=1.8$ and (inset) stable noise X_t. Note the infrequent but large "spiky" excursions in X_t and the consequent abrupt steps in Y_t.
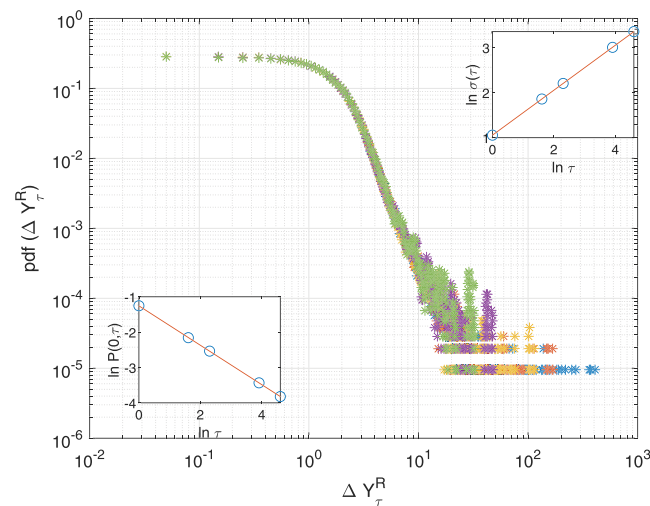
Figure 4. The main panel shows rescaled pdfs $P(\Delta Y^R, \tau)$ of $\Delta Y^R(\tau) = (\Delta Y - \hat{\mu}(\Delta Y))/(\Delta Y)^{1/\alpha}$, for positive ΔY^R, where the differencing lags were τ=1,5,10,50, and 100 for an α-stable motion with α=1.8. The left inset shows $P(0,\tau)$ versus τ and the right inset shows sample standard deviation $\hat{\sigma}$ versus τ.

There is a second important exponent, for which Mandelbrot advocated the term Joseph exponent J (p. 157 of Mandelbrot, 2002). The best known context in which J appears is as the exponent measured by Mandelbrot's diagnostic for LRD ("The Joseph effect") in time series, the rescaled range or R/S method originally employed by Hurst, which we will describe in more detail in section 2.3. J appears in various other places, one of which is as the exponent of the power law which quantifies the rate of growth of the sample standard deviation of the differences of Brownian motion $\Delta Y(\tau)=Y_{t+\tau}-Y_t$ as a function of the differencing scale τ, that is, $\hat{\sigma}(\Delta Y(\tau)) \sim \tau^J$, For the sample series of Figure 1 this is plotted in the inset of Figure 2, and J is found to be +0.5011, again very close to its theoretical value of 1/2 (Mandelbrot, 2002).

A great deal of confusion has resulted from the fact that in the original Gaussian processes where H and J were first studied, such as the fractional Brownian motion of section 2.3, variances are well behaved and H and J always take the same values as each other. Understandably, after its early coinage by Mandelbrot in the Gaussian context, the term Hurst exponent has been used by various authors to refer to either H or J, adding to the continuing confusion; indeed, Mandelbrot (2002) ruefully noted that "the letter H became overloaded with a host of conflicting meanings." We will thus follow his recommendation that " … in the general R/S context it is best to replace 'the old H' by the letter J, called [the] 'Joseph exponent'" and will not
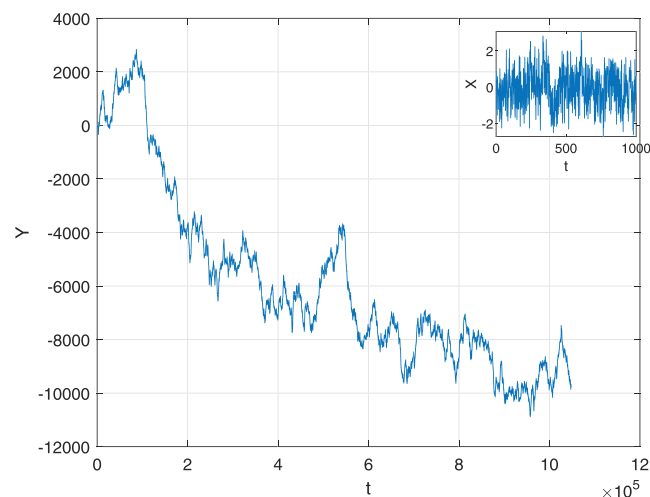
Figure 5. Simulated series of (inset) long-range dependent discrete fractional Gaussian noise X_t with d=0.3, and (main panel) fractional Brownian motion with H=1/2+d=0.5+0.3=0.8.
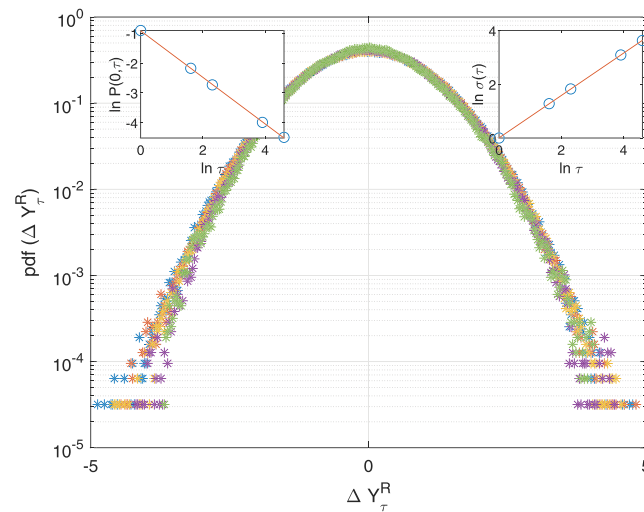
Figure 6. For the fractional Brownian motion $Y(t)$ in Figure 5, this figure shows scaling collapse of the probability distribution functions (pdfs) $P(\Delta Y^R, \tau)$ of $\Delta Y^R(\tau) = (\Delta Y - \hat{\mu}(\Delta Y))/\tau^{1/H}$, where the differencing lags were τ=1,5,10,50, and 100. The left inset shows $P(0, \tau)$ versus τ and the right inset shows sample standard deviation $\hat{\sigma}(\Delta Y)$ versus τ, with exponents close to -0.8 and $+0.8$, respectively.

use the term Hurst exponent further. We will illustrate this admittedly subtle, but nonetheless important, point in sections 2.2 and 2.4 by showing examples of processes where $H \neq J$.

The stationary (white noise) increments of self-similar classical Brownian motion form the reference against which Mandelbrot defined two kinds of anomalous behavior, which Mandelbrot and Wallis (1968) named the Noah effect (heavy tailed pdfs) and the Joseph effect (LRD in time). We will now examine these two effects, allowing first the pdf and then the ACF to have decaying power law tails, and then combining both effects.

2.2. The Noah Effect and the iid α-Stable Family: "Lévy Flights"

Stationary white d=0, stable $\alpha \neq 2$, noise: If one relaxes the intuitive assumption that the theoretical variance of the random process must be finite, a broader class of "α-stable" probability distributions is generated by generalizing the characteristic function to a stretched or compressed exponential $\sim \exp(-k^\alpha)$ with α lying in the range $0 \leq \alpha < 2$. Unlike the Gaussian, most of these distributions do not have a known analytic form, with rare exceptions like the α=1 (Cauchy) distribution (Sornette, 2003). However asymptotically, as
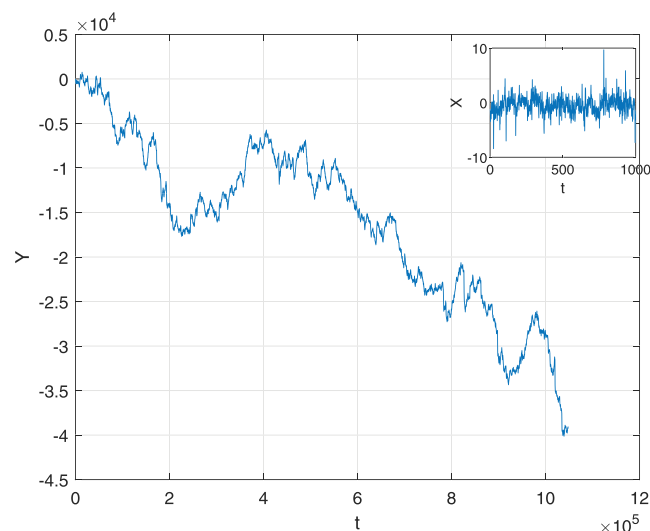
Figure 7. Simulated series of (inset) long-range dependent and heavy tailed discrete fractional stable noise X_t with d=0.3 and α=1.8, and (main panel) linear fractional stable motion Y_t with H=$1/\alpha$+d=$1/1.8$+0.3=0.8556.
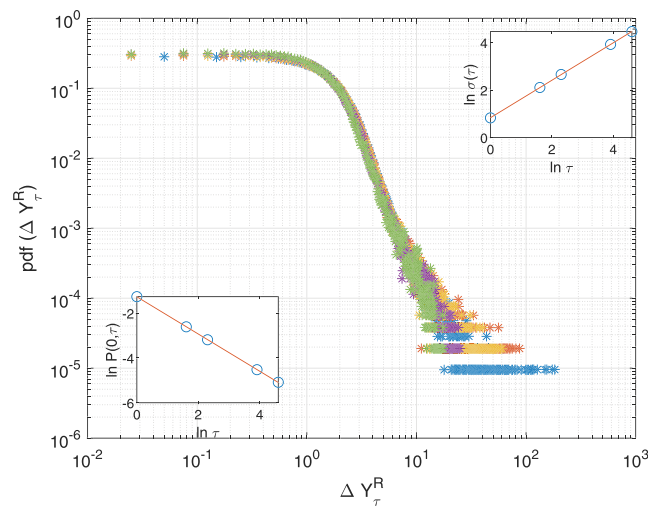
Figure 8. The main panel shows rescaled probability distribution functions (pdfs) $P(\Delta Y^R, \tau)$ of $\Delta Y^R(\tau) = (\Delta Y - \hat{\mu}(\Delta Y))$ $/(\Delta Y)^{1/\alpha}$, for positive ΔY^R, where the differencing lags were τ=1,5,10,50, and 100 for linear fractional stable motion with α=1.8 and d=0.3. The left inset shows $P(0,\tau)$ decaying with τ^{-H}, while the right inset shows sample standard deviation $\hat{\sigma}$ growing with τ^J.

$x\rightarrow\infty$ they all share power law tail behavior, with $P(x)\sim x^{-(1+\alpha)}$. Such α-stable noise is still stationary in the sense that its variance is theoretically constant (as is its mean in the α>1 case), but the infinite variance has many consequences, very strong fluctuations of the observed sample mean, and the "spiky" and "stepped" appearance of the X_t and Y_t time series, respectively, seen in the inset and main panel of Figure 3, where α=1.8. Inspired by the work of Paul Lévy, Mandelbrot (1963) introduced this model into econometrics as a description of fluctuations in the log differences of cotton prices.

Self-similar H=1/α stable motion: When integrated, the α-stable noise $X(t)$ becomes α-stable motion $Y(t)$. This is still self-similar but H is now 1/α rather than 2, illustrated in the left-hand inset of Figure 4 where $P(0,\tau)$ is seen to decay with an exponent of -0.5573, close to its theoretical value of 1/1.8=0.556. Measures of the Hurst exponent J still, however, return J=1/2, illustrated in the right-hand inset of Figure 4, where σ scales with τ^J with a scaling exponent of +0.4993. The pdfs $P(\Delta Y^R)$ still collapse onto each other if the rescaling operation used in the Gaussian case replaces division by $\hat{\sigma}$ with division by $\tau^{1/\alpha}$, confirming that this is the appropriate H. A doubly logarithmic plot is used to make the asymptotic power law tails of $P(\Delta Y^R)$ easier to see in Figure 4.

2.3. The Joseph Effect and the Fractional Gaussian Family

Stationary fractional ($d\neq$0), Gaussian (α=2), noise: If instead of a delta function one allows the ACF $R(\tau)$ of the model noise $X(t)$ to have a power law tail at large τ values, $R(\tau)\sim|\tau|^{2d-1}$ the noise time series is found to acquire the property of LRD, when $0<d<1/2$. Because $X(t)$ has the weak time stationarity needed to make the two-time ACF into a function depending only on the lag τ between the two times, the PSD is also well-defined (Brockwell & Davis, 2016), and so one can apply the Wiener-Khinchine theorem which relates the Fourier transform of the ACF to the PSD. One then finds a power law PSD $S(f)\sim f^{-2d}$ as $f\rightarrow0$ (a precise formulation is given in Beran et al., 2013). If we also assume a Gaussian amplitude distribution and random phase the resulting model is essentially the fractional Gaussian noise (fGn), originally introduced by Kolmogorov (1940), and studied by Mandelbrot, Wallis, and Van Ness in the late 1960s (Mandelbrot, 2002). The fGn with positive d is said to be persistent because a value of a given sign will tend to be followed by another of the same sign. It has not always been appreciated, including by me (Watkins, 2017a), that fGn with negative d, which corresponds to antipersistence, is not LRD. For detailed discussions of this point see Beran et al. (2013) and Raymond et al. (2003).

The pioneering application of fGn was to hydrology in a series of papers beginning with Mandelbrot and Wallis (1968) about Hurst's Nile river height data set (Graves et al., 2017; Mandelbrot, 2002), for which the rescaled adjusted range, or R/S, diagnostic was developed. The link between R/S and the concept of an ideal water reservoir is described in detail by Feder (1988). The procedure to calculate it is defined, for example,

by Embrechts and Maejima (2003): If $\{X_i\}$ is a sequence of random variables, and $Y_n = \sum_{i=1}^{n} X_i$ is their cumulative sum, one can explore how Y_n fluctuates by defining the "adjusted range" $R(l,k)$, where

$$R(l, k) = \max_{1 \leq i \leq k} \left(Y_{l+i} - Y_l - \frac{i}{k} \left(Y_{l+k} - Y_l \right) \right) - \min_{1 \leq i \leq k} \left(Y_{l+i} - Y_l - \frac{i}{k} \left(Y_{l+k} - Y_l \right) \right) \tag{4}$$

and a standard deviation $S(l,k)$ by

$$S(l, k) = \left(\frac{1}{k} \sum_{i=l+1}^{l+k} (X_i - \bar{X}_{l,k})^2 \right)^{1/2} \tag{5}$$

where $\bar{X}_{l,k} = k^{-1} \sum_{i=l+1}^{l+k} X_i$. The rescaled adjusted range, or R/S is given by $Q(l,k)=R(l,k)/S(l,k)$ and can be used to estimate J from a $\log Q$ versus $\log k$ plot, though more accurate diagnostics are now available (Beran et al., 2013). The observed departure of J from 1/2 in several natural time series including Nile river minima was the famous "Hurst effect" (Feder, 1988).

Many applications of fBm and fGn have followed (Beran et al., 2013; Graves et al., 2017), and very recently the prediction of time-averaged temperature series with the StocSIPS technique (Del Rio Amador, 2019; Lovejoy et al., 2018) has used an fGn model. It is important to realize, however, that if there is actually any nonrandom phase structure in a real data,set, the fGn model, by construction, will be blind to it.

Fractional ($H=d+1/2$) *Brownian motion*: When integrated, the fGn $X(t)$ gives fractional Brownian motion (fBm) $Y(t)$. Confusion sometimes arises about the difference between stationary fGn (parameterized by the memory parameter d, which runs between $-1/2$ and $1/2$) and self-similar but nonstationary fBm (parameterized by the self-similarity parameter H, which runs from 0 to 1). The two parameters are linked in these models by $H=d+1/2$ and in this Gaussian special case H is additionally equal to J. Figure 5 shows how X and Y differ, with the inset showing an fGn time series X_t, for $d=0.3$, which retains the bounded appearance of white noise, due to its stationarity and Gaussian distribution, but which shows slow trend-like fluctuations arising from the presence of LRD. The main panel of Figure 5 shows a time series of fBm where $H=0.8$, while the left and right panels of Figure 6 illustrate the equivalence, in this case, of H and J, with the measured values being $H=0.7819$ and $J=0.7878$, respectively. The main panel of Figure 6 shows that a good scaling collapse of the rescaled pdfs $P(\Delta Y^R)$ is obtained when the rescaling operation used in the Gaussian case has division by the sample standard deviation replaced with division by τ^H, confirming that this is the appropriate self-similarity exponent. I have used a semilog plot rather than a log-log plot here because of the Gaussian shape of the pdfs.

2.4. The Noah and Joseph Effects Combined in the Fractional α-Stable Family

Stationary, fractional ($d\neq0$), *stable* ($\alpha\neq2$), *noise*: Combining both heavy tails and memory gives fractional stable noise (Samorodnitsky and Taqqu, 1994), a two-parameter (α, d) stationary model. The time series of discrete fractional stable noise X_t illustrated in the inset of Figure 7, combines the "spikiness" present in the heavy-tailed simulation of section 2.2, with the same $\alpha=1.8$, and the slow trend-like fluctuations from LRD seen in the fractional Gaussian noise simulation of section 2.3, with the same $d=0.3$. However, as in the iid α-stable model the theoretical variance is infinite, which complicates the use of the many variance-based diagnostics such as R/S that have typically been used to infer and quantify LRD in fGn and other Gaussian models (Mandelbrot, 2002).

Linear fractional ($H=d+1/\alpha$) *stable motion*: When fractional stable noise is integrated, linear fractional stable motion (LFSM) is obtained, illustrated in Figure 7. When applied to LFSM the well-known diagnostics of the Joseph exponent like the original R/S method continue to measure J, not H. In this non-Gaussian case the self-similarity exponent H is no longer the same as J; instead, $H=1/\alpha+d$, while J remains $1/2+d$ as it would be for fBm. We can see this in the left and right insets to Figure 8, where $P(0)$ decays with exponent -0.8307, close to $-H=-0.8556$, while $\sigma(\Delta Y)$ grows with exponent 0.7895, close to 0.8. Mandelbrot himself had noted this effect as early as 1969 but unfortunately chose to present it (Mandelbrot & Wallis, 1969) as the "robustness of R/S" as a measure of LRD, rather than pointing out that J and H need not always be the same. Although this remains a frequently overlooked problem in applications of the Holder and Joseph exponents, good discussions of it now exist in the literature, notably Mercik et al. (2003), which will aid the reader considerably when approaching Mandelbrot's frequently inaccessible original papers (Mandelbrot, 2002).
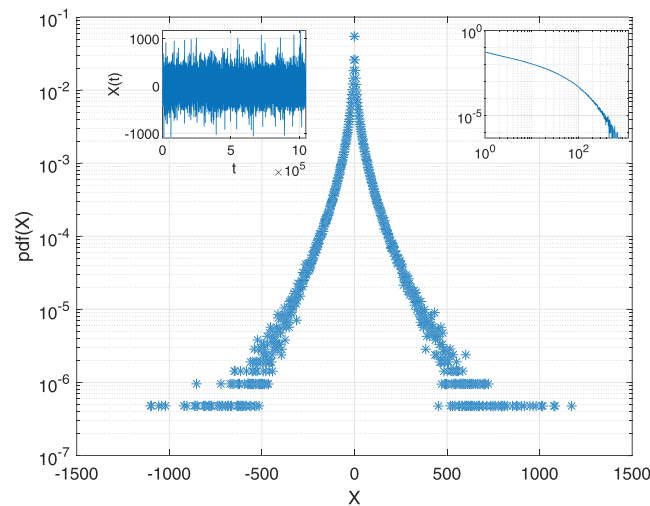
Figure 9. Time series, semilog probability distribution function (pdf) and loglog pdf calculated using the multiplicative rainfall model of (Wilson & Toumi, 2005), $\sigma_1=1$, $\sigma_2=5$, and $\sigma_3=10$.

3. Multiplicative Models and Turbulent Cascades

We now shift perspective from the typically one-dimensional time series models of the previous sections, to macroscopic fields, for which examples might be a global circulation model or spatial data from a geostationary observation satellite. The atmosphere would then be represented by a set of n-dimensional fields, which, at least theoretically, have time-evolving phases and amplitudes specified everywhere. The phase is no longer random but carries important information about coherent structures resulting from fluid physics. In such a system the spatial and temporal aspects of a pattern would not be presumed to be separate but rather entangled, as noted vividly for clouds by Bak and Chen (1989): " … it is hard to believe that long-range spatial and temporal correlations can exist independently. A local signal cannot be 'robust' and remain coherent over long times in the presence of any amount of noise, unless stabilized by the interactions with its environment. And a large, coherent spatial structure cannot disappear (or be created) instantly. [...] think of the temporal distribution of sunshine, which must be correlated with the spatial distribution of clouds, through the dynamics of meteorology."

Perhaps the richest geophysically relevant example of such a field is fluid turbulence (Frisch, 1995). Fully developed turbulence is notable for the complex structures it gives rise to in three-dimensional real space, famously drawn by da Vinci, and for its accompanying k space forward cascade, in which energy, injected at long wavelengths λ, cascades down the inertial range toward short wavelengths (or equivalently high wave numbers, $k(=2\pi/\lambda)$) where it is dissipated. To study the interplay of real space and k space in such a system using time series analysis, even in a controlled system like a wind tunnel one ideally wants several measurement sites. This is even more true for uncontrolled natural systems—one reason why network science is now so topical in geoscience (Donner et al., 2017). If limited to just one time series, then one will at least need to ensure that one uses a model capable of capturing the complicated phase and amplitude information in such a series. The well-known phenomenological models (Frisch, 1995) of the turbulent cascade are multiplicative by construction, as they capture the way in which energy in the inertial range at a given wavelength $E(k)$ is subdivided in a stochastic way to give that at the next level down $E(k-\Delta k)$.

Lovejoy and Schertzer (2013) have argued cogently that in such cases a multifractal picture will usually be the most natural model, where the single scaling exponents H and J are replaced by a continuum of exponents, the "multifractal spectrum." The concept of multifractality derives from the work of Mandelbrot (1999, 2004) and others (see, e.g., Chapter 17 of Falconer, 2003). A brief introduction, developing the multifractal concept after summarizing the additive models is given by Schmitt and Huang (2016), and a detailed discussion of the relevance to atmospheric science can be found in Lovejoy and Schertzer (2013).

Even without considering phase effects, though, one can begin to see how different multiplicative models are from additive ones, but how their phenomenology can still at first glance resemble some additive cases, in an even simpler example, studied by Wilson and Toumi (2005). Here just three Gaussian random variables are

multiplied, representing mass flux, specific humidity, and the precipitation efficiency in an extreme rainfall model. In the time series plotted in the left inset of Figure 9 the random numbers all had zero mean, but standard deviations of 1, 5, and 10, and the series is seen to fluctuate much more than a white noise would. The central panel shows that the resulting pdf has tails that are heavier than Gaussian, though fitted well by stretched exponentials, not power laws (Wilson & Toumi, 2005). The right-hand inset shows the same pdf in log-log axes and illustrates the by now well-recognized point (Sornette, 2003) that the ever-present temptation to infer a rolled off power law by eye should be resisted!

4. "1/f" Noise From Changes in Signal Level

Despite its theoretical elegance, it should be remembered that Mandelbrot's fBm was not the first explanation given for the Hurst effect, as recounted by Graves et al. (2017). A popular alternative view has been that the effect is usually preasymptotic and arises from processes with correlation times that are of the order of the observation length. The earliest attempt to mimic this, Hurst's own investigation of card-cutting models with a finite correlation length (Feder, 1988), remains instructive. The J exponent was seen to fall to 0.5 as the minimum lag used to fit an R/S versus lag plot was increased (see Figure 8.9 of Feder, 1988), an effect that would presumably not be seen with a true LRD signal like fBm.

The presence of breakpoints, for example, those arising from changed instrumentation, complicates the inference of trends, especially in data sets taken from uncontrolled natural systems. The problem of how best to distinguish between trends, shifts, and short-term memory of the $AR(1)$, or more generally the Autoregressive Moving Average ($ARMA$) types, remains an active topic in the recent climate statistics literature (e.g., Beaulieu & Killick, 2018; Poppick et al., 2017).

Even a single change point poses a problem for the inference of LRD (e.g., Katsev & L'Heureux, 2003). Consider a monofractal signal, with an expected J less than 1/2, say 1/3, "contaminated" by sharp jumps of the type seen in many natural time series, such as those caused by sector boundaries in the solar wind. The effective measured J can be *raised* toward 1/2 by this extra source of "roughness" in the time series. By separating such breakouts and examining their power spectrum, Borovsky (2010) was able to show their effect on the inferred power spectrum of solar wind turbulence. Unfortunately, though, change points cannot always be detected using a priori knowledge, and as Rooch et al. (2019) recently put it in their Working Paper: "...Conventional estimators of [J] easily lead to spurious detection of long memory if the time series includes a shift in the mean. This defect has fatal consequences in change-point problems: Tests for a level shift rely on [J], which needs to be estimated before, but this estimation is distorted by the level shift." In consequence, the problem of detecting LRD in the presence of breaks has become an active area in statistics in recent decades (e.g., Betken, 2016; Sibbertsen, 2004). In their recent work Rooch et al. (2019) propose an estimator for the J of a length n series based on averaging estimates from overlapping blocks of length (approximately a multiple of $\sqrt{n}$), but they note that a minimum of 100 observations per block should be available in the ARFIMA case.

The problem can be turned on its head, however, and particular models with power law distributions of the waiting times between change points can themselves be studied as models of 1/f noise. Because of the attention received by fGn it has very rarely been noticed that Mandelbrot proposed and published a series of three papers (Mandelbrot, 1999) on exactly this latter class of models in parallel with it, culminating in Mandelbrot (1967). These models, chronicled by Watkins (2017b) were very different from fGn, being, in an important sense, nonstationary. Mandelbrot described a stochastic process intended to caricature the observed intermittency in systems such as telephone line traffic, or fluid turbulence, and rather than the more familiar short-tailed distributions of the waiting times between transitions between discrete states of activity, he proposed long-tailed power laws. His key finding, in Mandelbrot (1967), was that traditional time series analysis techniques like the PSD and ACF would return a "1/f" spectrum for this model, but that the PSD $S(f)$ could be decomposed into a stationary part, and a random prefactor dependent on the observed time series' length. In this paper he pointed out that this resolved the so-called "1/f paradox"; the fact that a low-frequency cutoff would be needed to make such a power spectrum integrable, but was frequently not observed. (Mandelbrot, 1999) returned to the topic, arguing that rather than representing a true singularity in power at the lowest frequencies, the infrared catastrophe in the power spectral density seen in this model was a "mirage" resulting from the fact that the moments of the model varied in time in a step-like fashion, which he called "conditional covariance stationarity."

Unfortunately, Mandelbrot (1967) received far less contemporary attention than his papers on heavy tails in finance (Mandelbrot, 1963, 1997) or the series with Van Ness and Wallis in 1968–1969 on stationary fractional Gaussian models for LRD (Mandelbrot, 2002), gaining only about 20 citations in its first 20 years. It still seems to be little known, being, for example, not cited by the otherwise exhaustively detailed Beran et al. (2013), which nonetheless refers to newer work on this class of model under the name of "long-range count dependence" (LRcD). Although he revisited the papers with new commentaries in the volume of his Selecta (Mandelbrot, 1999) on multifractals and "1/f" noise, Mandelbrot himself neglected to mention them explicitly in his popular and historical account of the genesis of LRD in Mandelbrot and Hudson (1998). He clearly saw them as comprising a different strand of his work to that on fractional Brownian motion collected in another Selecta volume (Mandelbrot, 2002).

Since the 1960s, however, as described by Watkins (2017b) and Beran et al. (2013) fractional renewal models of the LRcD type, as well as the continuous time random walk (CTRW) with heavy-tailed waiting times (Kutner & Masoliver, 2017), both structurally very similar to Mandelbrot (1967, 1999) have subsequently been studied extensively by many authors. Applications have included computer science and internet traffic, and physics where they are now being actively considered as solutions to the "1/f" paradox, in the context of blinking quantum dots. In this latter context the time dependence of the random prefactor to "1/f" spectra is receiving renewed attention (e.g., Niemann et al., 2016), being related to the physical phenomenon of weak ergodicity breaking.

With the advantage of historical distance one can see the importance of both nonstationary fractional renewal models and stationary Gaussian fGn. In the valedictory essays of his Selecta N volume, Mandelbrot (1999) argued that this kind of structured nonstationarity as a mechanism for 1/f noise was an intermediate step between stationary LRD and more unstructured and arbitrary forms of nonstationarity and explicitly stated that it was a key step in his development toward the multifractal idea.

5. Beyond Mandelbrot's Walks and Noises: Mean Reverting Models for Dissipative Time Series

Mandelbrot's additive models extended two basic processes: (i) stationary white noise $X_{t+1}=\sigma\eta_t$, where at each time step the value of the discrete time series is drawn from a normal distribution of variance σ^2, preserving no information about the previous state, and (ii) nonstationary Brownian motion $X_t=X_{t-1}+\sigma\eta_t$ which adds a random variable at each time step to the previous X_t value, thus making full use of the value of the previous state of the system. Many physical and economic systems, however, require a model that lies between these two extremes, because they show a tendency for fluctuations to be damped out over time so that the time series reverts to a mean value. The classic way to achieve this in discrete models is via the $AR(1)$ model's ϕ_1 parameter, which lies in the range $|\phi|<1$, so that $X_t=\phi_1 X_{t-1}+\sigma\eta_t$. $AR(1)$ has a continuous counterpart, the Langevin equation in physics (Coffey & Kalmykov, 2017) for the velocity of a damped Brownian particle, whose solution has been studied extensively as the Ornstein-Uhlenbeck (OU) process, also known as continuous $AR(1)$ or $CAR(1)$ (Brockwell & Davis, 2016; Coffey & Kalmykov, 2017).

Extension of the $AR(1)$ model to pth order, so that $X_t=\phi_1 X_{t-1}+\phi_2 X_{t-2}\ldots+\phi_p X_{t-p}+\sigma\eta_t$ (Brockwell & Davis, 2016) allowed non-Markovian memory to be considered and fostered a useful continuing dialogue between users of such $AR(p)$ models and advocates of Mandelbrot's proposal of LRD, which effectively replaced the need for an $AR(\infty)$ model (Brockwell & Davis, 2016; Graves et al., 2017) with a single parameter d. The eventual introduction of the $ARFIMA(p,d,q)$ (Brockwell & Davis, 2016; Graves et al., 2017) model allowed Mandelbrot's d to be incorporated into flexible models which also incorporated shorter-ranged correlations via pth order autoregression and qth order moving average terms. Despite Mandelbrot's expressed distaste for the $ARFIMA$ model (Graves et al., 2017), it has become the standard way to incorporate LRD into time series modeling in many communities, notably statistics and econometrics (Brockwell & Davis, 2016). Recent progress has been made in the continuous counterpart, the $CARFIMA(p,d,q)$ model by Tsai (2009), and considerable rigor has now been attained (Kubilius et al., 2017) in the study of the fractional Ornstein-Uhlenbeck process, that is, OU driven by a fractional noise. However, I have previously argued (Watkins, 2013, 2017a) that the fractal extension of the Langevin equation of most potential interest to geoscience is the fractional Langevin equation (Coffey & Kalmykov, 2017), which as well as having a fractional noise term replaces the constant damping term in the Langevin equation by an integral over a time-dependent memory kernel, which is then taken to have a power law decay. Investigation of equations

of the fractional Langevin type in the global temperature modeling context has now begun (Lovejoy et al., 2018; Watkins et al., in preparation) as well as other closely related models (Rypdal & Rypdal, 2014).

In closing I would note that Nature, of course, is not compelled to actually follow any of these models closely, or at all, but evidence is accumulating that many of them are surprisingly effective. Whatever the final verdict on Mandelbrot's fractal time series models, I would argue that they have greatly expanded the quantitative descriptive palette available to geophysics.

Data Statement:

The short Matlab simulation and data analysis codes used to generate the figures are available from the author. Readers seeking best practice in code design and algorithms are however directed to a better resource such as the codes available at https://github.com/lanlankai for the book by Schmitt and Huang (2016).

References

Bak, P., & Chen, K. (1989). The physics of fractals. *Physica D*, *38*, 5–12.

Baker, G. L., & Gollub, J. P. (1990). *Chaotic dynamics*. An Introduction: Cambridge University Press.

Beaulieu, C., & Killick, R. (2018). Distinguishing trends and shifts from memory in climate data. *Journal of Climate*, *31*, 9519–9543.

Beran, J., Feng, Y, Ghosh, S., & Kulik, R. (2013). *Long-memory processes*. Probabilistic Properties and Statistical Methods: Springer.

Betken, A. (2016). Testing for change-points in long-range dependent time series by means of a self-normalized Wilcoxon test. *Journal of Time Series Analysis*, *37*, 785–809. https://doi.org/10.1111/jtsa.12187

Borovsky, J. E. (2010). Contribution of strong discontinuities to the power spectrum of the solar wind. *Physical Review Letters*, *105*, 111102.

Brockwell, P. J., & Davis, R. A. (2016). *Introduction to time series and forecasting* (3rd ed.). Cham, Switzerland: Springer.

Chapman, S. C., Hnat, B., Rowlands, G., & Watkins, N. W. (2005). Scaling collapse and structure functions: Identifying self affinity in finite length time series. *Nonlinear Processes in Geophysics*, *12*, 767–774.

Coffey, W. T., & Kalmykov, Y. P. (2017). *The Langevin equation: With applications to stochastic problems in physics, chemistry and electrical engineering* (4th ed.). Singapore: World Scientific.

Cvitanovic, P. (Ed.) (1984). *Universality in chaos* Cvitanovic, P.. Bristol, UK: Adam Hilger.

Del Rio Amador, L. (2019). Predicting the global temperature with the Stochastic Seasonal to Interannual Prediction System (StocSIPS). *Climate Dynamics*, *53*, 4373–4411. https://doi.org/10.1007/s00382-019-04791-4

Donner, R., Wiedermann, M., & Donges, J. (2017). Complex network techniques for climatological data analysis. in Franzke and O'Kane, 2017.

Embrechts, P., & Maejima, M. (2003). *Self-similar processes* . Princeton, NJ: Princeton University Press.

Falconer, K. (2003). *Fractal Geometry: Mathematical foundations and applications* (2nd ed.). Chichester, UK: Wiley.

Feder, J. (1988). *Fractals*. New York: Plenum.

Frisch, U. (1995). *Turbulence: The legacy of A. N. Kolmogorov*. Cambridge, UK: Cambridge University Press.

Graves, T., Gramacy, R. B., Franzke, C. L. E., & Watkins, N. W. (2017). A brief history of long memory: Hurst, Mandelbrot and the road to ARFIMA, 1951-1980. *Entropy*, *19*, 437. https://doi.org/10.3390/e19090437

Giles, J. (2004). Father of Fractals. *Nature*, *432* (7015), 266–267.

Katsev, S., & L'Heureux, I. (2003). Are Hurst exponents estimated from short or irregular time series meaningful? *Computers and Geosciences*, *29*, 1085–1089.

Klafter, J., Shlesinger, M. F., & Zumofen, G. (1996). Beyond Brownian motion. *Physics Today*, *49*(2), 33–39.

Kolmogorov, A. N. (1940). Wiener spirals and some other interesting curves in a Hilbert space. *Dokl. Akad. Nauk SSSR 26*, 115–118 (in Russian), reprinted in V. M. Tikhomirov (Ed.), Selected Works of A. N. Kolmogorov, Kluwer Academic Publishers, Dordrecht, 1991.

Kondrashov, D., Feliks, Y., & Ghil, M. (2005). Oscillatory modes of extended Nile River records (A.D. 622–1922). *Geophysical Research Letters*, *32*, L10702. https://doi.org/10.1029/2004GL022156

Koutsoyiannis, D. (2013). Hydrology and change. *Hydrological Sciences Journal*, *58*(6), 1177–1197.

Kubilius, K., Mishura, Y., & Ralchenko, K. (2017). *Parameter estimation in fractional diffusion models*. Cham, Switzerland: Springer.

Kutner, R., & Masoliver, J. (2017). The continuous time random walk, still trendy: Fifty-year history, state of art and outlook. *European Physical Journal B*, *90*, 50. https://doi.org/10.1140/epjb/e2016-70578-3

Lorenz, E. A. (1963). Deterministic nonperiodic flow. *Journal of Atmospheric Science*, *20*, 130–141. https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2.

Lovejoy, S. (1982). Area perimeter relations for rain and cloud areas. *Science*, *187*, 1035–7.

Lovejoy, S., Del Rio Amador, L., & Hébert, R. (2018). Harnessing b: Theory and Practice of the Stochastic Seasonal to Interannual Prediction System (StocSIPS). In A. Tsonis (Ed.), *Advances in Nonlinear Geosciences*. Cham, Switzerland: Springer.

Lovejoy, S., & Schertzer, D. (2013). *The weather and climate*. Emergent Laws and Multifractal Cascades: Cambridge University Press.

Mandal, P., Mabawonku, A. O., & Dimri, V. P. (2005). Self-organized fractal seismicity of reservoir triggered earthquakes in the Koyna-Warna Seismic Zone, Western India. *Pure and Applied Geophysics*, *162*, 73–90.

Mandelbrot, B. B. (1963). The variation of certain speculative prices. *The Journal of Business*, *36*(4), 394–419.

Mandelbrot, B. B. (1967). Some noises with 1/f spectrum, a bridge between direct current and white noise. *IEEE Transactions on Information Theory*, *IT-13*(2), 289–298.

Mandelbrot, B. B. (1982). The fractal geometry of nature. W. H. Freeman.

Mandelbrot, B. B. (1997). *Fractals and scaling in finance: Discontinuity, concentration risk, Selecta volume E*. New York: Springer-Verlag.

Mandelbrot, B. B. (1999). *Multifractals and 1/F noise: wild self-affinity in physics (1963–1976), Selecta volume N*. New York: Springer-Verlag.

Mandelbrot, B. B. (2002). *Gaussian self-affinity and fractals: Globality, the earth, 1/F noise and R/S, Selecta volume H*. New York: Springer-Verlag.

Mandelbrot, B. B. (2004). *Fractals and chaos: The Mandelbrot set and beyond, Selecta volume C*. New York: Springer-Verlag.

Mandelbrot, B. B., & Hudson, R. L. (1998). *The (mis)behaviour of markets*. London: Profile Books.

Mandelbrot, B. B., & Wallis, J. R. (1968). Noah, Joseph, and operational hydrology. *Water Resources Research*, *4*(5), 909–918.

Mandelbrot, B. B., & Wallis, J. R. (1969). Robustness of R/S in measuring noncyclic global statistical dependence. *Water Resources Research*, *5*, 967–988.

Mercik, S., Weron, K., Burnecki, K., & Weron, A. (2003). Enigma of self-similarity of fractional lévy stable motions. *Acta Physica Polonica B*, *34*, 3773–3791.

Niemann, M., Barkai, E., & Kantz, H. (2016). Renewal theory for a system with internal states. *Mathematical Modelling of Natural Phenomena*, *11*(3), 191–239.

Osso, A., Corral, A., & Llebot, J. E. (2010). Scaling of tropical cyclone dissipation. *Nature Physics*, *6*(9). https://doi.org/10.1038/nphys1725

Poppick, A., Moyer, E. J., & Stein, M. L. (2017). Estimating trends in the global mean temperature record. *Advances in Statistical Climatology, Meteorology and Oceanography*, *3*, 33–53. https://doi.org/10.5194/ascmo-3-33-2017

Raymond, G. M., Percival, D. B., & Bassingthwaighte, J. B. (2003). The spectra and periodograms of anti-correlated discrete fractional Gaussian noise. *Physica A*, *322*, 169–179.

Rodríguez-Iturbe, I., & Rinaldo, A. (1997). *Fractal river basins: Chance and self-organisation*. Cambridge, UK: Cambridge University Press.

Rooch, A., Zelo, I., & Fried, R. (2019). Estimation methods for the LRD parameter under a change in the mean. *Statistical Papers*, *60*(1), 313–347.

Rundle, J. B., & Klein, W. (1993). Scaling and critical phenomena in a cellular automaton slider-block model for earthquakes. *Journal of Statistical Physics*, *72*(1–2), 405–412.

Rypdal, M., & Rypdal, K. (2014). Long-memory effects in linear response models of Earth's temperature and implications for future global warming. *Journal of Climate*, *27*, 5240–5258.

Samorodnitsky, G., & Taqqu, M. S. (1994). *Stable non-Gaussian random processes*. New York: Chapman and Hall.

Schmitt, F. G., & Huang, Y. (2016). *Stochastic analysis of scaling time series: From turbulence theory to applications*. Cambridge, UK: Cambridge University Press.

Sharma, A. S. (1995). Assessing the magnetosphere's nonlinear behavior: Its dimension is low, its predictability, high. *Reviews of Geophysics*, *33*, S1. https://doi.org/10.1029/95RG00495

Sibbertsen, P. (2004). Long memory versus structural breaks: An overview. *Statistical Papers*, *45*, 465. https://doi.org/10.1007/BF02760564

Sornette, D. (2003). *Critical phenomena in natural sciences: Chaos, fractals, self-organization and disorder—Concepts and tools (Springer Series in Synergetics)* (2nd ed.). Berlin: Springer.

Stoev, S., & Taqqu, M. (2004). Simulation methods for linear fractional stable motion and FARIMA using the fast Fourier transform. *Fractals*, *12*, 95.

Tsai, H. (2009). On continuous-time autoregressive fractionally integrated moving average processes. *Bernoulli*, *15*(1), 178–194.

Turcotte, D. L. (1997). *Fractals and chaos in geology and geophysics* (2nd ed.). Cambridge, UK: Cambridge University Press.

Watkins, N. W. (2013). Bunched black (and grouped grey) swans: Dissipative and non-dissipative models of correlated extreme fluctuations in complex geosystems. *Geophysical Research Letters*, *40*, 4020–410. https://doi.org/10.1002/GRL.50103

Watkins, N. W. (2017a). Fractional stochastic models for heavy-tailed, and long-range dependent, fluctuations in physical systems. In C. L. E. Franzke & T. J. O'Kane (Eds.), *Nonlinear and stochastic climate dynamics* (pp. 340–368). Cambridge University Press.

Watkins, N. W. (2017b). On the continuing relevance of Mandelbrot's non-ergodic fractional renewal models of 1963 to 1967. *The European Physical Journal B*, *90*, 241.

Watkins, N. W., Pruessner, G., Chapman, S. C., Crosby, N. B., & Jensen, H. J. (2016). 25 years of self-organized criticality: Concepts and controversies. *Space Science Reviews*, *198*, 3–44. https://doi.org/10.1007/s11214-015-0155-x

Wilson, P. S., & Toumi, R. (2005). A fundamental probability distribution for heavy rainfall. *Geophysical Research Letters*, *32*, L14812. https://doi.org/10.1029/2005GL022465

Zaslavsky, G. (2002). Chaos, fractional kinetics, and anomalous transport. *Physics Reports*, *371*, 461–580.

Erratum

In the originally published version of this article, there were minor typographic errors. These errors have since been corrected and this version may be considered the authoritative version of record.