Multifractal surrogate-data generation algorithm that preserves pointwise Hölder regularity structure, with initial applications to turbulence

C. J. Keylock*
Sheffield Fluid Mechanics Group and Department of Civil and Structural Engineering, University of Sheffield, Mappin Street, Sheffield S1 3JD, United Kingdom
(Received 21 December 2016; published 13 March 2017)

An algorithm is described that can generate random variants of a time series while preserving the probability distribution of original values and the pointwise Hölder regularity. Thus, it preserves the multifractal properties of the data. Our algorithm is similar in principle to well-known algorithms based on the preservation of the Fourier amplitude spectrum and original values of a time series. However, it is underpinned by a dual-tree complex wavelet transform rather than a Fourier transform. Our method, which we term the iterated amplitude adjusted wavelet transform can be used to generate bootstrapped versions of multifractal data, and because it preserves the pointwise Hölder regularity but not the local Hölder regularity, it can be used to test hypotheses concerning the presence of oscillating singularities in a time series, an important feature of turbulence and econophysics data. Because the locations of the data values are randomized with respect to the multifractal structure, hypotheses about their mutual coupling can be tested, which is important for the velocity-intermittency structure of turbulence and self-regulating processes.

DOI: 10.1103/PhysRevE.95.032123

I. INTRODUCTION

Given a measurement of a complex and potentially nonlinear phenomenon in the form of a time series, it is often important and insightful to test the statistical significance of metrics that attempt to characterize and summarize the underlying data structure. Because a complex process does not necessarily have a Gaussian or even a known prior distribution function, conventional statistical analyses are not always readily applicable. Hence, the surrogate-data approach was developed within nonlinear physics in the 1990s as a means to treat this class of problems [1–5]. This has proven to be a popular methodology, with applications spanning many disciplines, including medical physics [6–8], fluid mechanics [9,10], and the geosciences [11,12].

Perhaps the best known algorithm is the iterated amplitude adjusted Fourier transform (IAAFT) method [3], which gives surrogates constrained to the original values in the data while also preserving the Fourier amplitude spectrum accurately. Consequently, differences between the data and surrogates will reside in a specific structure to the Fourier phases of the original data that cannot be replicated by chance in the synthetic surrogate data. Thus, hypotheses concerning the possible nonlinear properties of an observed time series may be answered by comparing the values of some metric that characterizes the nonlinearity (e.g., the correlation dimension, maximal Lyapunov exponent, derivative skewness, nonlinear prediction error) [13] between the data and the surrogates. Thus, having selected a statistical significance level α and assuming a two-tailed statistical test so that $\frac{2}{\alpha} - 1$ surrogates are generated, if the value of the metric for the data falls outside the range for the surrogates, then a significant difference exists.

A development of this general approach is termed gradual wavelet reconstruction (GWR) and envisages the original data to be at one end of a continuum ($\rho = 1$) and phase-randomized IAAFT data to be at the other ($\rho = 0$). Surrogate data along this continuum are generated by fixing in place some of the phase information in the original time series [14]. This is achieved by adopting a wavelet transform, and indeed, wavelets and windowed Fourier transform methods have been used in the past as a means to generate IAAFT-like surrogates [15] and for constraining the randomization in other ways [16,17]. With surrogates generated at different values for ρ it is then possible to generate bootstrapped confidence limits for data that preserve a sufficient degree of the nonlinearity present in the time series [18,19] or to test relative degrees of complexity on a given metric by determining the value for ρ above which there is no longer a significant difference between data and surrogates [20].

Given that one way in which a time series may depart from its IAAFT surrogates is through the presence of multifractality, a different approach to GWR for this case is to construct surrogates that preserve the multifractal characteristics of a time series [21]. Such surrogates can be used to generate bootstrapped variants of multifractal phenomena such as human cardiac dynamics [8], earth surface topography [22], precipitation [23], stock market price fluctuations [24], and quantum Hall systems [25]. They may also be used as a means for testing if the observed degree of multifractality departs significantly from that of a constant pointwise Hölder exponent. However, the existing multifractal algorithm of Paluš [21] does not impose the additional constraint that the values of the original data set are preserved, making the algorithm more akin to the original surrogate-data methods of Theiler and coworkers [1,2] rather than the IAAFT algorithm, where differences have to reside in the phase information.

This paper gives an algorithm that not only incorporates this feature but also approximates the ordering in time of the Hölder exponents for the original data. We demonstrate the utility of the method by applying it to the detection of

*Also at Department of Aeronautics, Imperial College London, Exhibition Road, London SW7 2AZ, United Kingdom; c.keylock@sheffield.ac.uk

oscillating singularities in time series [26,27], which is relevant to furthering our understanding of turbulence [28,29], and to analyzing turbulent systems where there is a dependence between the intermittency and the macroscale velocity, as originally discussed by Kolmogorov [30] and subsequently shown to be of relevance in various contexts [31–33]. Before this algorithm is explained and applied, we briefly review essential concepts of multifractality and Hölder exponents, leading to methods for their estimation. We then review surrogate algorithms for IAAFT surrogate generation as well as the method introduced by Paluš [21]. After presentation of our method, we test its relative effectiveness and then provide some example applications with relevance to turbulence physics.

II. HÖLDER REGULARITY

A. Definition of pointwise and local Hölder regularity

In this paper we primarily focus on the pointwise Hölder regularity $\alpha_p(t)$ of a function, the variation in which is approximated by methods for characterizing multifractality such as structure functions [34,35], universal multifractal exponents [36], and wavelet transform modulus maxima methods [37,38]. However, in addition, one may define the local Hölder regularity $\alpha_L(t)$ to characterize oscillating singularities [39,40], which are not distinguished from sharp singularities when using $\alpha_p(t)$, and can go further and define a function of such exponents, calculating all forms of regularity at a point [40]. We make use of $\alpha_L(t)$ when looking at the statistical detection of an oscillating singularity, as described below.

Given a signal u_t and a position t_0 and assuming that $\alpha_p \in \{0,\ldots,1\}$, then we seek values β_p for which

$$|u(t) - u(t_0)| \leqslant c_p |t - t_0|^{\beta_p}. \tag{1}$$

The value for $\alpha_p(t_0)$ is then the supremum of this set of legitimate β_p values. Following Seuret and Lévy Véhel [39], to define α_L, we commence by considering a ball $B(t_0,R)$ with a center at t_0 and radius R. If the constant c_L exists such that the inequality

$$|u(t_i) - u(t_j)| \leqslant c_L |t_i - t_j|^{\beta_L} \tag{2}$$

holds, where t_i,t_j $(i \neq j)$ fall within $B(t_0,R)$, then the choice for β_L is legitimate, and $\alpha_L(t_0,B(t_0,R))$ is given by the supremum of these β_L. Finally, the local Hölder exponent $\alpha_L(t_0)$ is given by

$$\alpha_L(t_0) = \lim_{R\to 0} \alpha_L(t_0,B(t_0,R)) \tag{3}$$

and $\alpha_L \leqslant \alpha_p$.

This formulation permits us to contrast two functions, a cusp and a chirp, or in the terminology of Hunt and Vassilicos [28], a point vortex and a spiral vortex:

$$\text{Cusp: } u(t) = |t - t_0|^{\alpha_+}, \tag{4}$$

$$\text{Chirp: } u(t) = |t - t_0|^{\alpha_+} \sin \frac{1}{|t - t_0|^{\alpha_\omega}}. \tag{5}$$

In the former case, at $t = 0$, $\alpha_p = \alpha_L \equiv \alpha_+$. In the latter case, $\alpha_p = \alpha_+$ and $\alpha_\omega = (\alpha_p/\alpha_L) - 1$. Hence, α_p measures the density of singularities but is insensitive to their type. On the other hand, α_L provides information on the topology of these singularities.

B. Combined estimation of α_p and α_L in a function space

There are a number of techniques that exist for estimating $\alpha_p(t)$ [40] and $\alpha_L(t)$ [26,27], but if both are to be determined simultaneously, it is important that the methods used are consistent. This can be ensured using a time domain, microlocal frontier method [40,41] that resolves both $\alpha_p(t_0)$ and $\alpha_L(t_0)$ simultaneously in the time domain. A function is said to belong to $K_{t_0}^{(s,s')}$ if, for $0 < \delta < 1/4$, $c_k > 0$ and for all (t_i,t_j) where $|t_i - t_0| < \delta$ and $|t_j - t_0| < \delta$ the following is true:

$$|u(t_i) - u(t_j)| \leqslant c_K |t_i - t_j|^{s+s'} (|t_i - t_j| + |t_i - t_0|)^{-s'/2} \times (|t_i - t_j| + |t_j - t_0|)^{-s'/2}. \tag{6}$$

A comparison to (1) and (2) shows that $K_{t_0}^{(s,s')}$ includes the definitions for both α_p and α_L. Setting $s' = 0$ recovers (2), and in the limit $\delta \to 0$, one obtains $s \leqslant \alpha_L(t_0)$, meaning that the supremum is $\alpha_L(t_0)$. Because the method is concerned with first order differences (increment statistics), both s and $-s'$ have a maximum of 1. (Note that higher order singularities can be studied by replacing the left-hand side of (6) with a higher order difference [41].) Hence, it follows that $s + s' = 1$ is an upper bound on the meaningful behavior of $K^{(s,s')}$. Forming the ratio between the left-hand side of (6) as the numerator and the right-hand side as the denominator, it is clear that this ratio tends to ∞ if $s + s' > \alpha_p(t_j)$ for fixed t_j and increasing t_i. Given that $\alpha_L(t_0) \leqslant \alpha_p(t_0)$, it follows that the value for s that marks the intersection between values on the convex frontier for $K_{t_0}^{(s,s')}$ and $s + s' = 0$ gives $\alpha_p(t_0)$. Similarly, the value for s where $K_{t_0}^{(s,s')}$ intersects $s' = 0$ gives $\alpha_L(t_0)$ [40]. This is the method adopted here for the simultaneous estimation of the pointwise and local exponents [42].

III. SURROGATE-DATA ALGORITHMS

Both the IAAFT method and the multifractal method of Paluš [21] are relevant to our approach. The IAAFT algorithm for a discrete time series x_t, $t = 1,\ldots,N$, may be stated as follows [3]:

(1) Store the squared amplitudes of the discrete Fourier transform of x_t (i.e., $X_f^2 = |\sum_1^N x_t\, e^{i2\pi f(t/N)}|^2$).

(2) Perform a random shuffle of x_t to give $x_t^{(j=0)}$.

(3) Then iterate a power spectrum step and a rank-order matching step on $x_t^{(j)}$ as follows:

(a) Take the Fourier transform of $x_t^{(j)}$ and replace the squared amplitudes with X_f^2 while retaining the phases. Given the initial random sort, this means that the spectrum should be preserved but with random phases. Invert the Fourier transform with the original amplitudes restored.

(b) Replace the values in the new series $x_t^{(j)}$ by those in x_t using a rank-order matching process. This preserves the set of original values in the data set but deteriorates the quality of spectral matching, which explains why the Fourier amplitudes are replicated only approximately.

(4) Repeat until a convergence criterion is fulfilled or any changes are too small to result in any reordering of the values from the previous iteration.

In this way, the histogram of the original data is preserved precisely, and the Fourier spectrum is approximated to a given error tolerance.

The multifractal approach [21] is conceptually similar to a discrete wavelet transform (DWT) based algorithm for realizing multifractal data [43,44]. Synthetic wavelet coefficients are generated on a dyadic tree that, when inverted, yields a data set with prescribed multifractal characteristics. If the chosen distribution function $P(\eta)$ is symmetric about zero, then, given a value for the wavelet coefficient at the largest scale $w_{0,0}$, the coefficients at scale j and position k are found by

$$w_{j,k} = \eta_{j,k}\, w_{j-1,\frac{1}{2}k}. \tag{7}$$

If $P(\eta)$ is asymmetric about zero, then this is modified to

$$w_{j,k} = \epsilon_{j,k}\, \eta_{j,k}\, w_{j-1,\frac{1}{2}k}, \tag{8}$$

where, with equal probability, $\epsilon_{j,k} = \pm 1$. The Paluš [21] algorithm commences with a DWT of a data set and, with scales $j \in \{0,1\}$ left fixed, then builds surrogate values for $\eta_{j,k}$ for $j > 1$, designated here by $\tilde{\eta}_{j,k}$ according to

$$\tilde{\eta}_{j,k} = \frac{w_{j,k}}{w_{j-1,\frac{1}{2}k}} \qquad j > 1. \tag{9}$$

At each scale j, the 2^j multipliers $\tilde{\eta}_{j,k}$ are then permuted to give values $\tilde{\mu}_{j,k}$. Setting $\tilde{w}_{j,k} = w_{j,k}$ for $j \in \{0,1\}$, the new wavelet coefficients are constructed for $j > 1$ by

$$\tilde{w}_{j,k} = \tilde{\mu}_{j,k}\, \tilde{w}_{j-1,\frac{1}{2}k}. \tag{10}$$

Amplitude adjustment is then used to recover the original distribution of wavelet coefficients at each scale. That is, at a scale c, $\tilde{w}_{j=c,k}$ are replaced by the equivalent ranked values from $w_{j=c,k}$. In this way, the permutations of the DWT are randomized in a manner that preserves the original hierarchical structure. From this randomized dyadic structure of the $w_{j,k}$, the inverse DWT can be employed to generate a new data sequence. Note that the amplitude adjustment step of this algorithm is applied to $w_{j,k}$. Consequently, the histogram of the new data sequence will not exactly replicate the values for the original time series.

Figure 1 shows an example multifractional Brownian motion $m(t)$ [45,46] generated from an underlying sinusoidal function for the pointwise Hölder regularity $\alpha_p(t)$ using the algorithm of Wood and Chan [47], $\alpha_p(t) = 0.33 + 0.2\sin(6\pi t)$, for the black solid line together with two example surrogates from the IAAFT algorithm (black and displaced upwards) and using the Paluš algorithm (black and displaced downwards). The IAAFT algorithm returns the correct average roughness because of the Fourier spectrum preservation. However, it clearly fails to capture the intermittency in the test signal. The Paluš algorithm does a better job in this respect, but it is clear visually that the superior surrogates are those in gray, which are produced using the algorithm presented in this paper. In addition, while the data values in these latter surrogates occur at random locations, the Hölder properties are clearly localized: the rough and smooth patches occur at the same positions in the data and surrogates. This is a desired property of our algorithm, as shown in the example applications below.

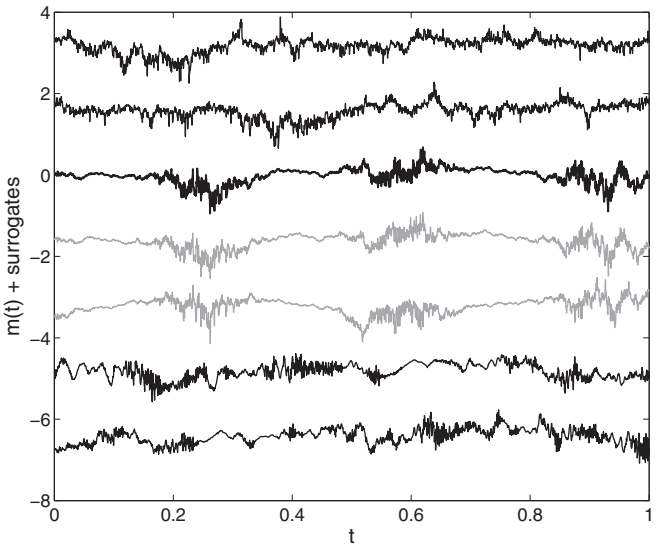

FIG. 1. A multifractional Brownian motion $m(t)$ is shown as a thick black line at the origin and is generated from $\alpha_p(t) = 0.33 + 0.2\sin(6\pi t)$. Two example IAAFT surrogates are shown in black and displaced vertically above $m(t)$, while two example Paluš surrogates are in black and displaced below $m(t)$. Surrogates generated by our algorithm are shown in gray and are immediately below $m(t)$.

IV. THE ITERATED AMPLITUDE ADJUSTED WAVELET TRANSFORM ALGORITHM

A pair of dyadic wavelet trees may be designed to form a Hilbert transform pair [48], and such wavelets have been proposed for analysis of turbulence [49] and waveform encoding [50]. More specifically, we employ the dual-tree complex DWT [51,52]. The analytic signal of some real signal $x(t)$ is given by $x_a(t) = x(t) + i\, x_H(t)$, where $x_H(t)$ is the Hilbert transform of $x(t)$, which is a convolution operator with a filter given by $h(t) = 1/(\pi\, t)$:

$$x_H(t) = \int_{-\infty}^{\infty} h(\tau)\, x(t-\tau) d\tau. \tag{11}$$

Because the Fourier transform of $h(t)$ lies completely in the imaginary plane, it follows that a Hilbert transform approach can be used to perform a complex-valued wavelet transform. Given two filters $g(t)$ and $h(t)$ and their Fourier transforms $G(\omega)$ and $H(\omega)$, it may be shown that if $G(\omega) = H(\omega)e^{-i\,\omega/2}$ for $|\omega| < \pi$, then their associated wavelets form a Hilbert pair [53], which can be achieved for orthogonal wavelets by offsetting the scaling filters by one half sample.

Hence, one approach would be to deploy two trees of linear phase filters of even length in one tree and odd in the other [51]. However, such filters lack orthogonality, and the subsampling structure is not particularly symmetric. Thus, Kingsbury [51] proposed the Q-shift dual tree where, below the coarsest scale, all filters are even length but no longer linear in phase. By designing the filters to have a delay of $\frac{1}{4}$ sample and by using the time reverse of one set of filters in the other tree, the required $\frac{1}{2}$ sample delay can be achieved. In

this paper we use symmetric, biorthogonal filters with support widths of 13 and 19 values for the first level of the algorithm and Q-shift filters with a support of 14 values for all other levels on the dual tree (case C in Kingsbury [51]). The Q-shift dual-tree approach retains properties that make undecimated transforms advantageous for use in surrogate generation, such as shift invariance [5], but at a computational cost that is merely double that for a standard discrete wavelet transform, $O(N)$, rather than $O(N \log_2 N)$, which is the case for an undecimated maximal overlap discrete wavelet transform [54].

Our algorithm may now be stated in a form that parallels the IAAFT. Given a signal of length $N = 2^J$ we proceed as follows:

(1) Perform a dual-tree complex DWT and obtain the amplitudes and phases over all J scales for the $k = 1, \ldots, 2^{j-1}$ complex-valued $w_{j,k}$ at each j.

(2) Randomly sort the original data and take the dual-tree complex DWT to produce randomized wavelet phases for each scale.

(3) Produce new $w_{j,k}$ by combining the original amplitudes with the randomized phases.

(4) Iterate the following steps until convergence is achieved:

(a) Perform the inverse wavelet transform to give a new time series, and then apply the same amplitude adjustment step used in the IAAFT algorithm.

(b) Take the dual-tree complex DWT and obtain the new phases; combine these with the original amplitudes to give the latest $w_{j,k}$.

As with the IAAFT algorithm, this method, which we term the iterated amplitude adjusted wavelet transform (IAAWT), preserves the probability distribution of original values in the data set, with the multifractal structure preserved up to a convergence criterion.

V. TESTING THE IAAWT ALGORITHM

A. Tests using signals with prescribed regularity structure

Figure 1 shows that the IAAWT algorithm replicates the multifractal characteristics of a signal when $\alpha_p(t)$ are given by a continuous and differentiable function of time. A more complex multifractal signal has its $\alpha_p(t)$ prescribed by a continuous and nondifferentiable function, such as the Weierstrass function [55], or by some hierarchical process. We consider the latter by generating a signal using the algorithm of Benzi *et al.* [43]. This provides a stringent test of the IAAWT algorithm because the similarities between the Benzi *et al.* [43] and Paluš [21] algorithms imply that the latter will yield better results than the IAAWT technique.

With reference to Eqs. (7) and (8), we use an empirical, discrete distribution function for $P(\eta)$, where $\eta \in \{0.4,0.5,0.6,0.7\}$ and the corresponding probabilities are $P(\eta) \in \{0.45,0.1,0.1,0.35\}$, and with $w_{0,0}$ selected randomly from this distribution, we then apply (8) to derive a dyadic tree of wavelet coefficients over 16 scales, which we then invert with a least asymmetric Daubechies wavelet with three vanishing moments to produce a signal

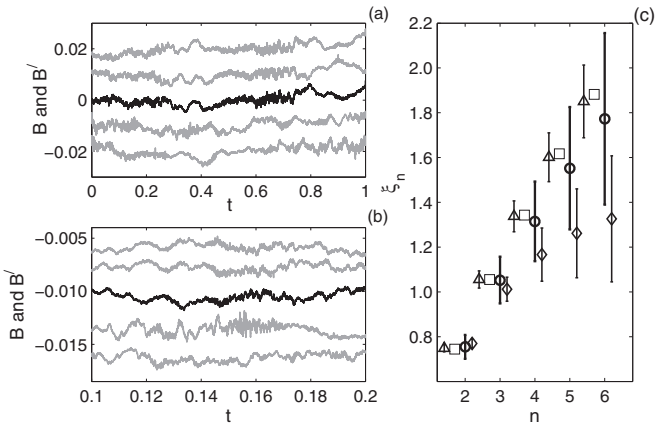

FIG. 2. (a) A multifractal signal B generated using the Benzi *et al.* [43] algorithm, shown as a black line. (b) Close-up of a portion of B. Surrogate series generated using the IAAWT algorithm are shown in gray and displaced above the black line, and those from the Paluš [21] algorithm are in gray and displaced below in (a) and (b). (c) Structure function exponents ξ_n as a function of moment number n, with error bars indicating ± 2 standard deviations. The dark lines with circles are for 20 different realizations of the process with constant $P(\eta)$. The squares give the structure function scaling for the black lines in (a) and (b), the triangles are the results for 25 IAAWT surrogates of this data series, and diamonds are the results for 25 Paluš surrogates. Values are displaced horizontally from the integer value of n for clarity.

$B(t)$ containing 2^{16} values. Nineteen further such signals were generated to characterize the inherent variability in the multifractal properties of stochastic signals with the same $P(\eta)$. The multifractal properties of these data were determined using the structure function approach [34,35], where, for a choice of moment exponent n, the structure function is $S_n(r) = \langle [B(t+r) - B(t)]^n \rangle$ and the scaling exponent $\xi(n)$ describes the power-law relation between r and $S_n(r)$ [34].

Figures 2(a) and 2(b) show an example signal B in black together with the surrogate series B' in gray. Figure 2(b) is simply a magnification of a relatively rough region of Fig. 2(a), showing that this section of the signal also contains regions of relatively high and low roughness, demonstrating the hierarchical nature of the multifractality of B. Note that both surrogate algorithms yield multifractal data, but the enhanced ability of the IAAWT algorithm to preserve the multifractal properties of the original data is quantified in Fig. 2(c), where the variability for all 20 different variants of B is shown by the circles (median value) and thick black lines (± 2 standard deviations). The structure function results for the data shown in Figs. 2(a) and 2(b) are given by squares, and structure function estimates for 25 surrogates for this signal using the IAAWT algorithm are shown by the triangles and error bars; those using the Paluš algorithm are shown by diamonds with error bars. Not only is the median for the IAAWT surrogates very close to the value for the data (shown as squares), the error in estimating ξ_n using the IAAWT is much less than that inherent in different realizations of B (circles and thick lines). The Paluš algorithm is both less accurate and less precise than our method.

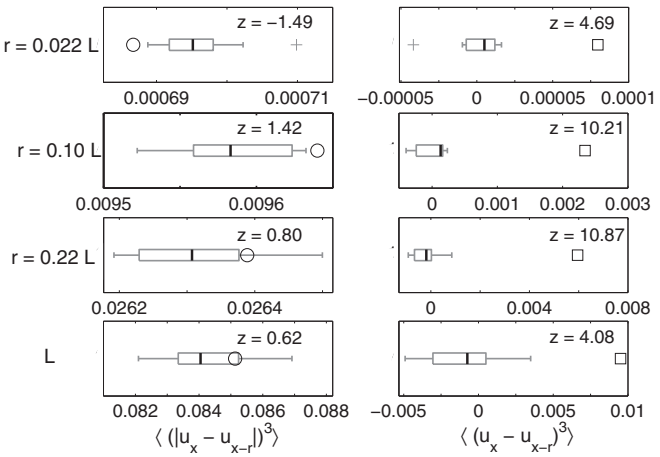

FIG. 3. An analysis of the third order moments of the velocity differences (increments) for a turbulence time series of the longitudinal velocity component u_x. The left-hand column (data as circles) shows the third moments of the absolute differences, while the right-hand column (data as squares) shows the third moments of the differences. The z value in each plot is the distance of the data from the mean of the surrogates as a function of the standard deviation of the surrogates (the z score). The values for the increment separation r range from the Taylor scale, $\sim 0.02L$, to the integral scale, L, and are given on the left. The box plots indicate the values for the surrogates: the median is shown by a thick vertical line, the edges of the box are the lower and upper quartiles, and the whiskers extend up to 1.5 times the interquartile range, with outliers shown by crosses.

B. Testing that characteristics other than multifractality are not preserved

To highlight that our algorithm is preserving only the multifractal structure, we examine the ability to preserve the skewness of the increments and skewness of the absolute part of the velocity increments for a turbulence time series in Fig. 3. The data series consists of 2^{17} samples of the longitudinal velocity component u_x measured at 0.02 m ($\sim$150 wall units above the bed of a wind tunnel in a developed boundary layer with a mean flow of 6 m s^{-1}, with more detail on the experimental design in the original paper [56]). Results are presented for choices of the separation r that vary from the Taylor scale to the integral scale L. Ten IAAWT surrogates were generated to form the box plots in Fig. 3. Because it is $\langle(|u_x - u_{x+r}|)^3\rangle$ (circles), rather than $\langle(u_x - u_{x+r})^3\rangle$ (squares), that is the relevant expression in the structure function formulation of multifractality [34], it follows that surrogates and data should lie much closer to one another in the left-hand column of Fig. 3 if our algorithm captures multifractal structure. The z scores included in each panel show this is the case: there is no significant difference between data and surrogates for any case on the left-hand side of Fig. 3 at the 5% level, while for all the cases on the right-hand side, there are clearly significant differences at this level. Where differences are seen on the left-hand side, they are at the very smallest scales, where there is inevitably some sensitivity to the random nature of the algorithm as increment statistics are taken over very small distances.

VI. EXAMPLE APPLICATIONS

There is a great range of potential applications of this algorithm as it provides a means for assessing confidence limits for any algorithm applied to multifractal data and, indeed, to techniques used to determine the multifractal properties of observed data. Here we consider two contrasting applications, both of which have potential relevance to turbulence physics.

A. Oscillating singularities in time series

Understanding the nature of any singularity or near singularity (dissipation potentially smoothing discontinuities) in turbulence is an important area connecting topology to the Navier-Stokes equations [28,57]. Related phenomena have been documented in studies of financial prices and rupture processes [58,59]. We have already defined relevant cusp and chirp signals in terms of their pointwise Hölder, $\alpha_p(t)$, and local Hölder, $\alpha_L(t)$, exponents in (5) and described the coupled estimation of these exponents. In addition, note that the set of $\alpha_p(t)$ values is often represented by its singularity spectrum $D(\alpha_p)$ [37], but various multifractal spectra can be defined: the Haussdorff multifractal spectrum [60], the Legendre transform spectrum [37], and the large deviation spectrum [61]. Prominent additional modes in the latter, resulting in a loss of convexity, indicate that α_L are an important part of the signal.

1. *Definition of artificial signals with a chirp*

Two test case signals are considered in this section: a fractional Brownian motion (fBm) with a chirp and a multifractional Brownian motion (mBm) with a chirp. They are defined for $0 \leqslant t \leqslant 1.2$ and consist of 24 000 discrete samples. The fBm has $\overline{\alpha}_p = 0.33$ (it is turbulence-like to the level of the second order structure function), while for the mBm we have $\alpha_p(t) = 0.33 + 0.15 \sin 16\pi t$. Both are discretized into 20 000 values over the support $0 \leqslant t \leqslant 1$. The chirp in both cases has a support of $-0.1 \leqslant \chi \leqslant 0.1$ discretized over 4000 values and is given by $|\chi|^{\alpha_+} \sin \frac{1}{|\chi|^{\alpha_\omega}}$, with $\alpha_\omega = 1.4$ and $\alpha_+ = 0.33$, using the notation in (5). Thus, $\alpha_p = 0.33$ and $\alpha_L \sim 0.14$. The chirp was inserted into the fBm or mBm near the center of the signal at a position that eliminated introducing any additional discontinuity (that the detection might be sensitive to).

Figure 4 shows these signals in black, focusing on the region of the chirp ($0.2 \leqslant t \leqslant 1$). IAAFT surrogates for these two signals are shown in gray in Figs. 4(b) and 4(d), and it is clear that the oscillating singularity and the intermittency have been destroyed. In contrast, Fig. 5 shows the signal from Fig. 4(c) together with three IAAWT surrogates. It is clear that the IAAWT algorithm preserves the structure of the mBm, although the detailed behavior of the oscillation is lost [with only the broad characteristics governed by $\alpha_p(t)$ preserved].

2. *Oscillating singularity detection*

In addition to the $K^{s,s'}$ method discussed in the Introduction and as noted above, the presence of oscillating singularities can be detected by a lack of convexity in the multifractal large deviation spectrum [60,61]. Figure 6 shows the singularity spectra for the fBm + chirp (top row) and the mBm + chirp (bottom row) signals. In both cases there is a lack of convexity to the multifractal spectrum, and in the latter case the secondary

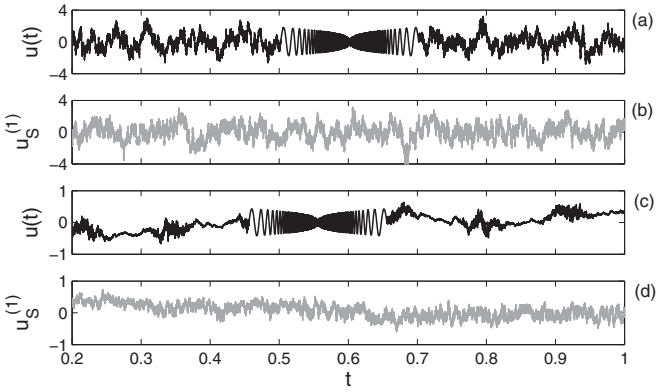

FIG. 4. (a) A fractional Brownian motion with an embedded chirp is shown in black. (b) An example IAAFT surrogate is shown in gray. (c) A multifractional Brownian motion with an embedded chirp is shown in black. (d) An IAAFT surrogate for (c) is shown in gray. The panels show a smaller segment of the full data set in each case, which has a support of $0 \leqslant t \leqslant 1.2$.

mode is as large as the primary mode centered at $\alpha_p \sim 0.33$. Figures 6(a) and 6(c) are compared to 19 IAAFT surrogates of the respective signals in gray, examples of which are given in Fig. 4. It is clear that convexity is recovered, indicating the removal of the oscillating singularity (the high values for α_p are a consequence of the edges of the chirp being smoother than $\alpha_p = 0.33$ in a pointwise sense). In contrast, the IAAWT surrogates in Figs. 6(b) and 6(d) are clearly retaining the general α_p characteristics of the oscillation, with no significant difference existing between $D(\alpha_p)$ for the data and surrogates.

Given that the general characteristics of the oscillation are retained in the IAAWT surrogates, we can use the $K^{s,s'}$ frontiers (Sec. IIB) to undertake a more detailed analysis and extract the oscillating singularity to be detected with statistical confidence. We estimated values for $\alpha_p(t)$ and $\alpha_L(t)$ for the mBm + chirp signal and for 19 IAAFT and IAAWT surrogates from the intersection of the $K^{s,s'}$ frontier (estimated with 40

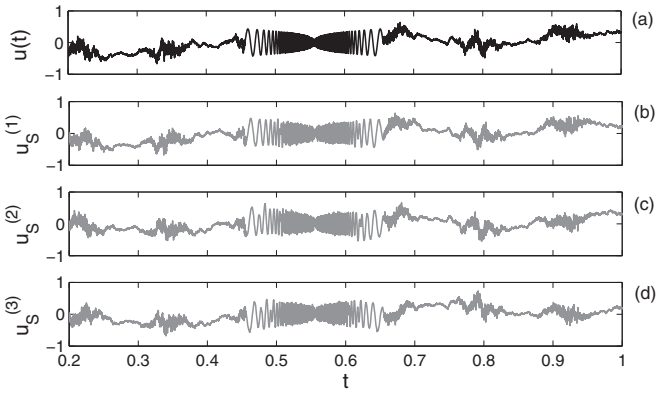

FIG. 5. (a) A multifractional Brownian motion with an embedded chirp is shown in black. (b)–(d) Three example IAAWT surrogates are shown in gray. The preservation of the pointwise Hölder exponents and the location of the oscillating components is clear. The panels show a smaller segment of the full data set in each case, which has a support of $0 \leqslant t \leqslant 1.2$.

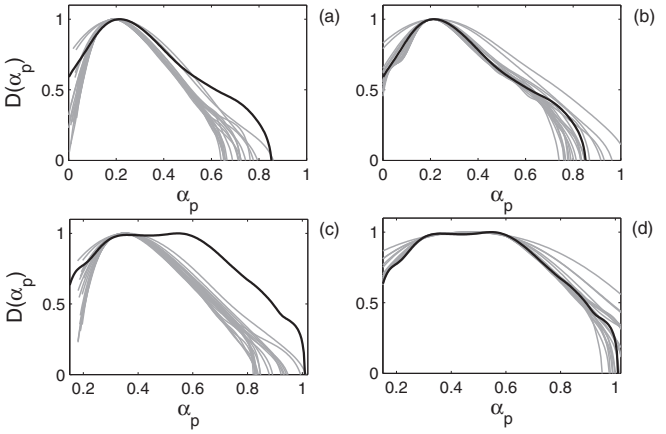

FIG. 6. Multifractal spectra $D(\alpha_p)$ estimated using the large deviation method (black) and compared to surrogate data (gray) for (a) and (b) a fractional Brownian motion with an embedded chirp and (c) and (d) a multifractional Brownian motion with an embedded chirp. (a) and (c) show IAAFT surrogate data, while IAAWT surrogates are used in (b) and (d). The multimodal structure of the spectrum indicates the presence of oscillating singularities in the original data in all cases, but this disappears in the IAAFT surrogates (although it is retained by the IAAWT surrogates).

points) with $s' = 0$ (α_L) and with $s + s' = 0$ (α_p). This was undertaken for 1001 positions, centered on the middle of the chirp and traversing double the support of the chirp (i.e., every eighth value was calculated). Given a position t_0, we may derive a z-score estimate for the pointwise and local Hölder exponents at this point. Using $\alpha_p(t_0)$ as an example,

$$z_{\alpha_p}(t_0) = \frac{\alpha_p(t_0) - [\overline{\alpha}_p(t_0)]_S}{\sigma[\alpha_p(t_0)]_S}, \tag{12}$$

where the S subscript indicates the operation is taken over the 19 values from the surrogate data, the overbar is the arithmetic mean, and σ is the standard deviation. Following Donoho and Johnstone [62], an expected upper bound on these z scores (unit standard deviation by definition) is $(2\log_e N/\sqrt{2})^{0.5}$.

The results for $z_{\alpha_p}(t)$ (black) and $z_{\alpha_L}(t)$ (gray) are shown in Fig. 7, using IAAFT [Fig. 7(a)] and IAAWT [Fig. 7(b)] surrogates. The expected upper bound on these values is shown by dashed lines in each panel. It is clear from Fig. 4(c) that the center of the chirp lies at $t \sim 0.555$. Because IAAFT surrogates preserve neither multifractal structure nor oscillating singularities, departures outside the expected upper bound arise at a variety of locations in Fig. 7(a) for both $z_{\alpha_p}(t)$ and $z_{\alpha_L}(t)$. In contrast, apart from one datum at $z_{\alpha_p}(t \sim 0.57)$, the z scores for the pointwise Hölder exponents lie within the expected bound, which is to be expected for a multifractal preserving algorithm. In contrast, z_{α_L} remains within the bound, except at the center of the chirp, where a difference is observed across several consecutive values. Hence, the IAAWT algorithm preserves $\alpha_p(t)$ characteristics and, because it also fixes in place the position of particular features, locates a chirp-like feature in the correct place. However, it does not intrinsically preserve α_L, meaning that the center of an oscillating singularity is extracted rather precisely.

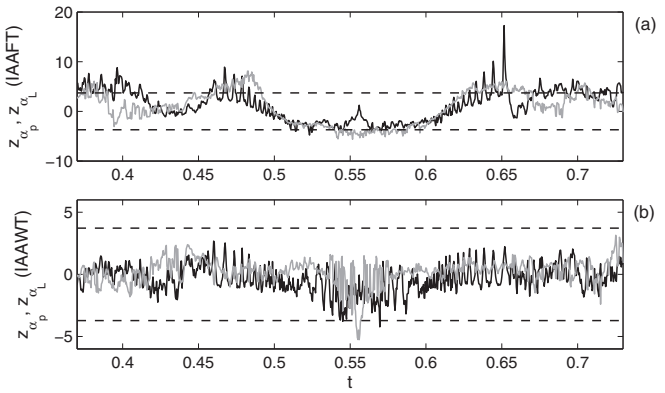

FIG. 7. Time series of z-score estimates based on the the original data and the mean and standard deviation for surrogates for α_p (black) and α_L (gray). The results are (a) for IAAFT surrogates and (b) for IAAWT surrogates. An almost sure threshold [62] is indicated by the dashed lines.

B. Velocity-intermittency analyses for turbulence

Kolmogorov's and Oboukhov's revised theories for turbulence, published in 1962 [30,63], dealt with Landau's objection to the original 1941 theory [64] through the substitution of a Gaussian distribution for the dissipation with a log-normal distribution. This mathematical adjustment had the physical consequence of resulting in intermittency in statistical models of the turbulence cascade, which were interpreted as being due to the action of vortices in the flow [65] and have been analyzed from a multifractal perspective [37,66,67]. However, another key change in the 1962 theory that has garnered less attention is the explicit recognition of potential dependencies of the coefficients in the theory on the macroscale structure of the flow [as stated for Eqs. (3) and (4) in Kolmogorov [30]]. Indeed, Frisch has argued that Kolmogorov was aware of such a potential dependence in 1941 but removed it from consideration at that time to derive a compact theory [68].

Given the potential importance of macroscale velocity to the cascade and the links between intermittency and coherent structures, potential velocity-intermittency coupling in experimentally measured flows may be analyzed using a method built around the notion of quadrants, which are typically used to describe the structure of boundary layers [69,70]. However, in the velocity-intermittency approach, the quadrants are based on the joint distribution function for the longitudinal velocity component u and its pointwise Hölder regularity $\alpha_p(u)$, rather than for u and the vertical velocity component v [33]. This method has had some success in classifying different canonical flows, demonstrating how turbulent flows forced in different ways have different characteristics [71], with results interpretable in terms of the behavior of coherent structures [72].

The velocity-intermittency quadrants are defined and analyzed as follows:

(1) For a given velocity time series $u(t)$, calculate the corresponding $\alpha_p(u)$ and scale these two terms by their respective means and standard deviations, i.e., $u^* = (u - \overline{u})/\sigma(u)$.

(2) Define threshold "hole sizes", $0 \leqslant H \leqslant H_{\max}$, where $H_{\max}$ is the largest value justifiable based on the number of values in the time series N and $H_{\max} \propto \log_e N$ and a threshold exceedance arises when $|u^*\alpha_p(u)^*| > H\sigma(u)\sigma(\alpha_p(u))$.

(3) For each H, form the probability of quadrant occupancy p_Q, which is the number of $u^* - \alpha_p(u)^*$ pairs exceeding H in a given quadrant Q divided by the number of pairs exceeding this H in all four quadrants.

(4) Approximate the relation between p_Q and H by its slope dp_Q/dH, and characterize a particular turbulent flow by its four values for dp_Q/dH.

Given a particular set of measurements for $u(t)$ as a function of time, the distribution function of $u(t)$ will have some impact on the joint distribution for u and $\alpha_p(u)$. For example, if $u(t)$ is positively skewed, then quadrants Q1 $[u' > 0, \alpha_p(u)' > 0]$ and Q4 $[u' > 0, \alpha_p(u)' < 0]$ are liable to dominate the extremal statistics. Hence, there is value in placing confidence limits on the derived values for dp_Q/dH, and IAAWT surrogates provide a means to do this as the preservation of the values for $u(t)$ and $\alpha_p(u)$ independently means one can determine if dp_Q/dH could have arisen by chance given the observed $u(t)^*$ and $\alpha_p(u)^*$. (Note that this is a test for velocity-intermittency given the velocity and intermittency. As the nature of the forcing will impact the structure for $\alpha_p(u)^*$, in particular, the observed dp_Q/dH may still be highly informative and significant physically even if their joint structure is not significant when conditioned on values for the marginals.)

Analysis of a turbulent wake

The data analyzed here were measured in a 1 m cross-section wind tunnel with a hot wire anemometer at 5 kHz in the wake of an $h = 100$ mm high fence, immersed in a neutral boundary layer with a thickness of ~200 mm. The fence consisted of a 10 mm bottom gap, followed by nine 5.3 mm horizontal elements, with a 5.3 mm space between each. Hence, the center of the fence proper is 55 mm above the wall ($y/h = 0.55$), and we recorded data for several hundred integral scales at distances downwind of the fence x in the range $1.25 \leqslant x/h \leqslant 10$. Further details on the wind tunnel, the experiments, and properties of the wake structure are given by Keylock *et al.* [56].

Results for the velocity-intermittency structure for these data at various x/h and $y/h = 0.55$ are shown in Fig. 8 as open triangles. The dominance of positive slopes for quadrant 4 $[u^* > 0, \alpha_p(u)^* < 0]$ is clear, meaning that the extremal events are of high relative velocity and intermittency. This may be contrasted with results for other flows shown in Fig. 9. For example, while the far field wake generated by flow about a cylinder (from the experiment described in Stresing and Peinke [32]) also shows a positive slope for quadrant 4 (the gray dotted and solid lines), the value for $dp_Q/dH \sim 0.03$ is much smaller. In fact, the greater positive slope for these data is in quadrant 1. In Fig. 9 it is the boundary-layer data near the wall that show the strongest positive slope in quadrant 4, with $dp_Q/dH \sim 0.07$. Given that the plane of shear for the fence wake is oriented in the longitudinal-vertical plane (i.e., similar to the Reynolds stress gradient in a boundary layer) rather than the longitudinal-transverse plane for the cylinder wake and that the Reynolds stresses in this wake

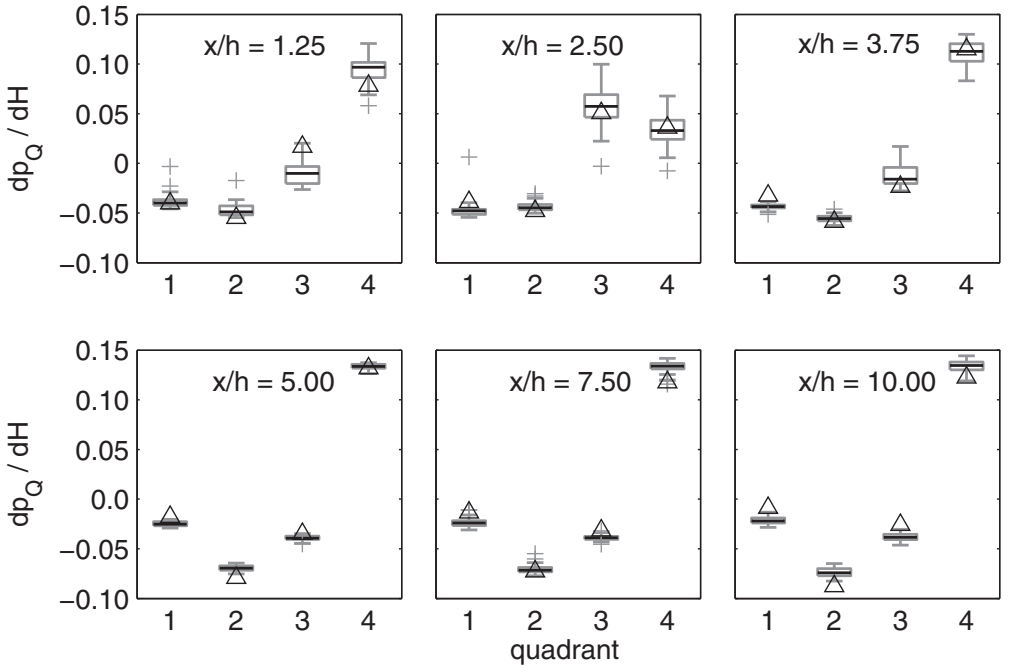

FIG. 8. Box plots of the results for the velocity-intermittency quadrant analysis at various dimensionless distances downwind of the fence. Values for the actual data are shown as open triangles. The boxes themselves are based on the values for 39 IAAWT surrogates, with outliers indicated by a cross, the box consisting of the median (center line) and first and third quartiles (bottom and top edges of the box, respectively), and the whiskers extending up to 1.5 times the interquartile deviation.

region are much greater than for the boundary-layer case, this quadrant 4 dominance for the fence wake is to be expected.

In addition to the results for the wind tunnel data in Fig. 8, we also show those obtained by generating 39 IAAWT surrogates for u at the six different positions. These results are summarized in box plot form, and it is clear that the surrogate data also exhibit the pattern of quadrant 4 dominance. Thus, the prescribed distribution function for u and the prescribed Hölder function constrain the data to a quadrant 4 dominating

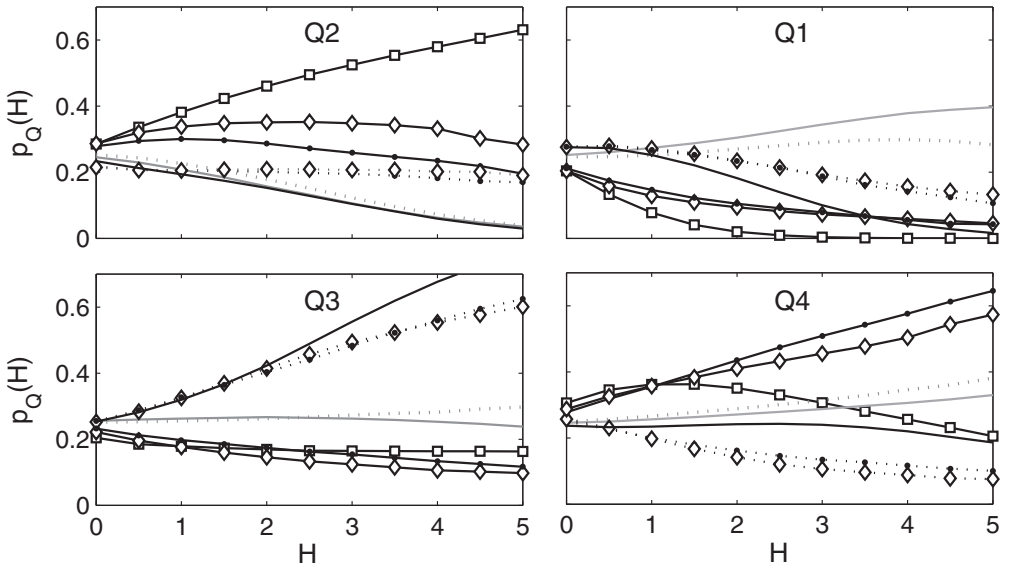

FIG. 9. An analysis of velocity intermittency from various experiments. The various lines correspond to data from a turbulent jet experiment [73] (solid black line with open squares), wake data at 8.5 m s^{-1} (gray dotted lines) and 24.3 m s^{-1} (gray solid lines) [32], and data below 150 wall units (solid black lines) and above 700 wall units (dotted black lines) at 6 m s^{-1} (open diamonds) and 8 m s^{-1} (solid circles) for a zero pressure gradient boundary layer [33]. Data for flow over bed forms [71] are shown as a solid black line. This figure is modified from C. J. Keylock *et al.* [Geophys. Res. Lett. **40**, 1351 (2013)]. Copyright the American Geophysical Union (AGU) and reproduced with the permission of AGU.

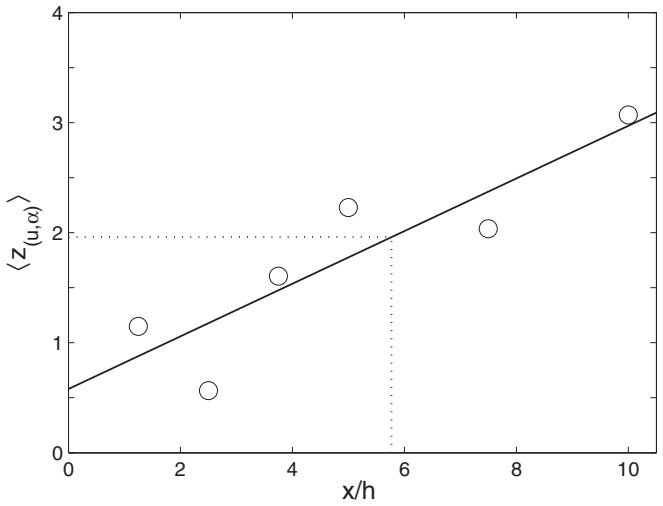

FIG. 10. A summary of the results in Fig. 8 where the distance between the data and the mean of the surrogates, normalized by the standard deviation of the surrogates (hence a z score $z_{(u,\alpha)}$), is reported, with an average taken over all four quadrants.

response. However, the extent to which the surrogates can select different values for dp_Q/dH is much more constrained for $x/h > 5$ than closer to the fence. If $z_{(u,\alpha)}(Q)$ is the z score for dp_Q/dH for a given quadrant, then $\langle z_{(u,\alpha)}\rangle$ is the average z score over all quadrants at a given x/h. This provides a summary measure of the extent to which the data and surrogates are different, and for $x/h \leqslant 10$, it is clear from Fig. 10 that this value grows downwind of the fence, with $\langle z_{(u,\alpha)}\rangle > 1.96$ at $x/h \gtrsim 6$. This is despite any major increase in $dp_{Q=4}/dH$ over this range of x/h, indicating that the velocity-intermittency structure continues to evolve relative to appropriately constrained surrogates even though the actual structure of the four values for dp_Q/dH is relatively invariant for $x/h > 5$.

VII. CONCLUSION

We have presented a surrogate-data algorithm designed to preserve not only the multifractal characteristics of a time series (the probability function for the pointwise Hölder regularity α_p) but also the observed distribution of values in the time series. Compared to an existing algorithm [21], ours replicates the distribution function for the Hölder exponents more accurately (Fig. 2) while also preserving the distribution function of the raw data values, making our algorithm more akin to the well-known IAAFT algorithm [3].

Hence, not only can this algorithm be used to generate bootstrapped data for multifractal processes, it may be used for analysis of processes where the outputs contain higher structure than a random multifractal process [61]. We have presented two such examples: the detection of oscillating singularities in a time series and the analysis of turbulence data that exhibit velocity and intermittency coupling. The latter example is particularly exciting in the context of recent extensions of fractional Brownian and multifractional Brownian motions to *self-regulating processes* where, as with our turbulence analysis, local regularity is a function of amplitude [74,75]. Our work complements these studies, which establish the mathematical basis for the formulation and estimation of such processes by providing a means to test if any observed coupling between regularity and amplitude is significant. Thus, it provides a practical criterion for assessing if the modeling of biophysical time series, planetary topography, or turbulence processes with complex forcings can be enhanced by modeling them as self-regulating processes.

ACKNOWLEDGMENTS

This work was supported by NERC Grant No. NE/F00415X/1, EPSRC Grant No. EP/K007688/1, and Royal Academy of Engineering/Leverhulme Senior Research Fellowship LTSRF1516-12-89.

[1] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. Farmer, Phys. D (Amsterdam, Neth.) **58**, 77 (1992).
[2] D. Prichard and J. Theiler, Phys. Rev. Lett. **73**, 951 (1994).
[3] T. Schreiber and A. Schmitz, Phys. Rev. Lett. **77**, 635 (1996).
[4] M. Small, D. Yu, and R. G. Harrison, Phys. Rev. Lett. **87**, 188101 (2001).
[5] C. J. Keylock, Phys. Rev. E **73**, 036707 (2006).
[6] J. Theiler and P. E. Rapp, Electroencephalogr. Clin. Neurophysiol. **98**, 213 (1996).
[7] R. B. Govindan, K. Narayanan, and M. S. Gopinathan, Chaos **8**, 495 (1998).
[8] P. C. Ivanov, L. A. N. Amaral, A. L. Goldberger, S. Havlin, M. G. Rosenblum, Z. R. Struzik, and H. E. Stanley, Nature (London) **399**, 461 (1999).
[9] S. Basu, E. Foufoula-Georgiou, and F. Porté-Agel, Phys. Rev. E **70**, 026310 (2004).
[10] D. Poggi, A. Porporato, L. Ridolfi, J. D. Albertson, and G. G. Katul, Geophys. Res. Lett. **31**, L05102 (2004).
[11] D. M. Volobuev and N. G. Makarenko, Sol. Phys. **249**, 121 (2008).
[12] C. J. Keylock, S. N. Lane, and K. S. Richards, J. Geophys. Res. **119**, 264 (2014).
[13] T. Schreiber and A. Schmitz, Phys. Rev. E **55**, 5443 (1997).
[14] C. J. Keylock, Nonlinear Processes Geophys. **17**, 615 (2010).
[15] M. Breakspear, M. Brammer, and P. Robinson, Phys. D (Amsterdam, Neth.) **182**, 1 (2003).
[16] C. J. Keylock, Phys. D (Amsterdam, Neth.) **225**, 219 (2007).
[17] P. Borgnat and P. Flandrin, J. Stat. Mech. (2009) P01001.
[18] C. J. Keylock, T. E. Tokyay, and G. Constantinescu, J. Turbul. **12**, N45 (2011).
[19] C. J. Keylock, Water Resour. Res. **48**, W12521 (2012).
[20] C. J. Keylock, A. Singh, and E. Foufoula-Georgiou, J. Geophys. Res. **119**, 682 (2014).
[21] M. Paluš, Phys. Rev. Lett. **101**, 134101 (2008).
[22] J. Gagnon, S. Lovejoy, and D. Schertzer, Europhys. Lett. **62**, 801 (2003).
[23] V. Venugopal, S. Roux, E. Foufoula-Georgiou, and A. Arnéodo, Phys. Lett. A **348**, 335 (2006).
[24] K. Matia, Y. Ashkenazy, and H. Stanley, Europhys. Lett. **61**, 422 (2003).

[25] R. Klesse and M. Metzler, Europhys. Lett. **32**, 229 (1995).
[26] S. Jaffard and Y. Meyer, Mem. Am. Math. Soc. **123**, 1 (1996).
[27] A. Arnéodo, E. Bacry, S. Jaffard, and J. F. Muzy, J. Fourier Anal. Appl. **4**, 159 (1998).
[28] J. C. R. Hunt and J. C. Vassilicos, Proc. R. Soc. London, Ser. A **435**, 505 (1991).
[29] F. Nicolleau and J. C. Vassilicos, Philos. Trans. R. Soc. London, Ser. A **357**, 2439 (1999).
[30] A. N. Kolmogorov, J. Fluid Mech. **13**, 82 (1962).
[31] A. A. Praskovsky, E. B. Gledzer, M. Y. Karyakin, and Y. Zhou, J. Fluid Mech. **248**, 493 (1993).
[32] R. Stresing and J. Peinke, New J. Phys. **12**, 103046 (2010).
[33] C. J. Keylock, K. Nishimura, and J. Peinke, J. Geophys. Res. **117**, F01037 (2012).
[34] U. Frisch and G. Parisi, in *Turbulence and Predictability in Geophysical Fluid Dynamics and Climate Dynamics*, edited by M. Ghil, R. Benzi, and G. Parisi (North-Holland, Amsterdam, 1985), pp. 84–88.
[35] R. Benzi, S. Ciliberto, R. Tripiccione, C. Baudet, F. Massaioli, and S. Succi, Phys. Rev. E **48**, R29(R) (1993).
[36] D. Schertzer and S. Lovejoy, J. Geophys. Res. **92**, 9693 (1987).
[37] J. F. Muzy, E. Bacry, and A. Arnéodo, Phys. Rev. Lett. **67**, 3515 (1991).
[38] E. Bacry, J. Muzy, and A. Arnéodo, J. Stat. Phys. **70**, 635 (1993).
[39] S. Seuret and J. Lévy Véhel, Appl. Comput. Harmonic Anal. **13**, 263 (2002).
[40] K. Kolwankar and J. Lévy Véhel, J. Fourier Anal. Appl. **8**, 319 (2002).
[41] S. Seuret and J. Lévy Véhel, J. Fourier Anal. Appl. **9**, 473 (2003).
[42] Institut National de Recherche en Informatique et en Automatique, FRACLAB toolbox, http://fraclab.saclay.inria.fr/
[43] R. Benzi, L. Biferale, A. Crisanti, G. Paladin, M. Vergassola, and A. Vulpiani, Phys. D (Amsterdam, Neth.) **65**, 352 (1993).
[44] A. Arnéodo, E. Bacry, and J. F. Muzy, J. Math. Phys. **39**, 4142 (1998).
[45] A. Benassi, S. Jaffard, and D. Roux, Rev. Mat. Iberoam. **13**, 19 (1997).
[46] R. Peltier and J. Lévy Véhel, Institut National de Recherche en Informatique et en Automatique, Research Report No. 2645, 1995 (unpublished).
[47] A. T. A. Wood and G. Chan, J. Comput. Graphical Stat. **3**, 409 (1994).
[48] I. Selesnick, IEEE Trans. Signal Process. **50**, 1144 (2002).
[49] P. Abry and P. Flandrin, in *Proceedings of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis* (IEEE, New York, 1994), pp. 225–228.
[50] E. Ozturk, O. Kucur, and G. Atkin, in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing* (IEEE, New York, 2000), p. 2641.
[51] N. Kingsbury, Appl. Comput. Harmonic Anal. **10**, 234 (2001).
[52] I. Selesnick, R. Baraniuk, and N. Kingsbury, IEEE Signal Process. Mag. **22**, 123 (2005).
[53] I. Selesnick, IEEE Signal Process. Lett. **8**, 170 (2001).
[54] J. Liang and T. W. Parks, IEEE Trans. Signal Process. **44**, 225 (1996).
[55] K. Falconer, *The Geometry of Fractal Sets* (Oxford University Press, Oxford, 1994).
[56] C. J. Keylock, K. Nishimura, M. Nemoto, and Y. Ito, Environ. Fluid Mech. **12**, 227 (2012).
[57] H. K. Moffat, Simple topological aspects of turbulent vorticity dynamics, in *Turbulence and Chaotic Phenomena in Fluids*, edited by T. Tatsumi (Elsevier, Amsterdam, 1984), p. 223.
[58] K. Ide and D. Sornette, Phys. A (Amsterdam, Neth.) **307**, 63 (2002).
[59] W. X. Zhou and D. Sornette, Phys. A (Amsterdam, Neth.) **330**, 543 (2003).
[60] S. Seuret, Math. Nachr. **279**, 1195 (2006).
[61] J. M. Aubry and S. Jaffard, Commun. Math. Phys. **227**, 483 (2002).
[62] D. L. Donoho and I. M. Johnstone, Biometrika **81**, 425 (1994).
[63] A. M. Oboukhov, J. Fluid Mech. **13**, 77 (1962).
[64] A. N. Kolmogorov, Dokl. Akad. Nauk SSSR **30**, 299 (1941).
[65] U. Frisch, P. L. Sulem, and M. Nelkin, J. Fluid Mech. **87**, 719 (1978).
[66] C. Meneveau and K. R. Sreenivasan, Phys. Rev. Lett. **59**, 1424 (1987).
[67] U. Frisch and M. Vergassola, Europhys. Lett. **14**, 439 (1991).
[68] U. Frisch, J. Bec, and E. Aurell, Phys. Fluids **17**, 081706 (2005).
[69] H. Nakagawa and I. Nezu, J. Fluid Mech. **80**, 99 (1977).
[70] D. G. Bogard and W. G. Tiederman, J. Fluid Mech. **162**, 389 (1986).
[71] C. J. Keylock, A. Singh, and E. Foufoula-Georgiou, Geophys. Res. Lett. **40**, 1351 (2013).
[72] C. J. Keylock, K. S. Chang, and G. S. Constantinescu, J. Fluid Mech. **805**, 656 (2016).
[73] C. Renner, J. Peinke, and R. Friedrich, J. Fluid Mech. **433**, 383 (2001).
[74] J. Lévy Véhel, Nonlinear Processes Geophys. **20**, 643 (2013).
[75] A. Echelard, J. Lévy Véhel, and A. Philippe, Scand. J. Stat. **42**, 485 (2015).