AGU100 ADVANCING EARTH AND SPACE SCIENCE

Earth and Space Science

REVIEW ARTICLE
10.1029/2018EA000435

Special Section:
Nonlinear Systems in Geophysics: Past Accomplishments and Future Challenges

Key Points:
- Hypothesis testing with surrogate data is described
- Temporal and spatial data are considered
- Examples make use of gradual reconstruction tools

Correspondence to:
C. J. Keylock,
c.j.keylock@lboro.ac.uk

Citation:
Keylock C. J. (2019). Hypothesis testing for nonlinear phenomena in the geosciences using synthetic, surrogate data. *Earth and Space Science*, *6*, 41–58. https://doi.org/10.1029/2018EA000435

Received 30 JUL 2018
Accepted 7 DEC 2018
Accepted article online 15 DEC 2018
Published online 19 JAN 2019

Hypothesis Testing for Nonlinear Phenomena in the Geosciences Using Synthetic, Surrogate Data

C. J. Keylock[1]

[1]School of Architecture, Building and Civil Engineering, Loughborough University, Loughborough, UK

Abstract Studying nonlinear and potentially chaotic phenomena in geophysics from measured signals is problematic when system noise interferes with the dynamic processes that one is trying to infer. In such circumstances, a framework for statistical hypothesis testing is necessary but the nonlinear nature of the phenomena studied makes the formulation of standard hypothesis tests, such as analysis of variance, problematic as they are based on underlying linear, Gaussian assumptions. One approach to this problem is the method of surrogate data, which is the technique explained in this paper. In particular, we focus on (i) hypothesis testing for nonlinearity by generating linearized surrogates as a null hypothesis, (ii) a variant of this that is perhaps more appropriate for image data where structural nonlinearities are common and should be retained in the surrogates, and (iii) gradual reconstruction where we systematically constrain the surrogates until there is no significant difference between data and surrogates and use this to understand geophysical processes. In addition to time series of sunspot activity, solutions to the Lorenz equations, and spatial maps of enstrophy in a turbulent channel flow, two examples are considered in detail. The first concerns gradual wavelet reconstruction testing of the significance of a specific vortical flow structure from turbulence time series acquired at a point. In the second, the degree of nonlinearity in the spatial profiles of river curvature is shown to be affected by the occurrence of meander cutoff processes but in a more complex fashion than previously envisaged.

Plain Language Summary Complexity and nonlinearity in physics and the geosciences are at the heart of understanding a great number of processes. It can therefore be difficult to infer exactly what is happening in a system due to this complexity. This paper examines various methods for generating surrogate data for testing hypotheses about nonlinearity in data. That is, the surrogates provide the null model and one looks to see if the real data has a value significantly different to the surrogates. Examples are given for how a variable changes in time, how several variables measured at the same point change in time, and how the spatial values for a variable change. In particular, we consider a method called *gradual wavelet reconstruction* where we can systematically change the properties of the surrogates until there is no longer a difference between data and surrogates and use this as a way to understand geophysical processes. Examples of the use of this technique are given for understanding scour caused by turbulent flows transporting sediment and how river meandering dynamics effect the curvature of the river channel.

1. Introduction

There is no double-blind, replicated trial for a tsunami, an earthquake, a tornado or an ionospheric perturbation (Simpson, 1963). Numerical models provide a means to study the dynamics of such phenomena but are dependent upon the relevant process mechanics being understood. Thus, we need to complement numerical investigation with field and laboratory work to gain as much information as possible about the dynamics and contingent circumstances affecting such an event, thereby informing the development of the more refined models of the future (Paniconi & Putti, 2015).

However, the process of measurement requires decisions to be taken on the variables to be measured, the acquisition frequency, and the dimensionality of the data, which means that we only gain a partial insight into the physics of the phenomenon under consideration. Given the complex, nonlinear, and nonlocal nature of many of these processes (Arens et al., 2008; Goncharenko et al., 2010; Hannachi et al., 2017; Tsinober, 2013), as well as the unique contexts in which they operate due to the historical development of the salient boundary conditions (Church & Slaymaker, 1989), it is not a trivial task to characterize the nature of this

nonlinearity (Aguirre & Billings, 1995). It therefore follows that it is not sufficient to apply some measure of nonlinearity, such as those reviewed by Schreiber and Schmitz (1997), to a geophysical system and state that the value is x; rather, one should also test to see if this value is statistically significantly different to what one might expect by chance. However, as already stated, we have no double-blind replicated trial for the natural event so we cannot undertake more "experiments" for field measurements, for example, and there are very rarely appropriate statistical tests for particular hypotheses regarding nonlinear phenomena. Therefore, how should we proceed? The suggested approach in this paper is through the use of synthetic, surrogate data that are built with certain properties and to which the observed data may be compared (Kantz & Schreiber, 1997; Schreiber & Schmitz, 1996; Theiler et al., 1992). After discussing the types of data that may be studied, and the way in which hypotheses may be tested, we consider two examples where the use of surrogate data-based hypothesis testing has helped to (a) further understanding of geophysical processes and (b) permitted refined inference of processes from observations of landscape. A particular focus of this work is the gradual reconstruction framework for undertaking further investigations of data once significant nonlinearity has been detected. The primary focus of this article is on time series, or one-dimensional spatial data, although we do also discuss spatial imagery. For a general overview of key aspects of nonlinear time series, the reader is recommended to take a look at Bradley and Kantz (2015).

2. Synthetic, Surrogate Data Algorithms

Following the groundbreaking work of Ed Lorenz in the 1960s on chaos theory (Lorenz, 1963), a corpus of work emerged in the 1970s and 1980s that derived other strange attractors from simple sets of coupled ordinary differential equations (Hénon, 1976; Rössler, 1976), studied these dynamics in some detail (Pomeau & Manneville, 1980), and developed a suite of methods for the analysis of nonlinear systems (Grassberger & Procaccia, 1983; Kaplan & Yorke, 1979; Takens, 1981; Wolf et al., 1985). However, assuming such a method has been employed, it is still not clear what the meaning of the result is given noise in the measurement process. Error bars may be employed, but that requires a model for the error, which may not be known for a non-Gaussian measure of a nonlinear system. A more recent development has been the formulation of the theory of pullback attractors for the analysis of nonautonomous dynamical systems where the forcing is time dependent (Pierini et al., 2018). The case where this forcing is stochastic in time leads to a random attractor (Chekroun et al., 2011), which may be analyzed by an ensemble of model simulations (Van Schaeybroeck & Vannitsem, 2016). These techniques are extremely useful for characterizing the variability in the attractor dynamics as a consequence of temporal variability from model simulations but are less applicable to experimental problems where the aim is to infer something about the dynamics from the measured signal.

Hence, as stated by Theiler et al. (1992, p. 78), an alternative is to

" … specify a well-defined underlying linear process or null hypothesis and to determine the distribution of the quantity we are interested in (dimension say) for an ensemble of surrogate data sets which are just different realizations of the hypothesized linear stochastic process. Then, rather than estimate error bars on the dimension of the original data, we put error bars on the value given by the surrogates."

Here the use of "dimension" means the correlation dimension for the attractor for a potentially chaotic phenomenon (Grassberger & Procaccia, 1983). The simplest way to generate a linearized variant of a source data set is by a modified Fourier phase randomization procedure. A linear model for a phenomenon (here assumed to be a time series, u_t) may be expressed as an autoregressive process (Brockwell & Davis, 2013)

$$u_t = a_0 + \sum_{q=1}^{Q} a_q u_{t-q} + \sigma\epsilon, \tag{1}$$

where the a_q are coefficients, σ captures the standard deviation of the random error, and ϵ is a zero mean, unit variance Gaussian noise. Use of the Wiener-Khintchine theorem means that in the Fourier domain, this autoregressive model is equivalent to specifying the amplitudes of the Fourier transform for u_t. However, nothing is stated about the Fourier phases. This observation resulted in the first proposal for a rigorous surrogate data algorithm by Theiler et al. (1992)—the amplitude-adjusted Fourier transform (AAFT) method. In addition to a phase randomization, this method also recognized that surrogates needed to mimic the observed values for a data series so that the surrogate data could be a legitimate realization of the process studied. These ideas underpinned what is perhaps the most widely used surrogate data algorithm, which has seen subsequent applications in economics, neuroscience, and geophysics—the iterated AAFT (IAAFT).

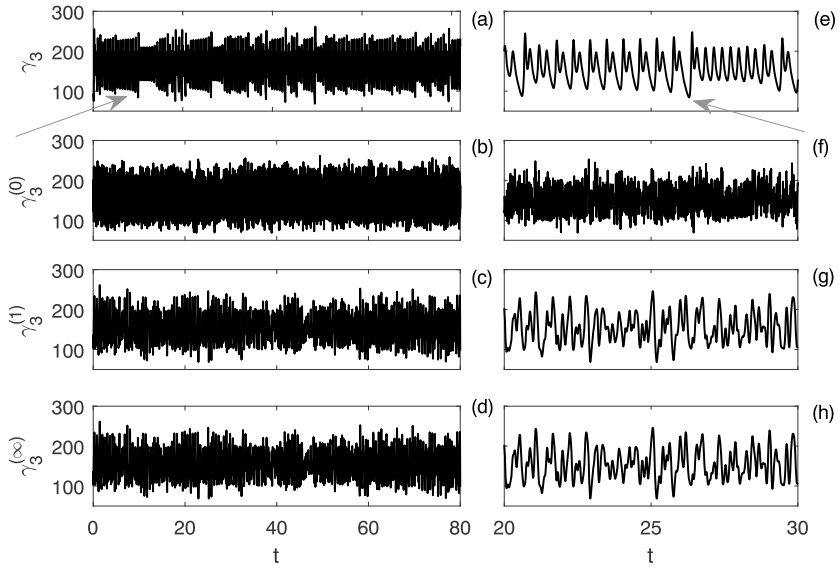

Figure 1. A time series for γ_3, the third variable in the Lorenz system, and its surrogate series are shown at coarse (panels a to d) and more refined (panels e to h) resolutions. Panels (a) and (e) show the original data. In panels (b) and (f) are shown the random shuffle step, $\gamma_3^{(0)}$ of the iterated amplitude-adjusted Fourier transform algorithm. The first complete iteration of the iterated amplitude-adjusted Fourier transform algorithm, $\gamma_3^{(1)}$, is shown in (c) and (g), and the final, converged iteration, $\gamma_3^{(\infty)}$, is given in (d) and (h). The two gray arrows relate to features described in the text.

2.1. The Iterated Variant of the AAFT Method

The IAAFT method was proposed by Schreiber and Schmitz (1996) as an improved surrogate generation algorithm. Given a time series, $u(t)$, one first stores these values in ranked order, and the amplitudes of its Fourier transform,

$$|F|(\omega) = \left| \sum_{t=0}^{N-1} u_t e^{i2\pi\omega t/N} \right|, \tag{2}$$

where ω is frequency. The initial surrogate, $u_t^{(0)}$ is a random shuffle of u_t. The algorithm then proceeds through the following steps:

1. Impose the correct periodicities by taking the Fourier transform of $u_t^{(0)}$, retaining the phases, and replacing the amplitudes by $|F|(\omega)$.
2. Take the inverse Fourier transform to produce a new time series, which will have the desired periodic structure but will no longer retain the original values of u_t. Hence, impose the original values via an amplitude adjustment step to produce $u_t^{(1)}$. The amplitude adjustment consists of ranking the observed values and substituting them for the value from the original data with the same position in the ranked list.

These two steps are repeated until $u_t^{(\infty)}$ is generated, a surrogate with the same values as u_t and a representation of the Fourier spectrum that is sufficiently precise to a selected error tolerance. It is this data set that gives us the surrogate, that is, $\hat{u}_t \equiv u_t^{(\infty)}$.

We consider the generation of an example IAAFT surrogate in Figure 1 based on the Lorenz (1963) equations but using a value for the Rayleigh parameter, k_2, that is much higher than normal, resulting in a dynamical system that has been considered as a simple model for intermittent turbulence (Pomeau & Manneville, 1980). Specifically, we solve

$$\begin{aligned} \frac{d\gamma_1}{dt} &= k_1(\gamma_2 - \gamma_1), \\ \frac{d\gamma_2}{dt} &= -\gamma_1\gamma_3 + k_2\gamma_1 - \gamma_2, \\ \frac{d\gamma_3}{dt} &= \gamma_1\gamma_2 - k_3\gamma_3, \end{aligned} \tag{3}$$

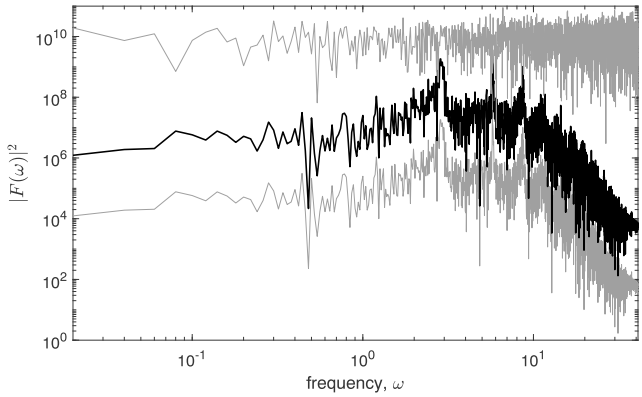

Figure 2. The square of the Fourier amplitudes as a function of frequency for γ_3 (black), $\gamma_3^{(\infty)}$ (gray and displaced vertically by dividing by 100), and $\gamma_3^{(0)}$ (gray and displaced vertically by multiplying by 1,000).

with $k_1 = 10$, $k_2 = 167$, and $k_3 = 8/3$ and extracted a time series for γ_3 consisting of 8,192 samples, obtained at 100 samples per unit time, following an initial simulation for 1,000 time steps to lose any effect from initial conditions. The original series and then the different steps of the IAAFT algorithm are shown in the different rows of Figure 1, with the right-hand panels magnifying a portion of the left-hand panels. While all rows consist of the same values for the data, it is clear that the random shuffle in the second row has no appropriate autocorrelative structure. However, this is visually broadly correct by just the first iteration in the third row and the fourth row is not visually too different to the third, indicating a rapid convergence rate for this time series. The values for $|F(\omega)|^2$ in Figure 2 highlight that the Fourier amplitudes are captured with excellent fidelity in the surrogate, which is displaced vertically downward as one of the series is not visible otherwise. For example, the error in the representation of the primary peak at $\omega = 2.88$ is 0.0125%. That the random shuffle, before the imposition of the Fourier amplitudes yields a white noise is clear from the upper gray line in Figure 2.

Despite the excellent preservation of the Fourier amplitudes, and while the surrogates and original data consist of exactly the same values, it is clear that there are some differences in the shape of the time series. The two gray arrows in Figure 1 show two such aspects of the signal: At the low frequencies emphasized in the left panel, there is a clear tendency for a gradual increase in the amplitudes of the higher-frequency oscillations before a sudden reduction in amplitude. At the higher frequencies emphasized in the right panel, the rising limb of one of these high-frequency oscillations appears to be much steeper than the recessional limb. Such features are not seen in the surrogate series. This would lead to the hypothesis that these are a significant aspect to the nonlinear nature of the process driving this signal. Thus, a measure of the signal asymmetry, such as that presented below, could be applied to the original data and surrogates to see if this hypothesis is borne out statistically.

While the IAAFT algorithm generally converges well, there may be issues with particular data sets. This could motivate one to embed the algorithm within a global convergence framework such as simulated annealing. However, a more straightforward method to improve convergence was pointed out by Venema et al. (2006) who imposed the original values in a gradual fashion. Another issue with the technique can emerge as a consequence of a lack of matching of the beginning and end of the data set. Because the Fourier transform is periodic, a harmonic is propagated into the surrogates if there is a major discrepancy in the first and last values of the original data set (implying a dramatic, near-instantaneous change from the Nth value to the first in a periodic domain). An example of this phenomenon is shown in Figure 12 of Keylock (2007). This can be handled by choosing a subsample of the original data such that the Nth and first values are similar, but this may be restrictive for environmental data that are hard to acquire. An alternative is to adopt a wavelet transform for surrogate generation, with the concomitant improvements that occur because of the localized nature of a wavelet with finite support (Keylock, 2007; Mallat, 1999; Percival & Walden, 2000).

2.2. Dealing With Multiple Time Series

Given that the basis for the IAAFT algorithm is the generation of linearized surrogates, it would be useful to formulate a variant suitable for a measurement program where multiple time series are acquired

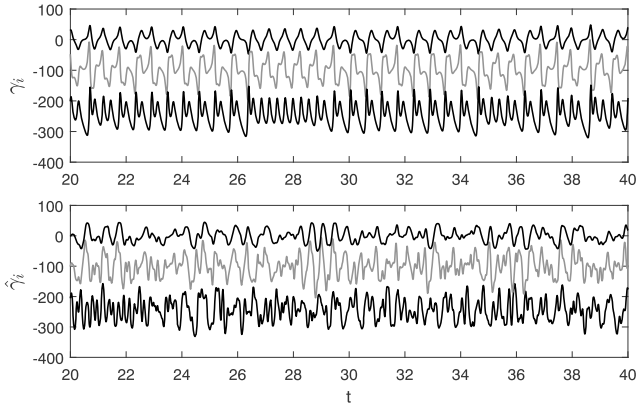

Figure 3. A subsection of the time series for the three terms, γ_i, in the Lorenz equations is shown in the upper panel with a realization in the lower panel using the iterated amplitude-adjusted Fourier transform algorithm that retains the original cross correlation between the surrogate series. In each panel, from top to bottom are shown γ_1 and γ_2 (in gray) and then γ_3. They are displaced vertically by 0, −100, and −250, respectively.

simultaneously. If the IAAFT algorithm in the form described above is adopted, then the null hypothesis is effectively that

There is no significant difference between the n time series and n linear variants drawn from the same distribution function and sampled independently.

The last part of this statement is problematic for a null hypothesis of linearity for multiple time series because the variance-covariance matrix is clearly a property of the linear system and covariances will not be preserved, except by chance, using IAAFT surrogates. Going further than variances and covariances, it is the cross-correlation structure that is crucial if one wishes to work with the null hypothesis that

There is no significant difference between the n time series and n linear variants drawn from the same distribution function and sampled from the same linear system.

The way to handle this was pointed out by Prichard and Theiler (1994) who noted that the key quantity is the difference in the Fourier phases between any given pair of signals. Hence, assuming that one has two series, in addition to storing the values and Fourier amplitudes for both series in the IAAFT algorithm, one also stores the Fourier phase differences between them. When the first series is randomly shuffled to produce $u_t^{(0)}$, the second series is initialized from the phases of this random shuffle with the phase difference imposed. An example of the use of this cross-correlation preserving algorithm is provided in Figure 3 where the upper panel shows the time evolution of the three terms in the Lorenz equations, γ_i, and the lower panel shows three corresponding surrogate series, $\hat{\gamma}_i$. For the original data, the maximum cross correlation between γ_1 and γ_2 occurs at a lag of $\tau = 0.04$ and is $R_\tau^{\max}(\gamma_1, \gamma_2) = 0.91$, while $R_\tau^{\max}(\gamma_1, \gamma_3) = 0.12$ and $R_\tau^{\max}(\gamma_2, \gamma_3) = 0.13$, occurring at $\tau = -6.48$ and $\tau = -4.88$, respectively. All these statistics were preserved to close to machine precision in the surrogates. That is, the standard deviation of $R_\tau^{\max}(\hat{\gamma}_1, \hat{\gamma}_2)$ for 39 surrogates including those shown was 2×10^{-16} with a mean of 0.91 as required.

2.3. Two-Dimensional Data

To illustrate the generation of surrogates for two-dimensional data, we show in Figure 4a a plane in the longitudinal (x) and transverse (z) directions of the enstrophy for a direct numerical simulation of a turbulent channel flow at a channel centerline Reynolds number of 22,625 (Graham et al., 2016; Li et al., 2008). The height from the lower boundary of this plane is $y^+ = 180$, where the plus indicates inner-scale (wall) units, according to $y^+ = yu_\tau/\nu$, where $\nu = 5 \times 10^{-5}$ is the (dimensionless) kinematic viscosity and $u_\tau = 4.997 \times 10^{-2}$ is the (dimensionless) friction velocity at the wall. With the velocity gradient tensor given by

$$\mathbf{A} = \frac{\partial u_i}{\partial x_j} \equiv \begin{pmatrix} \frac{\partial u_1}{\partial x_1} & \frac{\partial u_1}{\partial x_2} & \frac{\partial u_1}{\partial x_3} \\ \frac{\partial u_2}{\partial x_1} & \frac{\partial u_2}{\partial x_2} & \frac{\partial u_2}{\partial x_3} \\ \frac{\partial u_3}{\partial x_1} & \frac{\partial u_3}{\partial x_2} & \frac{\partial u_3}{\partial x_3} \end{pmatrix}, \tag{4}$$

where x is a length and u is an orthogonal velocity component, one may decompose $\mathbf{A}$ into strain, $\mathbf{S}$, and rotation, Ω, components

$$\begin{aligned} \mathbf{S} &= \frac{1}{2}(\mathbf{A} + \mathbf{A}^*), \\ \mathbf{\Omega} &= \frac{1}{2}(\mathbf{A} - \mathbf{A}^*). \end{aligned} \tag{5}$$

The enstrophy is then given by $\frac{1}{2}||\mathbf{\Omega}||^2$, where $||\ldots||$ is the Frobenius norm, that is, $||\mathbf{\Omega}|| = \sqrt{\mathrm{tr}(\mathbf{\Omega}\mathbf{\Omega}^*)}$, where the asterisk indicates the (complex) conjugate. The field in each panel of Figure 4 consists of 256×256 values obtained in a region of extent πh, where h is the channel half height. Here we generate surrogates for the scalar enstrophy field using a two-dimensional variant of the time series methods. An alternative approach is to study values for the enstrophy at a point relative to surrogates of the velocity gradient tensor itself, which requires a method for generating synthetic velocity gradient tensors (Keylock, 2017a). However, because the

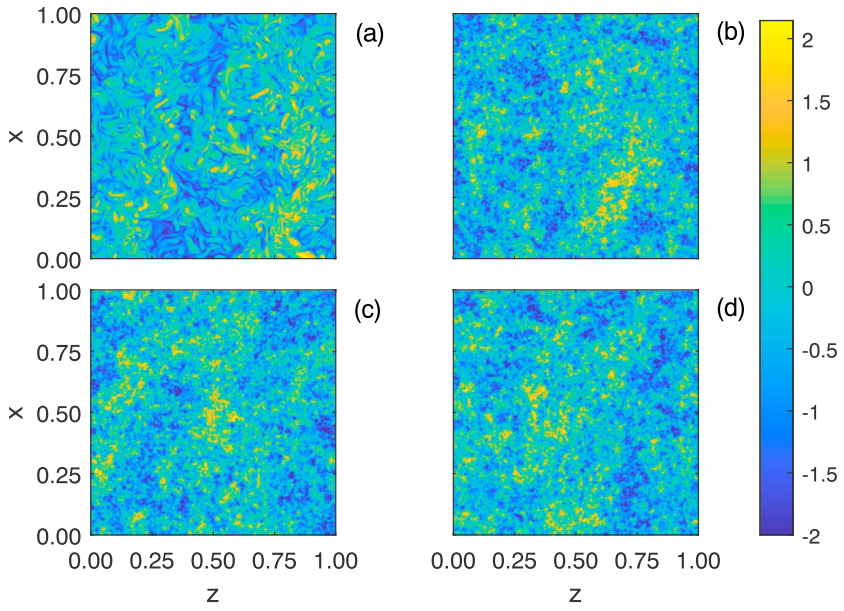

Figure 4. A plane indicating the enstrophy in a turbulent channel flow obtained from a direct numerical simulation is shown in (a). Three IAAFT surrogates are shown in panels (b) to (d). The scalebar indicates the values in terms of the base 10 logarithm of the dimensionless inverse time squared. The extent of the plane in the horizontal and transverse directions is $\pi \times h$, where h is the dimensionless channel half height.

mathematical basis for this approach (linear algebraic manipulation) is rather different to the Fourier- and wavelet-based methods focused upon in this paper, we do not discuss this technique further here.

Three example IAAFT surrogate data sets are shown in Figures 4b–4d, and it is clear that the surrogate data have the appropriate values and the appropriate sizes of the flow structures, while randomizing the locations of high and low enstrophy regions. The two-dimensional IAAFT was applied by Venema et al. (2006) for studying the properties of cloud fields and also used as a synthetic cloud generator. In our experience, the gradual convergence for the IAAFT algorithm becomes more necessary with 2-D data than for time series or one-dimensional spatial series. The approach we tend to adopt of the methods suggested by Venema et al. (2006) is to impose 20% of the original values in the amplitude adjustment step until convergence occurs with these values and then include an additional 20% of the values, continuing until all are included. With strongly extremal distributions for the original data, it may also facilitate convergence by working with the logarithm of the original data when building surrogates.

However, there is an issue with the use of IAAFT surrogates for imagery, which is that the sharp interfaces that define lineaments or delimit the foreground from the background are not perhaps of primary scientific interest but are nonetheless nonlinear features of the imagery. Thus, it would be preferable to have a surrogate algorithm that preserved these features, while randomizing the locations of the values in the image and other nonlinear properties. For example, because of the swirling nature of high enstrophy events, reentrant angles in the boundary between high and low enstrophy regions will reduce the typical separation between enstrophy maxima and minima compared to IAAFT surrogates that give more rounded regions (Figures 4b–4d). This is shown in the top row of Figure 5 where the probability is given for different values for the distance, D, defined for each local enstrophy maximum in the Figure 4a with a value greater than 10 (dimensionless) as the distance to the closest local enstrophy minimum with a value less than 0.1 (i.e., in the $\log_{10}$ units used, the spatial separation between 10^1 and 10^{-1} maxima and minima). It is clear that the IAAFT surrogates cannot mimic the observed D because of the absence of the nonlinear, swirling shape in the IAAFT surrogates. Results are given in both normalized (right panels) and nonnormalized (left panels) form as the exact number of local maxima is not necessarily constant between data and surrogates. That there is very little difference between the left and right panels shows that this is a minor effect.

Sharp interfaces are of particular importance for enstrophy fields because, when one writes the evolution equations for Ω^2 and $\mathbf{S}^2$, the total strain, the nonlocal contributions from the pressure Hessian impact directly upon the total strain field rather than the enstrophy field (Tsinober, 2001). The pressure Hessian is

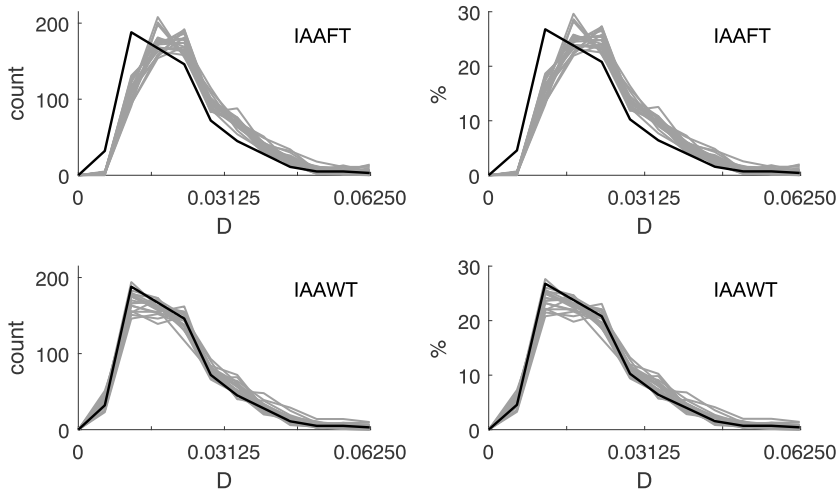

Figure 5. Values for the quantity D, the minimum distance between each regional enstrophy maximum and its closest minimum are shown in black, expressed both as raw counts (left-hand panels) and as a percentage of the number of identified maxima (right-hand panels). Values for 19 surrogates are shown as gray lines in each panel with IAAFT surrogates in the top row and IAAWT surrogates in the bottom row. Values for D are in terms of πh, where h is the dimensionless channel half height. IAAWT = iterated amplitude-adjusted wavelet transform; IAAFT = iterated amplitude-adjusted Fourier transform.

defined as $H^{(p)} = \frac{\partial^2 p}{\partial x_i \partial x_j}$ and is the relevant contributing term to the equation for the Lagrangian transport of the velocity gradient tensor, which is derived by taking the spatial gradient of the Navier-Stokes momentum equation

$$\frac{\partial}{\partial t}\mathbf{A} + \vec{\mathbf{u}} \cdot \nabla \mathbf{A} = -\mathbf{A}^2 - \mathbf{H}^{(p)} + \nu \nabla^2 \mathbf{A}. \tag{6}$$

This means that the strain field appears more diffused at high Reynolds number because the pressure contributions to the dynamics are integrated over a volume surrounding the selected spatial location (Ohkitani & Kishiba, 1995), while the evolution of the enstrophy is only a direct function of the local straining of enstrophy and the (small) viscous contribution.

A synthetic data method was recently proposed that retains the objective of preserving edge characteristics in the surrogates, and this is termed the iterated amplitude-adjusted wavelet transform (IAAWT) technique (Keylock, 2017b). Example surrogate fields generated with this algorithm are shown in Figure 6. The root-mean-square errors between the original image and the IAAFT and IAAWT surrogates are not too different (means of 10.59 and 8.88, respectively), but it is very clear that the IAAWT surrogates preserve the image structure to a much better degree, while still clearly randomizing the locations where enstrophy maxima and minima occur. That there is no significant difference between the original data and IAAWT surrogates in the bottom row of Figure 5 provides quantitative support for the better preservation of the shape of the high enstrophy regions.

The appealing thing about the IAAWT algorithm is that it is qualitatively identical to that for the IAAFT method. The difference arises in the mathematical transform that underpins it, with the Fourier transform replaced by a complex-valued discrete wavelet transform (Kingsbury, 2001). The consequence of this is that the multifractal structure of the original image is fixed to a level given by the convergence error tolerance in the algorithm, thus, fixes in place the image structure. Because the wavelet phases are still randomized, the locations of the maxima and minima, and other image properties are randomized subject to the multifractal preserving constraint as seen in Figures 6b–6d.

2.4. The Gradual Reconstruction Philosophy

Given a rejection of the null hypothesis using linearized, IAAFT surrogate data, the natural follow-up questions include

- How nonlinear are the data?
- How does the degree of nonlinearity vary as a function of the strength of the forcing applied to the system?

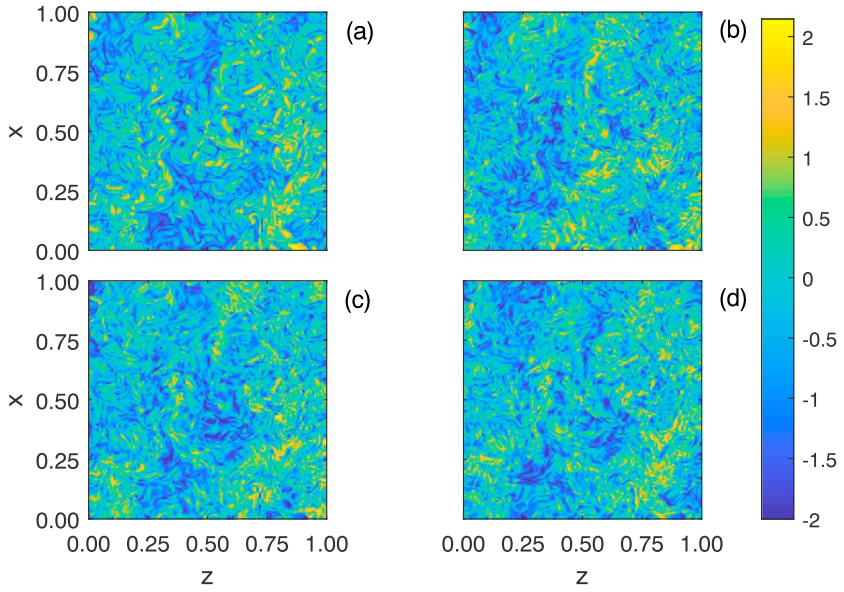

Figure 6. A plane indicating the enstrophy in a turbulent channel flow is shown in (a) obtained from a direct numerical simulation, and three iterated amplitude-adjusted wavelet transform surrogates are shown in panels (b) to (d). The scalebar indicates the values in terms of the base 10 logarithm of the dimensionless inverse time squared. The horizontal and vertical extent of the plane is $\pi \times h$, where h is the dimensionless channel half height.

In addition, if a set of surrogate data are generated that capture the nonlinearity in the original data at a given significance level, then they can potentially be used as a means to generate boot-strapped versions of those original data that preserve key nonlinear aspects.

The gradual reconstruction framework (Keylock, 2010) was introduced to tackle these questions. The idea is to define a continuum of surrogate data generators defined over, $0 \leq \rho \leq 1$, where $\rho = 0$ is a linearized, IAAFT surrogate and $\rho = 1$ is the original data set itself. The continuum is populated by constraining the degree of Fourier phase randomization from no constraints ($\rho = 0$) to fully constrained ($\rho = 1$). To do this requires a modification to the surrogate generation method, and the approach taken was to use a wavelet basis. Wavelets had been used for surrogate generation before in medical signal processing (Breakspear et al., 2004; Bullmore et al., 2001). However, these methods were really wavelet implementations of the AAFT or IAAFT methods. Recalling that a wavelet transform of a time series is a simultaneous, time-frequency decomposition (Foufoula-Georgiou et al., 1994; Mallat, 1999), it was first noted by Keylock (2006) that if the IAAFT algorithm is applied to each scale (frequency band) of an undecimated wavelet transform (one where there are N wavelet coefficients at each scale, $j = 1, \ldots, J$, where N is the number of samples in the time series), a randomization method can be developed that has no impact on the energy at a given scale, because the exact values are preserved, and is a legitimate realization of the wavelet transform at this scale because the preservation of the autocorrelation mimics the support of the wavelet. An alignment step may also be added that builds back in the interscale coupling (Prichard & Theiler, 1994). This idea was first used for testing aspects of signal stationarity, a topic subsequently developed using windowed Fourier transforms by Borgnat, Flandrin and coworkers (Borgnat & Flandrin, 2009; Borgnat et al., 2010). In a recent review paper on surrogate data methods, Lancaster et al. (2018) tested a number of methods at different noise levels with regards to their ability to detect the peak in the wavelet bispectrum between two test signals. This becomes more difficult with increasing noise because of the harmonics produced. They found that our approach (the wavelet IAAFT or WIAAFT method) performed very well in this regard (Figure 33 of their study).

Given that the wavelet decomposition provides information on the energy in the signal in time and frequency, by fixing in place the largest n squared wavelet coefficients of a total of $N \times J$ coefficients, such that a given proportion of the total energy of the original signal, ρ, is fixed, it is possible to appropriately constrain the degree of randomization of the Fourier phases when the wavelet transform is inverted (Keylock, 2007). This requires a wavelet transform that is exactly invertible, which is why we moved away from the continuous wavelet transform, commonly adopted in geophysics (Torrence & Compo, 1998) to the use of

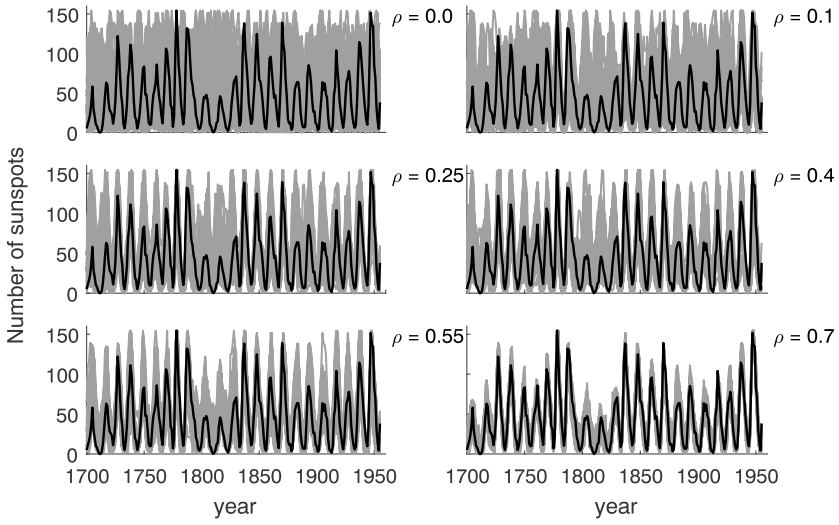

Figure 7. Surrogate data for 256 years of sunspot activity using gradual wavelet reconstruction. The original data are shown in black in each panel, while 39 surrogates are shown in gray, for the stipulated value of the gradual wavelet reconstruction parameter, ρ. Data taken from the original data product by WDC-SILSO, Royal Observatory of Belgium, Brussels, http://www.sidc.be/silso/home (Van der Linden & SIDC team, 2018).

the maximal overlap discrete wavelet transform (Percival & Walden, 2000), which is constructed using a hierarchy of filter banks in the same way as a discrete wavelet transform but is also undecimated like the continuous wavelet transform. See Keylock (2010) for further discussion on the merits of this transform for surrogate data generation.

An example of how surrogate data can be generated by this gradual wavelet reconstruction (GWR) method is given in Figure 7 based on 256 years of sunspot activity from 1700 to 1955. When $\rho = 0$ and 39 IAAFT surrogates are generated (top-left panel), the maximum value for sunspot activity could occur anywhere in time. As the phases become locked in as more and more wavelet coefficients are fixed in place, the maximum is first restricted to the peaks of the 11-year Hale cycle for sunspot activity ($\rho = 0.25$), is then prevented from occurring during the Dalton Minimum (1796–1820) at $\rho = 0.55$ and finally becomes locked to either one of the two largest values arising in 1947 and 1778, respectively ($\rho = 0.70$). Clearly, if there was a significant difference between data and surrogates based on testing with IAAFT surrogates, as $\rho \rightarrow 1$ the probability of rejecting a null hypothesis decreases. Thus, it is possible to define a value for ρ, denoted ρ_{crit} above which a significant difference is not observed. This can then characterize how nonlinear the data are, as explained more thoroughly in the next section.

Of course, this approach can be applied directly to two-dimensional data but may also be combined with the IAAWT algorithm introduced in section 2.3 rather than the IAAFT algorithm. In this case, it is the wavelet phase that is constrained rather than the Fourier phase in the randomization. This technique, termed gradual multifractal resolution was introduced by Keylock (2018) in a time series context and is used in a spatial sense for the first time in Figure 8. Here φ is the control parameter used to index gradual multifractal resolution in a directly analogous fashion to the manner than ρ indexes the GWR technique. The value for ϵ stated next to the panels for the surrogates is the root-mean-square error between the original image and the synthetic field shown. These values may be compared directly to the mean values for the IAAFT and IAAWT surrogates of 10.59 and 8.88 stated earlier, and it is clear that ϵ decays rapidly with φ as the value is only 1.96 at $\varphi = 0.25$. This explains the visual similarity between the original data (top-left panel) and the surrogates.

3. Undertaking Hypothesis Testing With Surrogate Data

3.1. Basic Procedures

The number of surrogate data needed for a particular statistical test depends on the significance level required and the nature of the test (one tailed or two tailed). A two-tailed test is one where the alternate hypothesis is that there is a significant difference, irrespective of its nature. A one-tailed test specifies the nature of this difference (greater than or less than). For a significance level, α, one needs to generate at least

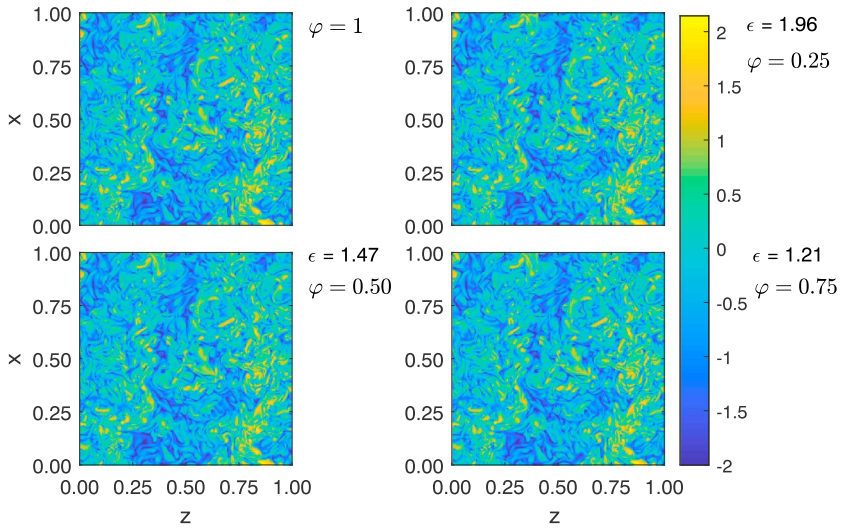

Figure 8. Gradual multifractal reconstruction of the enstrophy field according to the gradual multifractal reconstruction parameter, φ. The $\varphi = 0$ end of the continuum is the iterated amplitude-adjusted wavelet transform algorithm used in Figure 6, while $\varphi = 1$ are the original data. Intermediate values for φ are shown in the other panels together with ϵ, the root-mean-square error over the image for the difference in the values between the original field and the surrogate shown.

$(1/\alpha) - 1$ surrogates for the one-tailed case and $(2/\alpha) - 1$ for the two-tailed case. A greater number of surrogates improves the statistical power of the test.

Given some measure of the nonlinearity of the data set (see below), calculated for the data and surrogates, if the value for the data is greater or less than that for all the surrogates (assuming a two-tailed test), then a significant difference may be deemed to exist. Alternatively, if the values for the measure for the surrogates are distributed according to a particular distribution function, one may use the fitted distribution to define a distance between the original data and the surrogates. Thus, if the value for the measure for the surrogates was Gaussian, with M indicating the measure, o the original data and s the surrogates, the distance would be $z(M) = (|M^o - \overline{M^s}|)/\sigma(M^s)$, where the overbar is an average and σ is a standard deviation, and $z(M) > 1.96$ would imply a significant difference at the 5% level for a two-tailed test.

As an example, there are 19 surrogates used in each panel of Figure 5. With a null hypothesis that the average value for D is not significantly less for the original data than the linearized, IAAFT enstrophy images, we find we can reject this hypothesis at the 5% level but fail to reject it at this level using the IAAWT surrogates.

3.2. Measures Adopted

The key aspect to hypothesis testing with surrogate data is the formulation of the null hypothesis, which will imply some measure of the data, which needs to be calculated for both data and surrogates. A very useful overview of a number of different methods for characterizing nonlinear or complex systems was provided by Schreiber and Schmitz (1997), who examined methods closely associated with studying chaotic systems through embedding such as the correlation dimension (Grassberger & Procaccia, 1983) as well as nonlinear prediction error, higher-order cumulants (e.g., $\langle u_t u_{t-\tau} u_{t-2\tau} \rangle$ for lag τ), and what they called the time reversibility, which is also known as the time asymmetry

$$M_A = \frac{\left\langle |u_t - u_{t+\tau}|^3 \right\rangle}{\left\langle |u_t - u_{t+\tau}|^2 \right\rangle^{\frac{3}{2}}}. \tag{7}$$

They found that the nonlinear prediction error was an effective measure to choose in general, while Small and Judd (1998) emphasized correlation dimension as a useful test statistic. We would concur with this suggestion when testing for chaos in time series but would also suggest that M_A is particularly useful when studying the shape of a signal rather than the potential connection to underpinning dynamical equations. A relevant example is analyzing the asymmetry in river hydrographs due to the steep rising limb and more gradual recessional limb (Keylock, 2012).

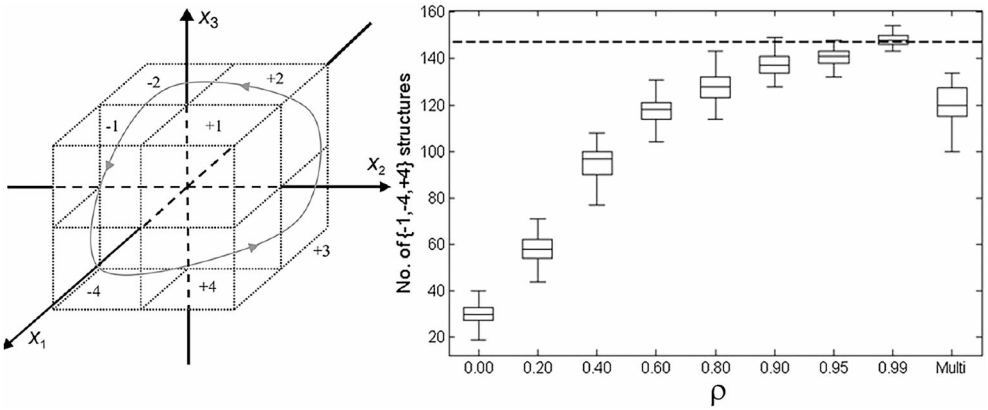

Figure 9. Turbulent structures evaluated from a flume experiment. The right-hand panel indicates the distribution functions for the values of surrogate data at different values for the control parameter ρ for studying the vortical structures in turbulence using a symbolic sequence method. The horizontal dashed line shows the actual number of observed instances of the identified critical sequence. "Multi" on the abscissa shows surrogates generated using the multivariate time series approach of Prichard and Theiler (1994). The left-hand panel shows the coordinate and convention systems for defining the vortical symbolic sequences based on the fluctuating part of three orthogonal velocity components. The typical path through these data for the case studied is shown by the gray line. This figure is a modified version of two figures that appeared in Keylock, Lane, et al. (2014). It is reproduced with permission of the AGU.

Note that sometimes a measure innately yields a single value for testing, such as the maximal Lyapunov exponent for a data set, while others are obtained for multiple points in time or space and the final summary measure is commonly a statistical moment of these multiple values. For example, when examining the shape of bedforms developed in an experimental study of river processes at different discharges, we took two approaches (Keylock et al., 2014). In the first, we used M_A applied to the increments of the neighboring elevations (values for z at $\tau = 1$), that is, in (7), $u_t - u_{t+\tau}$ was replaced by $z_t - z_{t+1}$ or $z_x - z_{x+1}$, depending if we sampled the time series of bedforms advected beneath a probe (t), or a spatial transect at the end of the experiment (x). In the second approach, we retained the values for the increment skewness as a function of spatial position,

$$M_A(x) = |z_x - z_{x+1}|^3. \tag{8}$$

As an alternative to a statistical expectation, we undertook a state-space embedding of this spatial series, $M_A(x)$ (Grassberger & Procaccia, 1983; Takens, 1981). A transportation distance metric (Moeckel & Murray, 1997) was then used to evaluate the distance in this embedding space between the probability distribution for the original data and each of the surrogates, $D^T(o, s_i)$, and that between each of the surrogates, $D^T(s_i, s_j)$, (see Small & Tse, 2004, for a discussion of how to select the embedding dimension and embedding lag in such methods). The existence of a significant difference in the central moment for the set of $D^T(o, s_i)$ compared to $D^T(s_i, s_j)$ then indicated the embedded distribution function of $M_A(x)$ for the data was significantly different to that for the surrogates.

3.3. Degree of Nonlinearity as Established by ρ_{crit}

With GWR, given the rejection of the null hypothesis for the IAAFT surrogates, the degree of nonlinearity may be established by constructing surrogates in the interval $0 < \rho < 1$ and then examining the existence of any significant difference as ρ decreases. Thus, using Figure 9 and excluding the right-most boxplot, which is generated in a different way, we see that at $\rho \geq 0.9$ there is no significant difference between data and surrogates because the boxplots of the values for the measure adopted overlap with the value for the original data as given by the dashed line. This high value for $\rho_{\text{crit}} = 0.9$ implies a complex system where a high degree of the phases in the original data must be fixed in the surrogates in order to replicate the observed behavior.

Another way in which ρ_{crit} can be used is to determine, for a particular data set, how sensitive to particular forms of nonlinearity the data are. For example, Keylock (2010) looked at the relation between the time series for the longitudinal and vertical components of the velocity in a neutral turbulence boundary layer

Table 1
Definition of Octants Derived From the Signs of the Fluctuating Velocities

Octant	u_1' Longitudinal	u_2' Transverse	u_3' Vertical
−1	+	−	+
+1	+	+	+
−2	−	−	+
+2	−	+	+
−3	−	−	−
+3	−	+	−
−4	+	−	−
+4	+	+	−

as a function of ρ. It was found that the phase synchronization between the velocity components, used as a measure of nonlinear association (Kreuz et al., 2007), resulted in $\rho_{crit} \sim 0.1$, while when applied to the Hölder series extracted from each time series (used as a means to determine the multifractal structure of the signal and the turbulence intermittency (Keylock, 2009), the value was $\rho_{crit} \sim 0.4$. Thus, greater phase information had to be retained in the surrogates to get the correct relation between the series in terms of the turbulence intermittency. One other distinct application of the method has been as a means to perturb inlet boundary conditions to a numerical simulation (Keylock et al., 2011). In this case, for $\rho \geq \rho_{crit}$ there is no significant difference between the value for a specific physical quantity at a particular position and that found for the original boundary conditions, and the sensitivity of the flow variable can be characterized as a consequence.

4. Example Applications

4.1. Inferring Turbulence Structure From Single-Point Instrumentation

As it provides a well-known source of nonlinearity in the physics of many geoscientific phenomena, turbulence is a common choice of application for both developing methods in nonlinear science (Frisch & Parisi, 1985; Takens, 1981) and using those methods to understand process dynamics. This includes the application of surrogate data methods (Basu et al., 2007; Keylock et al., 2015; Poggi & Katul, 2009). The example in this section is concerned with the common situation in practice where inferences on the dynamics are based on what may be determined from measurements of several velocity components at a single point. Near boundaries, it is common to decompose the flow into quadrants based on the signs of the fluctuating values for the vertical and horizontal velocity components (Bogard & Tiederman, 1986; Brunet et al., 1994; Willmarth & Lu, 1972) because these components generate the dominant Reynolds stresses in an idealized boundary layer. However, in the natural environment, eddy shedding from a rough boundary means that the transverse velocity component is also potentially significant. Thus, with a Reynolds decomposition applied to each velocity component to obtain a fluctuating velocity signal by mean subtraction, $u_i' = u_i - \bar{u}_i$, we can classify into octants according to Table 1 and Figure 9.

Keylock et al. (2014) grouped consecutive occurrences of the same type into an octant event and then studied the sequences of these events for flow structures generated downstream of a parallel-channel confluence (Best & Roy, 1991; Bradbrook et al., 1998) with an erodible bed. At the location where sediment entrainment was at a maximum, the dominant octant sequence was $\{-2, -1, -4, +4, +3, +2, -2\}$, as shown by the gray line in the left panel of Figure 9. The core to this structure was the $\{-1, -4, +4\}$ part as determined by the high Markovian transition probabilities between these states. This $\{-1, -4, +4\}$ sequence occurred 147 times in 300 s of data, which corresponded precisely to the dominant period to the shed turbulence, which was 2 s. This implied that the scour at this location was driven by a coherent structure consisting of a set of octants rather than driven by Reynolds stress (to which quadrant 4 contributes positively, but quadrant 1 contributes negatively), or the individual quadrants/octants. Various research questions then followed from these observations:

1. To what extent could 147 occurrences of the $\{-1, -4, +4\}$ have arisen by chance?
2. To what extent is the transverse component a significant part to this structure, or does it really reduce to a quadrant formulation?

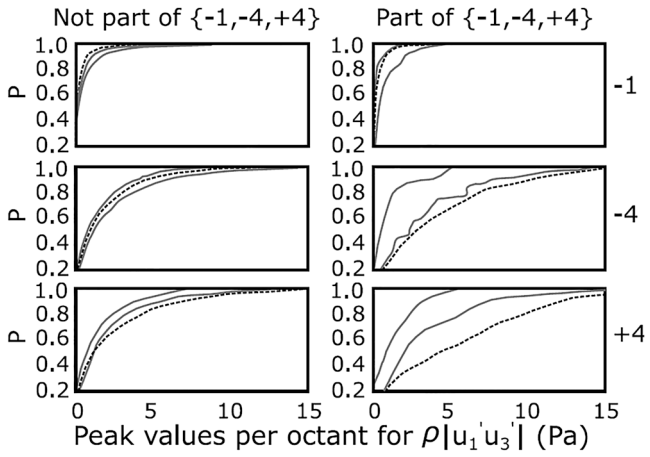

Figure 10. Cumulative probability plots of peak values per octant for the instantaneous stress (longitudinal-vertical component) truncated at a lower cumulative probability of 0.2 to enhance visibility. The gray lines in each plot delimit the lower and upper bounds of the results for 19 surrogates data sets generated with the IAAFT algorithm and the dashed line is obtained from the data. From top to bottom, octants −1, −4, and +4 are shown, with the left and right plots discriminating between occurrences within the −1, −4, and +4 sequence or not. This figure is modified from Keylock, Lane, et al. (2014). It is reproduced with permission of the AGU.

3. Is scour driven by the frequency of occurrence of the flow structures represented by this octant sequence, high stresses in general for these octants, or the stresses induced by flow structures with this octant sequence?

Concerning the former, the right-hand panel of Figure 9 shows the GWR results for the number of $\{-1, -4, +4\}$ occurrences based on 39 surrogates for each ρ. That is, the GWR algorithm was applied to each velocity component independently (so the cross correlation was allowed to build up as ρ increased) and then the octants and octant sequences were extracted. This panel also shows the results for IAAFT surrogates with cross correlation preserved in the algorithm as described in section 2.2, as indicated by "Multi" on the abscissa. It is clear that 147 occurrences could not arise by chance for a linear signal irrespective of if the cross-correlation structure is preserved, and $\rho_{\text{crit}} = 0.9$ at the 5% level emphasizes this.

In terms of the second question, it was found that the $\{1, 4\}$ quadrants occurred 318 times in 300 s. This does not match the dominant periodicity to the flow and, while it was significantly different to the maximum value from IAAFT surrogates (259), was only so by 23%, as opposed to the 268% difference for $\{-1, -4, +4\}$ occurrences at $\rho = 0$. Related to the second question are the results shown in Figure 10 for the peak instantaneous stresses in an octant event at the location for maximum scour. When comparing the −4 and +4 octants, which are clearly the same quadrant, the stress distributions are rather different compared to IAAFT surrogates. Considering the left-hand panels and therefore ignoring the octants that arise in the $\{-1, -4, +4\}$ sequence, while the distribution function for $\rho|u_1'u_3'|$ for +4 does yield more high stress events than for −4, the more obvious difference is that the distribution for the latter is not significantly different to IAAFT surrogates.

This figure also helps to answer the third question as it is clear that the −4 and +4 octants when occurring in the $\{-1, -4, +4\}$ sequence (right-hand panel) exert higher stresses than when they arise in a different sequence, both in absolute terms and relative to IAAFT surrogates. Thus, coherent structures are more important for sediment entrainment than particular individual events. However, what was also shown by Keylock, Lane, et al. (2014) was that when compared to IAAFT surrogates with the cross correlation preserved, no octant was significantly different to the surrogates irrespective of the octant sequence. Hence, the nonlinear structure of the velocity time series is critical to explaining scour at a parallel channel confluence, but the key role of the nonlinearity is to enhance the frequency of the $\{-1, -4, +4\}$ octant sequences that generate the peak stresses. These peak stresses can be captured by linear IAAFT surrogates with the appropriate cross-correlation structure, but the frequency of these structures cannot.

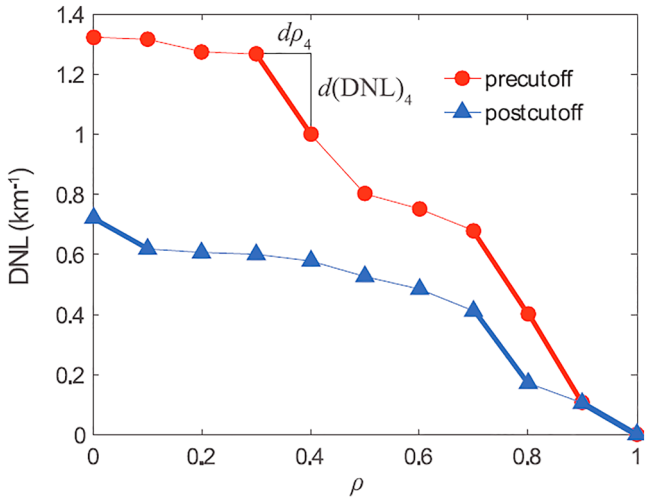

Figure 11. The degree of nonlinearity (*DNL*) of river meander curvature sequences is shown as a function of ρ for precutoff (circles) and postcutoff (triangles) centerline curvatures, where ρ is the GWR parameter controlling the amount of nonlinearity preserved in the surrogates. For both the precutoff and postcutoff curvatures, three bold line segments highlight the three largest change increments that contribute the most to the total *DNL* (defined at $\rho = 0$), as illustrated for the fourth increment, $i = 4$. This figure is modified from Schwenk and Foufoula-Georgiou (2017). It is reproduced with permission of the AGU.

4.2. Nonlinearity in River Meander Curvature

Recent work by Schwenk and Foufoula-Georgiou (2017) has looked at the extent to which river meander bend curvature profiles (curvature as a function of distance downstream) extracted from aerial photography, satellite imagery, and numerical simulation encode nonlinear variability in channel form as a consequence of the processes acting. As described in general terms in section 3.2, their approach was based on a state space embedding of the curvature time series and an evaluation of the distance between embedded probability distributions using the transportation distance, D_T. The definition of the degree of nonlinearity, *DNL*, is then simply the absolute difference in mean values:

$$DNL = |\overline{D^T(o, s_i)} - \overline{D^T(s_i, s_j)}|. \tag{9}$$

The surrogates underpinning *DNL* were generated according to GWR and the results for the numerical simulation are shown in Figure 11 for two time periods: after 450 years of simulation and before meander cutoffs began to occur (square symbols) and after 7,000 years when the channel was in an approximate steady state (triangles). In both cases, for $0.7 \leq \rho \leq 0.9$ the *DNL* value drops dramatically indicating that surrogates for both signals require a significant proportion of the phase information to be fixed in place before they can begin to mimic the characteristics of the modeled meander bend evolution. What is also very clear is the higher *DNL* values for the precutoff time frame, highlighting the important role of meander bend cutoffs in reducing the nonlinearity of the curvature series (Camporeale et al., 2008), even though the sudden changing of the channel slope and, thus, discharge and bed shear stresses can promote nonlinear effects. It is also the case that for both data sets the three segments highlighted in bold (thus, 30% of the wavelet energy) accounted for just over 60% of the total *DNL*. However, an examination of which coefficients in the wavelet transform were fixed in these segments found that in the precutoff case these coefficients corresponded to the fully developed meander bends with only 5% of the *DNL* arising from the small and large scales associated with cutoff activity. In contrast, for the later time period, 46% of the *DNL* for these segments was associated with these cutoff-related scales.

As a consequence, this study succeeded in two important ways relative to previous work:

- The use of a hypothesis testing framework resulted in an ability to detect and quantify significant form nonlinearity due to meandering. Thus, it was determined for the first time that process nonlinearities are encoded in meander planform structure.
- The use of GWR meant that the manner in which cutoffs reduce planform nonlinearity could be resolved, and the answer was dependent on the spatial configuration of meanders. Because cutoffs spread the

spectral energy of the curvature signal into a range of frequencies, the energy represented by fully developed meander bends was reduced. However, this was only an 8% reduction, while the contribution of these scales to the *DNL* declined by 41%.

5. Summary and Conclusion

Linearized surrogate data such as those obtained with the IAAFT algorithm are a very useful tool for the geoscientist interested in elucidating the importance of nonlinear effects in a great many fields. While there has typically been an emphasis on the study of time or spatial series data, there is no reason that imagery or higher dimensionality data cannot be studied in this way, as we have shown with the analysis of enstrophy in a turbulent channel flow. However, many features in imagery and higher dimensional analysis have sharp interfaces that are typically not realized from a linear realization of the data. This means that more nuanced hypotheses about properties of such data need to fix the interfaces in place in the surrogates to preserve such elements. The recently introduced IAAWT algorithm (Keylock, 2017b) does this by preserving multifractal structure and may be used, for example, for studying the coupling between velocity and intermittency in turbulence time series (Keylock et al., 2012).

In this paper, we have also outlined the utility of the gradual reconstruction philosophy for generating surrogate data to populate a continuum from a starting null hypothesis (such as the data are a sample from a linear process as tested with the IAAFT algorithm) to the data themselves. Not only does this provide a means to characterize how nonlinear a data set is from the determination of ρ_{crit} but then studying how ρ_{crit} varies as a function of an experimental control variable can result in scientific conclusions about how the importance of nonlinearity changes (Keylock et al., 2014).

While this gradual reconstruction framework is a general formalism, it is tailored to preservation of statistical measures rather than dynamical measures of the data at the $\rho, \varphi = 0$ limit (i.e., the Fourier amplitudes or multifractal structure rather than, say the maximal Lyapunov exponent). This means that in order to generate synthetic data that preserve the relevant dynamical properties, it is necessary to first find the critical value for the parameter. This may take some time and for certain measures to which the data are very sensitive may constrain variability to relatively few degrees of freedom, particularly if the original values for the data are retained in the surrogates. In such cases it would be advantageous for surrogate data methods to exist that are targeted at preserving such dynamical measures directly. To do this for multiple time series then requires the design of surrogate generation methods where the measure of coupling between the time series to be preserved is more refined than the Fourier phase difference that is employed when the null hypothesis is that the data are realizations from a linear system (Prichard & Theiler, 1994). For example, an obvious extension to the IAAWT algorithm (Keylock, 2017b) would be to preserve the wavelet phase differences between series so that the locations of "bursts" of activity detected across more than one time series can be fixed in the synthetic data.

One important recent innovation in the analysis of nonlinear systems is the use of network representations of system dynamics (Boers et al., 2015; Tejedor et al., 2018; Zhang et al., 2017). Such methods have shown that they are able to capture dynamical properties of continuous dynamical systems, such as the maximal Lyapunov exponent (McCullough et al., 2015). However, an as yet underexplored area is the fusion of network representations and surrogate data and the manner in which the network evolves, both in terms of the internode connectivity and the strength of interaction between nodes. This is also an area where a gradual reconstruction approach is of potential utility, in that it can be used to generate time series for which the internode connectivity is identical to the original data but for which the strength of interaction still varies. Thus, the surrogates provide a means to study the sensitivity of the dynamics with respect to data with identical network topology.

Another area that may benefit from a greater fusion between surrogate data methods and traditional approaches is ensemble model forecasting. Model ensembles are used as a means to deal with the uncertainty in model formulations and data, as well as the nonlinear dynamics of the associated systems, with ensemble spread used as a measure of model uncertainty (Grimit & Mass, 2007). While recent work has considered the statistical characterization of the error in such ensembles (Van Schaeybroeck & Vannitsem, 2016), how the variance derived from such an ensemble relates to that for appropriately constrained surrogate outputs from one model is not a topic that has been explored sufficiently. However, what this permits is a capability to assess the variance of the ensemble, the surrogates for each ensemble member and that for all

surrogates of all ensemble members. This provides alternate perspectives on the way in which the ensemble spread may be evaluated because a variance can be associated with each ensemble member. Application of extreme value theory in this context is also relevant, and we touched on this issue in work that placed confidence intervals on generalized extreme value distribution fits to hydrological maxima using constrained surrogates (Keylock, 2012). What will perhaps be of increasing interest in the future is the application of extreme value theory to the Poincaré returns of the strange attractor and how these then impact on observables like precipitation and temperature (Faranda et al., 2017). An interesting potential development is then the construction of appropriate synthetic data directly within phase space, rather than for (multiple) time series of system outputs. This difference in philosophy can be seen in work in turbulence, where recent approaches model the dynamics in the space of the invariants of the tensor (Wilczek & Meneveau, 2014). Thus, surrogate data for the tensor (Keylock, 2017a) are of greater utility in this case than those for time series outputs, such as the IAAFT and IAAWT algorithms.

Hence, as nonlinear analysis methods develop, and new forms of data become more readily available, the development of surrogate data methods needs to keep pace to facilitate the statistical interpretation of experimental or modelling results. More generally, it is hoped that the existing methods presented and discussed in this paper can be employed more widely in geoscience to enhance understanding of nonlinear systems.

Acronyms

AAFT amplitude-adjusted Fourier transform
CWT continuous wavelet transform
DNL degree of nonlinearity
GWR gradual wavelet reconstruction
IAAFT iterated AAFT
IAAWT iterated amplitude-adjusted wavelet transform

Acknowledgments
The author is grateful to Jon Schwenk for providing Figure 11. This work was partially supported by NERC grant NE/F00415X/1, EPSRC grant EP/K007688/1, and Royal Academy of Engineering/Leverhulme Senior Research Fellowship LTSRF1516-12-89. The sunspot data used in the paper are available from http://www.sidc.be/sunspot-data/ and the Johns Hopkins Turbulence Database is at http://turbulence.pha.jhu.edu.

References

Aguirre, L. A., & Billings, S. A. (1995). Identification of models for chaotic systems from noisy data: Implications for performance and nonlinear filtering. *Physica D: Nonlinear Phenomena*, *85*, 239–258.

Arens, S. J. T., Sullivan, P. F., & Welker, J. M. (2008). Nonlinear responses to nitrogen and strong interactions with nitrogen and phosphorus additions drastically alter the structure and function of a high arctic ecosystem. *Journal of Geophysical Research*, *113*, G03S09. https://doi.org/10.1029/2007JG000508

Basu, S., Foufoula-Georgiou, E., Lashermes, B., & Arneodo, A. (2007). Estimating intermittency exponent in neutrally stratified atmospheric surface layer flows: A robust framework based on magnitude cumulant analysis and surrogate analyses. *Physics of Fluids*, *19*, 115102.

Best, J. L., & Roy, A. G. (1991). Mixing-layer distortion at the confluence of channels of different depth. *Nature*, *350*, 411–413.

Boers, N., Donner, R. V., Bookhagen, B., & Kurths, J. (2015). Complex network analysis helps to identify impacts of the El Niño Southern Oscillation on moisture divergence in South America. *Climate Dynamic*, *45*(3-4), 619–632.

Bogard, D. G., & Tiederman, W. G. (1986). Burst detection with single-point velocity measurements. *Journal of Fluid Mechanics*, *162*, 389–413.

Borgnat, P., & Flandrin, P. (2009). Stationarization via surrogates. *Journal of Statistical Mechanics*, *1*, P01001.

Borgnat, P., Flandrin, P., Honeine, P., Richard, C., & Xiao, J. (2010). Testing stationarity with surrogates: A time-frequency approach. *IEEE Transactions on Signal Processing*, *58*(7), 3459–3470.

Bradbrook, K. F., Biron, P. M., Lane, S. N., Richards, K. S., & Roy, A. G. (1998). Investigation of controls on secondary circulation in a simple confluence geometry using a three-dimensional numerical model. *Hydrological Processes*, *12*, 1371–1396.

Bradley, E., & Kantz, H. (2015). Nonlinear time-series analysis revisited. *Chaos*, *25*(9), 097610.

Breakspear, M., Brammer, M. J., Bullmore, E. T., Das, P., & Williams, L. M. (2004). Spatiotemporal wavelet resampling for functional neuroimaging data. *Human Brain Mapping*, *23*, 1–25.

Brockwell, P. J., & Davis, R. A. (2013). *Time series: Theory and methods*. Verlag New York: Springer Science & Business Media.

Brunet, Y., Finnigan, J. J., & Raupach, M. R. (1994). A wind tunnel study of air flow in waving wheat: Single-point velocity statistics. *Boundary-Layer Meteorology*, *70*, 95–132.

Bullmore, E., Long, C., Suckling, J., Fadili, J., Calvert, G., Zelaya, F., et al. (2001). Colored noise and computational inference in neurophysiological (fMRI) time series analysis: Resampling methods in time and wavelet domains. *Human Brain Mapping*, *12*, 61–78.

Camporeale, C., Perucca, E., & Ridolfi, L. (2008). Significance of cutoff in meandering river dynamics. *Journal of Geophysical Research*, *113*, F01001. https://doi.org/10.1029/2006JF000694

Chekroun, M. D., Simonnet, E., & Ghil, M. (2011). Stochastic climate dynamics: Random attractors and time-dependent invariant measures. *Physica D: Nonlinear Phenomena*, *240*, 1685–1700.

Church, M., & Slaymaker, O. (1989). Disequilibrium of holocene sediment yield in glaciated British Columbia. *Nature*, *337*(6206), 452–454.

Faranda, D., Messori, G., Alvarez-Castro, M. C., & Yiou, P. (2017). Dynamical properties and extremes of Northern Hemisphere climate fields over the past 60 years. *Nonlinear Processes in Geophysics*, *24*, 713–725. https://doi.org/10.5194/npg-24-713-2017

Foufoula-Georgiou, E., Kumar, P., & Chui, C. K. (1994). *Wavelets in geophysics* (pp. 373). New York: Academic Press.

Frisch, U., & Parisi, G.The singularity structure of fully developed turbulence. In M. Ghil, R. Benzi, & G. Parisi (Eds.), *Turbulence and predictability in geophysical fluid dynamics and climate dynamics* (pp. 84–88). North Holland: Proceedings of the International School of Physics "Enrico Fermi", (1985).

Goncharenko, L. P., Chau, J. L., Liu, H.-L., & Coster, A. J. (2010). Unexpected connections between the stratosphere and ionosphere. *Geophysical Research Letters*, *37*, L10101. https://doi.org/10.1029/2010GL043125

Graham, J., Kanov, K., Yang, X. I. A., Lee, M. K., Malaya, N., Lalescu, C. C., et al. (2016). A Web services-accessible database of turbulent channel flow and its use for testing a new integral wall model for LES. *Journal of Turbulence*, *17*(2), 181–215.

Grassberger, P., & Procaccia, I. (1983). Measuring the strangeness of strange attractors. *Physica D: Nonlinear Phenomena*, *9*, 189–208. https://doi.org/10.1016/0167-2789(83)90298-1

Grimit, E. P., & Mass, C. F. (2007). Measuring the ensemble spread-error relationship with a probabilistic approach: Stochastic ensemble results. *Monthly Weather Review*, *135*, 203–221. https://doi.org/10.1175/MWR3262.1

Hannachi, A., Straus, D. M., Franzke, C. L., Corti, S., & Woollings, T. (2017). Low frequency nonlinearity and regime behavior in the Northern Hemisphere extratropical atmosphere. *Reviews of Geophysics*, *55*, 199–234. https://doi.org/10.1002/2015RG000509

Hénon, M. (1976). A two-dimensional mapping with a strange attractor. *Communications in Mathematical Physics*, *50*(1), 69–77.

Kantz, H., & Schreiber, T. (1997). *Nonlinear time series analysis*. New York, NY, USA: Cambridge University Press.

Kaplan, J., & Yorke, J. (1979). Chaotic behavior of multidimensional difference equations. In H. O. Peitgen (Ed.), *Functional differential equations and the approximation of fixed points. Lecture notes in mathematics* (Vol. 730, pp. 204–227). Berlin: Springer.

Keylock, C. J. (2006). Constrained surrogate time series with preservation of the mean and variance structure. *Physical Review E*, *73*, 036707.

Keylock, C. J. (2007). A wavelet-based method for surrogate data generation. *Physica D: Nonlinear Phenomena*, *225*, 219–228.

Keylock, C. J. (2009). Evaluating the dimensionality and significance of active periods in turbulent environmental flows defined using Lipshitz/Hölder regularity. *Environmental Fluid Mechanics*, *9*, 509–523.

Keylock, C. J. (2010). Characterizing the structure of nonlinear systems using gradual wavelet reconstruction. *Nonlinear Processes in Geophysics*, *17*, 615–632.

Keylock, C. J. (2012). A resampling method for generating synthetic hydrological time series with preservation of cross-correlative structure and higher order properties. *Water Resources Research*, *48*, W12521. https://doi.org/10.1029/2012WR011923

Keylock, C. J. (2017a). Synthetic velocity gradient tensors and the identification of statistically significant aspects of the structure of turbulence. *Physical Review Fluids*, *2*, 084607.

Keylock, C. J. (2017b). Multifractal surrogate-data generation algorithm that preserves pointwise holder regularity structure, with initial applications to turbulence. *Physical Review E*, *95*, 032123.

Keylock, C. J. (2018). Gradual multifractal reconstruction of time-series: Formulation of the method and an application to the coupling between stock market indices and their Hölder exponents. *Physica D: Nonlinear Phenomena*, *368*, 1–9.

Keylock, C. J., Lane, S. N., & Richards, K. S. (2014). Quadrant/octant sequencing and the role of coherent structures in bed load sediment entrainment. *Journal of Geophysical Research: Earth Surface*, *19*, 264–286. https://doi.org/10.1002/2012JF002698

Keylock, C. J., Nishimura, K., & Peinke, J. (2012). A classification scheme for turbulence based on the velocity-intermittency structure with an application to near-wall flow and with implications for bedload transport. *Journal of Geophysical Research*, *117*(F01037). https://doi.org/10.1029/2011JF002127

Keylock, C. J., Singh, A., & Foufoula-Georgiou, E. (2014). The complexity of gravel-bed river topography examined with gradual wavelet reconstruction. *Journal of Geophysical Research: Earth Surface*, *119*, 682–700. https://doi.org/10.1002/2013JF002999

Keylock, C. J., Stresing, R., & Peinke, J. (2015). Gradual wavelet reconstruction of the velocity increments for turbulent wakes. *Physics of Fluids*, *27*, 025104.

Keylock, C. J., Tokyay, T. E., & Constantinescu, G. (2011). A method for characterising the sensitivity of turbulent flow fields to the structure of inlet turbulence. *Journal of Turbulence*, *12*, N45. https://doi.org/10.1080/14685248.2011.636047

Kingsbury, N. (2001). Complex wavelets for shift invariant analysis and filtering of signals. *Applied and Computational Harmonic Analysis*, *10*, 234–253.

Kreuz, T., Mormann, F., Andrzejak, R. G., Kraskov, A., Lehnertz, K., & Grassberger, P. (2007). Measuring synchronization in coupled model systems: A comparison of different approaches. *Physica D: Nonlinear Phenomena*, *225*, 29–42.

Lancaster, G., Iatsenko, D., Piddea, A., Ticcinellia, V., & Stefanovskaa, A. (2018). Surrogate data for hypothesis testing of physical systems. *Physics Reports*, *748*, 1–60.

Li, Y., Perlman, E., Wan, M., Yang, Y., Burns, R., Meneveau, C., et al. (2008). A public turbulence database cluster and applications to study Lagrangian evolution of velocity increments in turbulence. *Journal of Turbulence*, *9*(31), 1–29.

Lorenz, E. N. (1963). Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, *20*, 130–141.

Mallat, S. (1999). *A wavelet tour of signal processing*. New York: Academic Press.

McCullough, M., Small, M., Stemler, T., & Iu, H. H.-C. (2015). Time lagged ordinal partition networks for capturing dynamics of continuous dynamical systems. *Chaos*, *25*(053101). https://doi.org/10.1063/1.4919075

Moeckel, R., & Murray, B. (1997). Measuring the distance between time series. *Physica D: Nonlinear Phenomena*, *102*, 187–194.

Ohkitani, K., & Kishiba, S. (1995). Nonlocal nature of vortex stretching in an inviscid fluid. *Physics of Fluids*, *7*(2), 411–421.

Paniconi, C., & Putti, M. (2015). Physically based modeling in catchment hydrology at 50: Survey and outlook. *Water Resources Research*, *51*, 7090–7129. https://doi.org/10.1002/2015WR017780

Percival, D. B., & Walden, A. T. (2000). *Wavelet methods for times series analysis*. Cambridge, U.K.: Cambridge University Press.

Pierini, S., Chekroun, M. D., & Ghil, M. (2018). The onset of chaos in nonautonomous dissipative dynamical systems: A low-order ocean-model case study. *Nonlinear Processes in Geophysics*, *25*, 671–692. https://doi.org/10.5194/npg-25-671-2018

Poggi, D., & Katul, G. (2009). Flume experiments on intermittency and zero-crossing properties of canopy turbulence. *Physics of Fluids*, *21*(6), 065103.

Pomeau, Y., & Manneville, P. (1980). Intermittent transition to turbulence in dissipative dynamical systems. *Communications in Mathematical Physics*, *74*, 189–197.

Prichard, D., & Theiler, J. (1994). Generating surrogate data for time series with several simultaneously measured variables. *Physical Review Letters*, *73*, 951–954.

Rössler, O. E. (1976). An equation for continuous chaos. *Physics Letters A*, *57*(5), 397–398.

Schreiber, T., & Schmitz, A. (1996). Improved surrogate data for nonlinearity tests. *Physical Review Letters*, *77*, 635–638.

Schreiber, T., & Schmitz, A. (1997). Discrimination power of measures for nonlinearity in a time series. *Physical Review E*, *55*, 5443–5447.

Schwenk, J., & Foufoula-Georgiou, E. (2017). Are process nonlinearities encoded in meandering river planform morphology? *Journal of Geophysical Research: Earth Surface*, *122*, 1534–1552. https://doi.org/10.1002/2016JF003929

Simpson, G. G. (1963). Historical science. In C. C. Albritton (Ed.), *The fabric of geology* (pp. 24–48). Reading, MA: Freeman-Cooper.

Small, M., & Judd, K. (1998). Correlation dimension: A pivotal statistic for non-constrained realizations of composite hypotheses in surrogate data analysis. *Physica D: Nonlinear Phenomena*, *120*, 386–400.
Small, M., & Tse, C. K. (2004). Optimal embedding parameters: A modelling paradigm. *Physica D: Nonlinear Phenomena*, *194*, 283–296.
Takens, F. (1981). Detecting strange attractors in turbulence, *Lecture notes in mathematics* (Vol. 898, pp. 366–381). Berlin: Springer.
Tejedor, A., Longjas, A., Passalacqua, P., Moreno, Y., & Foufoula-Georgiou, E. (2018). Multiplex networks: A framework for studying multi-process multi-scale connectivity via coupled-network theory with an application to river deltas. *Geophysical Research Letters*, *45*, 9681–9689. https://doi.org/10.1029/2018GL078355
Theiler, J., Eubank, S., Longtin, A., Galdrikian, B., & Farmer, J. D. (1992). Testing for nonlinearity in time-series: The method of surrogate data. *Physica D: Nonlinear Phenomena*, *58*, 77–94.
Torrence, C., & Compo, C. P. (1998). A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society*, *79*, 61–78.
Tsinober, A. (2001). Vortex stretching versus production of strain/dissipation. In J. C. R. Hunt, & J. C. Vassilicos (Eds.), *Turbulence structure and vortex dynamics* (pp. 164–191): Cambridge University Press.
Tsinober, A. (2013). *The essence of turbulence as a physical phenomenon* (pp. 169). Netherlands: Springer.
Van Schaeybroeck, B., & Vannitsem, S. (2016). A probabilistic approach to forecast the uncertainty with ensemble spread. *Monthly Weather Review*, *144*, 451–468. https://doi.org/10.1175/MWR-D-14-00312.1
Van der Linden, R. A. M., & SIDC team (2018). Online catalogue of the sunspot index. http://www.sidc.be/sunspot-data/
Venema, V., Ament, F., & Simmer, C. (2006). A stochastic iterative amplitude adjusted fourier transform algorithm with improved accuracy. *Nonlinear Processes in Geophysics*, *13*, 321–328.
Venema, V., Meyer, S., Garcia, S. G., Kniffka, A., Simmer, C., Crewell, S., et al. (2006). Surrogate cloud fields generated with the iterative amplitude adapted Fourier transform algorithm. *Tellus A*, *58*, 104–120.
Wilczek, M., & Meneveau, C. (2014). Pressure Hessian and viscous contributions to velocity gradient statistics based on Gaussian random fields. *Journal of Fluid Mechanics*, *756*, 191–225.
Willmarth, W. W., & Lu, S. S. (1972). Structure of the Reynolds stress near the wall. *Journal of Fluid Mechanics*, *55*(1), 65–92.
Wolf, A., Swift, A., Jack, B., Swinney, H. L., & Vastano, J. A. (1985). Determining Lyapunov exponents from a time series. *Physica D: Nonlinear Phenomena*, *16*(3), 285–317. https://doi.org/10.1016/0167-2789(85)90011-9
Zhang, J., Zhou, J., Tang, M., Guo, H., Small, M., & Zou, Y. (2017). Constructing ordinal partition transition networks from multivariate time series. *Scientific Reports*, *7*, 7795. https://doi.org/10.1038/s41598-017-08245-x