Bootstrapping multifractals:
Surrogate data from random cascades on wavelet dyadic trees

Milan Paluš

Institute of Computer Science, Academy of Sciences of the Czech Republic, Pod vodárenskou věží 2, 182 07 Prague 8, Czech Republic; E-mail: mp@cs.cas.cz

A method for random resampling of time series from multiscale processes is proposed. Bootstrapped series – realizations of surrogate data obtained from random cascades on wavelet dyadic trees preserve multifractal properties of input data, namely interactions among scales and nonlinear dependence structures. The proposed approach opens the possibility for rigorous Monte-Carlo testing of nonlinear dependence within, with, between or among time series from multifractal processes.

Phys. Rev. Lett. 101, 134101 (2008)

PACS numbers: 05.45.Tp, 05.45.Df, 89.75.Da, 89.75.-k

Estimation of any quantity from experimental data, with the aim to characterize an underlying process or its change, is incomplete without assessing confidence of the obtained values or significance of their difference from natural variability. With increasing performance and availability of powerful computers, Efron [1] proposed to replace (not always possible) analytical derivations based on (not always realistic) narrow assumptions by computational estimation of empirical distributions of quantities under interest using so-called Monte-Carlo randomization procedures. In statistics, the term "bootstrap" [2] is coined for random resampling of experimental data, usually with the aim to estimate confidence intervals ("error bars"). Theoretically different, but sometimes technically similar applications of the resampling approaches have been developed in the field of hypothesis testing. The latter has entered physics and nonlinear dynamics with the question of detection of chaotic dynamics in experimental data [3]. With the aim to prove that nonlinearity (and possibly, chaos) is present in analyzed data, "surrogate data" are constructed which preserve "linear properties" of the analyzed data but otherwise are realizations of a random process. The standard approach [3] uses the fast Fourier transform (FFT). Randomization of the phases of the complex Fourier coefficients and the inverse FFT provides realizations of a Gaussian process reproducing the sample spectrum and autocorrelation function of the analyzed data. Common preservation of spectra and amplitude distributions are solved by appropriate amplitude transformation and iterative procedures [3]. Breakspear et al. [4] have introduced surrogate data based on the wavelet transform [5]. The randomization is performed by one of the following three ways of manipulating the wavelet coefficients within each scale: (i) random permutation; (ii) cyclic rotation with a random offset; and (iii) block resampling, i.e., random permutation of blocks of the wavelet coefficients. Keylock [6] combines both the techniques in the sense that the wavelet coefficients within each scale undergo the iterative amplitude-adjusted FFT randomization combined with cyclic rotation in order to align extrema in coefficient values.

Generally, all these approaches reproduce the "linear properties" (the first and the second moments) of analyzed data in combinations with some constraints. Possible nonlinear dependence between a signal $s(t)$ and its history $s(t-\eta)$ is destroyed, as well as interactions among various scales in a potentially hierarchical, multiscale process. Multiscale processes that exhibit hierarchical information flow or energy transfer from large to small scales, successfully described by using the multifractal concepts (see [7] and references therein) have been observed in diverse fields from turbulence to finance [8], through cardiovascular physiology [9] or hydrology, meteorology and climatology [10]. Angelini et al. [11] express the need for resampling techniques in evaluating data from atmospheric turbulence and other hierarchical processes. They apply a sophisticated block resampling of the wavelet coefficients, however, multifractal properties of the tested data are only partially reproduced in the resampled data [11]. The "twin" surrogates [12] reproduce nonlinear dependence in trajectories, using the recurrence properties of dynamical systems evolving on or near attracting sets, however, they are not suitable for randomization of multiscale processes violating the recurrence condition.

In this letter we propose a method for random resampling of time series from multifractal processes in the sense that the resampled data replicate the multifractal properties of the original (input) data. The method reproduces the interactions among scales, so that multifractal spectra as well as nonlinear dependence structures are preserved. The proposed construction of such, let us call them *multifractal surrogate data,* is based on the idea of synthesis of multifractal signals using an orthonormal wavelet basis proposed by Arneodo et al. [7].

Let us consider a set $\{\psi_{j,k}\}$ of periodic wavelets that form an orthonormal basis of $\mathbb{L}^2([0,L])$. Thus any function $f \in \mathbb{L}^2([0,L])$ can be written as

$$f(x) = \sum_{j=0}^{+\infty} \sum_{k=0}^{2^{j-1}} c_{j,k} \psi_{j,k}(x), \tag{1}$$

where $c_{j,k} = \langle \psi_{j,k} | f \rangle = \int_L \psi_{j,k}(x) f(x) dx$, $\psi_{j,k} =$

$2^{j/2}\psi(2^{-j}x - k)$. To construct a self-similar process whose properties are defined multiplicatively from coarse to fine scales, Arneodo et al. [7] propose to define a cascade using the dyadic tree structure of coefficients of a discrete wavelet transform (DWT thereafter) in the following way

$$\begin{aligned} c_{0,0} &= 1, \\ c_{j,2k} &= W^{(l)}_{j-1,k} c_{j-1,k}, \\ c_{j,2k+1} &= W^{(r)}_{j-1,k} c_{j-1,k}, \end{aligned} \tag{2}$$

for all j and k, $j \geq 1$, $0 \leq k < 2^{j-1}$ and where $W^{(s)}_{j,k}$ ($s = l$ or $s = r$) are independent identically distributed real valued random variables. Applying an inverse DWT to the coefficients (2) we obtain a realization of a multifractal process [7].

Now, let us consider that a measurement of an experimental signal, a time series $\{s(t)\}$, $t = 1, \ldots, N$; $N = 2^n$, underwent a DWT [13] which yielded the set of the wavelet coefficients $\{c_{j,k}\}$, $0 \leq j \leq n-1$; $k = 1, 2$ for $j = 0$, and $1 \leq k \leq 2^j$ for $j \geq 1$. To obtain a randomization $\{s^{(\varrho)}(t)\}$ of the original series $\{s(t)\}$, let us compute, for each scale $j \geq 2$, the multiplicators $M_{j,k}$ as

$$\begin{aligned} M_{j,2k} &= c_{j,2k}/c_{j-1,k}, \\ M_{j,2k+1} &= c_{j,2k+1}/c_{j-1,k}. \end{aligned} \tag{3}$$

Then, keeping unchanged $c_{0,1}$, $c_{0,2}$, $c_{1,1}$ and $c_{1,2}$, the coefficients of all finer scales are recursively constructed according to the cascade

$$\begin{aligned} \tilde{c}_{j,2k} &= \mu_{j,2k}\tilde{c}_{j-1,k}, \\ \tilde{c}_{j,2k+1} &= \mu_{j,2k+1}\tilde{c}_{j-1,k}, \end{aligned} \tag{4}$$

where the multiplicators $\mu_{j,k}$ for each scale $j \geq 2$ are obtained as a random permutation of the 2^j original multiplicators $M_{j,k}$. Finally, for each scale $j \geq 2$, a new set of coefficients $\{\hat{c}_{j,k}\}$ is obtained by rearranging the sorted original $\{c_{j,k}\}$ according to the ordering of $\{\tilde{c}_{j,k}\}$. This transformation is known as the amplitude adjustment procedure in the surrogate data methodology [3] and is defined as a one-to-one correspondence between the elements of the sorted series of $\{c_{j,k}\}$ and $\{\tilde{c}_{j,k}\}$. As a result, for each scale $j \geq 2$, we obtain a random permutation of the original $\{c_{j,k}\}$, however, the permutations on different scales are not independent. The new coefficients preserve the statistical relationships among the scales due to the recurrent cascade from coarse to finer scales. Using the inverse DWT we obtain a realization $\{s^{(\varrho)}(t)\}$ of the multifractal surrogate data. Repeating the randomization procedure we can generate a number of the surrogate data realizations $\{s^{(\varrho)}(t)\}$, $\varrho = 1, \ldots$.

Using $|W^{(s)}_{j,k}|$ with a log-normal distribution and DWT with the DAUB12 basis [13] we generate a set of realizations of a multifractal process for which we can express its scaling function $\tau(q)$ as well as its multifractal spectrum $D(\alpha)$ as analytic formulae [7]. In a numerical simulation we characterize the generated time series by $\tau(q)$ estimated according to the wavelet transform modulus maxima approach [14] using the continuous wavelet transform with the complex Morlet wavelet basis [15]. In Figs. 1a,c we compare the theoretical $\tau(q)$ curve obtained from the analytic formula (solid line) with means (full circles) obtained from 1000 independent realizations of the same multifractal process. The variability of the process is illustrated by the vertical bars drawn from the 2.5th to the 97.5th percentile of the $\tau(q)$ values distribution obtained from the 1000 generated process realizations. The length of the generated time series was either 65536 (Fig. 1a) or 4096 (Fig. 1c) samples.

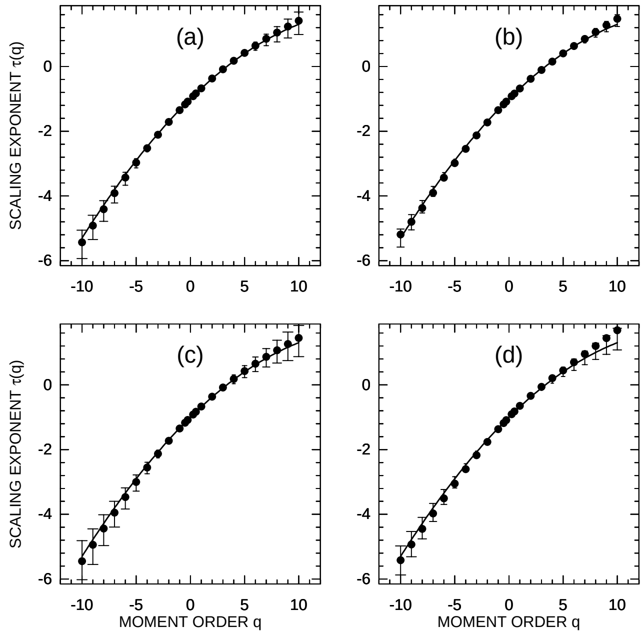

FIG. 1: The scaling function $\tau(q)$ for the multifractal process with log-normally distributed multiplicators. (a,c): Means (full circles) and 95% of distributions (vertical bars) of 1000 realizations of the process. (b,d): $\tau(q)$ values (full circles) for a particular realization of the multifractal process; and 95% of distribution (vertical bars) of 1000 realizations of the multifractal surrogate data constructed from the one process realization. The used time series length was 65536 (a,b) and 4096 (c,d) samples. The solid line shows theoretical values according to the analytic formula.

To study properties of the proposed multifractal surrogate procedure we take one, randomly chosen realization of the generated multifractal process and consider it as the input data for the randomization procedure. In Figs. 1b,d the full circles present the $\tau(q)$ values estimated from the chosen process realization, while the variability of the surrogate data is shown by the vertical bars drawn from the 2.5th to the 97.5th percentile of the $\tau(q)$ distribution obtained from 1000 realizations of the surrogate data. We can see that the surrogate procedure slightly underestimates the natural variance of the process in the case of the longer time series, however, in any

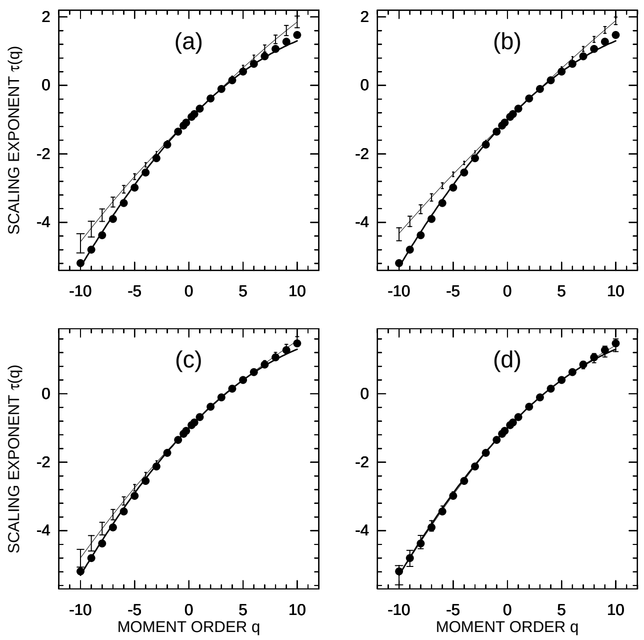

FIG. 2: The scaling function $\tau(q)$ for a particular realization of the multifractal process (full circles) and 95% of distribution (vertical bars) of 1000 realizations of (a) IAFT, (b) permutated wavelet coefficients, (c) cyclically rotated wavelet coefficients; and (d) multifractal surrogate data. The surrogate means are connected by the thin solid lines, the thick solid lines show the theoretical $\tau(q)$.

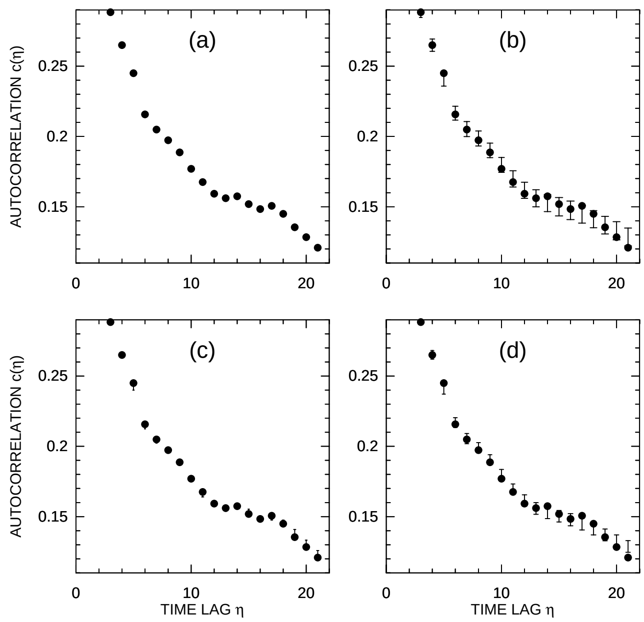

FIG. 3: Autocorrelation function of a realization of the multifractal process (full circles) and 95% of distribution (vertical bars) of 1000 realizations of (a) IAFT, (b) permutated wavelet coefficients, (c) cyclically rotated wavelet coefficients; and (d) multifractal surrogate data.

case 95% of the surrogate $\tau(q)$ distribution includes both the input data value and the theoretical $\tau(q)$ value for all q's.

The proposed multifractal surrogate data procedure is compared with the well-known surrogate procedures in Fig. 2 which is drawn in the same way as Figs. 1b,d, but, in addition, the surrogate means are connected by the thin solid lines. The iterated amplitude adjusted FFT (IAFT) surrogates reflect well the monofractal scaling, however, for $q > 6$ and $q < -3$ both the theoretical $\tau(q)$, as well as $\tau(q)$ estimated from the input data significantly differ from the IAFT surrogate data (Fig. 2a). Similar behavior can be observed using the wavelet surrogate data obtained by random permutation of the wavelet coefficients (Fig. 2b). The surrogate data obtained by the cyclic rotation of the wavelet coefficients (Fig. 2c) are closer to the original data, however, some statistically significant differences can be observed. Only the introduced multifractal surrogate data fit the scaling curve $\tau(q)$ of the input data and of the process itself (Fig. 2d, the thin line, connecting the surrogate means almost everywhere coincides with the theoretical $\tau(q)$ curve).

The IAFT surrogate data are constructed to fit the autocorrelation function (ACF) of the input data so that they replicate ACF of the input data with very small variance (in Fig. 3a the vertical bars giving the 95% of the variability of the IAFT surrogate ACF are covered by the full circles, giving the ACF of the input data). Also the other compared surrogate algorithms replicate the ACF, they just have different variability (Fig. 3). While the ACF reflects the linear dependence between a signal $s(t)$ and its future $s(t+\eta)$, for measuring possible nonlinear dependence we use the mutual information $I(s(t); s(t+\eta))$, in a short notation $I(\eta)$, given as the difference $H(s(t)) + H(s(t+\eta)) - H(s(t), s(t+\eta))$ between the marginal and joint Shannon entropies [16]. Only the proposed multifractal surrogate data are able to reproduce the nonlinear dependence between $s(t)$ and $s(t+\eta)$, as documented by $I(\eta)$ (cf. Fig. 4d with Figs. 4a–c). This property qualitatively differs the proposed multifractal surrogate data from other previously known techniques suitable for multifractal processes. Using the proposed randomization method, rigorous statistical testing and inference concerning nonlinear dependence in, with, between or among multifractal processes is possible. For instance, studying synchronization phenomena in cardio-respiratory interactions in mammals [17], the heart rhythm is characterized by inter-beat intervals (so-called RR intervals, RRI thereafter). To infer the direction of coupling, i.e., to decide whether the respiration influences the heart rhythm or vice-versa, a statistical test involving the conditional mutual information and surrogate data can be used [18]. In anaesthetized rats a simple random permutation of the RRI has been found sufficient [19]. In vigilant mammals, however, the dynamics of RRI in time is a complex, multiscale process with multifractal properties [9]. The auto-mutual information $I(\eta)$ of one variable influences the conditional mutual information used for inference of causality between two vari-

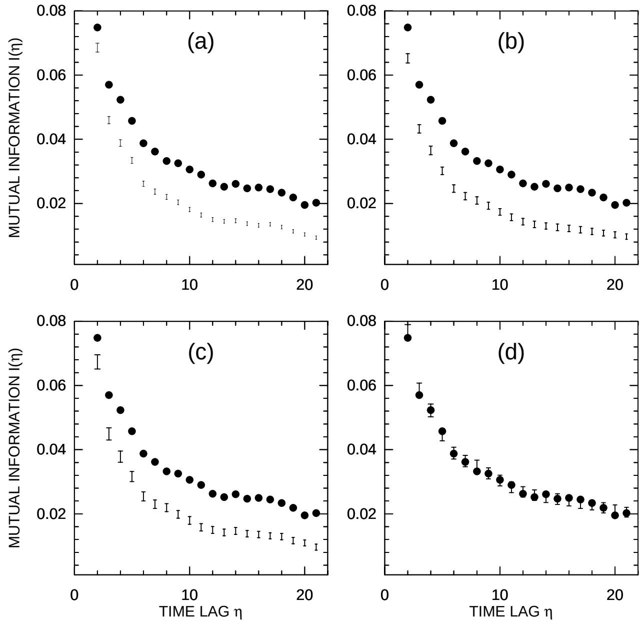

FIG. 4: (Auto)mutual information function of a realization of the multifractal process (full circles) and 95% of distribution (vertical bars) of 1000 realizations of (a) IAFT, (b) permutated wavelet coefficients, (c) cyclically rotated wavelet coefficients; and (d) multifractal surrogate data.

ables, thus none of the previously known randomization methods is appropriate for the randomization of the multifractal RRI. This problem will be discussed in detail elsewhere, here we just note the typical surrogate data application of the proposed algorithm: The RRI are randomized in order to destroy possible dependence on the respiratory rhythm, however, the multifractal properties influencing nonlinear dependence within and with RRI are preserved. In a different area, claims of multifractal properties of some meteorological and hydrological data were supported by tests using FFT-based surrogate data [10]. Nagarajan [20] recently opposed that non-Gaussian innovations were sufficient to reject the FFT-based surrogate null hypothesis in such tests, without the presence of multifractality. Such disputes could be resolved by estimation of confidence intervals of measures of multifractality using our randomization approach as a genuinely multifractal bootstrap method. Considering multifractality of financial data [8], the proposed randomization scheme can be used for Monte Carlo evaluations of financial derivatives, e.g. for option pricing [21]. The introduced method can find applications in estimation and inference using data from complex, multiscale processes, observed in various fields of physical, biological and social sciences.

This study was supported by the EC FP6 project BRACCIA (Contract No 517133 NEST), and in part by the Institutional Research Plans AV0Z10300504, MSM002160849 and ASCR project 1ET110190504.

[1] B. Efron, SIAM Rev. **21** 460 (1979).

[2] B. Efron, R. Tibshirani, *An Introduction to the Bootstrap,* Chapman & Hall, London, 1993.

[3] J. Theiler, S. Eubank, A. Longtin et al., Physica **D 58** 77 (1992); D. Prichard, J. Theiler, Phys. Rev. Lett. **73** 951 (1994); M. Paluš, Physica **D 80** 186 (1995); J. Theiler, D. Prichard, Physica **D 94** 221 (1996); T. Schreiber, A. Schmitz, Physica **D 142**(3-4) 346 (2000).

[4] M. Breakspear, M. Brammer, P.A. Robinson, Physica **D 182** 1 (2003).

[5] I. Daubechies, Ten Lectures on Wavelets, SIAM, Philadelphia, 1992.

[6] C.J. Keylock, Phys. Rev. E **73** 036707 (2006); Physica **D 225** 219 (2007).

[7] A. Arneodo, E. Bacry, J.F. Muzy, J. Math. Phys. **39**(8) 4142 (1998).

[8] R.M. Mantegna, H.E. Stanley, Nature **383** 587 (1996); K. Matia, Y. Ashkenazy, H.E. Stanley, Europhys. Lett. **61**(3) 422 (2003).

[9] P.C. Ivanov, L.A.N. Amaral, A.L. Goldberger et al., Nature **399** 461 (1999); P.C. Ivanov, L.A.N. Amaral, A.L. Goldberger et al., Chaos **11**(3) 641 (2001).

[10] V.N. Livina, Y. Ashkenazy, P. Braun et al., Phys. Rev. E **67** 042101 (2003); Y. Ashkenazy, D.R. Baker, H. Gildor et al., Geophys. Res. Lett. **30**(22) 2146 (2003).

[11] C. Angelini, D. Cava, G. Katul et al., Physica **D 207** 24 (2005).

[12] M. Thiel, M.C. Romano, J. Kurths et al., Europhys. Lett. **75**(4) 535 (2006).

[13] W.H. Press, B.P. Flannery, S.A. Teukolsky and W.T. Vetterling, *Numerical Recipes in FORTRAN 77: The Art of Scientific Computing* (Cambridge Univ. Press, Cambridge 1992), Chapter 13.10.

[14] A. Arneodo, E. Bacry, J. Muzy, Physica **A 213** 232 (1995); E. Bacry, J. Muzy, A. Arneodo, J. Stat. Phys. **70** 635 (1993); S. Mallat, *A Wavelet Tour of Signal Processing,* Academic Press, San Diego, CA (1998); B. Audit, E. Bacry, J.F. Muzy, A. Arneodo, IEEE Trans. Inf. Theor. **48**(11) 2938 (2002).

[15] C. Torrence, G.P. Compo, Bull. Amer. Meteorological Society **79**(1) 61 (1998).

[16] T.M. Cover, J.A. Thomas, *Elements of Information Theory* (J. Wiley & Sons, New York, 1991); K. Hlaváčková-Schindler et al., Phys. Rep. **441** 1 (2007).

[17] C. Schäfer, M.G. Rosenblum, J. Kurths, H.-H. Abel, Nature **392** 239 (1998); Phys. Rev. E **60** 857 (1999); A. Stefanovska, H. Haken, P. V. E. McClintock et al., Phys. Rev. Lett. **85(22)** 4831 (2000).

[18] M. Paluš, M. Vejmelka, Phys. Rev. E **75** 056211 (2007).

[19] B. Musizza et al., J. Physiol. (London) **580** 315 (2007).

[20] R. Nagarajan, Physica **A 370** 355 (2006); Physica **A 366** 530 (2006); Physica **A 374** 281 (2007).

[21] F. Black, M. Scholes, J. Polit. Econom. **81** 637 (1973); J. Voit, Physica **A 321** 286 (2003).