Signal Processing 105 (2014) 243–257

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Scaling range automated selection for wavelet leader multifractal analysis

Roberto F. Leonarduzzi [a,b,*,1], María E. Torres [a,b], Patrice Abry [c]

[a] Laboratorio de Señales y Dinámicas No Lineales, Universidad Nacional de Entre Ríos, Entre Ríos, Argentina
[b] CONICET, Argentina
[c] Physics Dept. (CNRS UMR 5672), at École Normale Supérieure de Lyon, 69364 Lyon Cedex 07, France

ARTICLE INFO

Article history:
Received 22 December 2013
Received in revised form
26 April 2014
Accepted 4 June 2014
Available online 12 June 2014

Keywords:
Automated scaling range selection
Bootstrap
Wavelet leaders
Multifractal analysis

ABSTRACT

Scale invariance and multifractal analysis constitute paradigms nowadays widely used for real-world data characterization. In essence, they amount to assuming power law behaviors of well-chosen multiresolution quantities as functions of the analysis scale. The exponents of these power laws, the scaling exponents, are then measured and involved in classical signal processing tasks. Yet, the practical estimation of such exponents implies the selection of a range of scales where the power law behaviors hold, a difficult task with yet crucial impact on performance. In the present contribution, a nonparametric bootstrap based procedure is devised to achieve scaling range automated selection. It is shown to be effective and relevant in practice. Its performance, benefits and computational costs are assessed by means of Monte Carlo simulations. It is applied to synthetic multifractal processes and shown to yield robust and accurate estimation of multifractal parameters, despite various difficulties such as noise corruption or inter-subject variability. Finally, its potential is illustrated at work for the analysis of adult heart rate variability on a large database.

1. Motivation, context and contribution

1.1. Scale invariance and multifractal analysis

After Mandelbrot's seminal intuitions and contributions [1,2], the paradigm of *scale invariance*, also referred to as *scaling*, or sometimes *fractal*, has been used to model and/or analyze the temporal dynamics of many different real-world data sets produced by applications of very different natures, including biomedical [3], internet [4], physics [5], geophysics [6], finance [7], etc.

Irrespective of the details of the considered applications, the scale invariance concept amounts to assuming that no specific scale plays a dominant role in the temporal dynamics of the data, and that instead time scales, spread within a large range, are all equally contributing to data temporal dynamics. For such situations, data analysis should no longer consist in identifying preferred scales, but instead in essentially quantifying mechanisms that relate scales ones to the others. Assessing scale invariance requires the use of multiresolution quantities, $T_X(a,t)$, i.e., quantities that depend jointly on time (or space) and scale.

* Corresponding author. Postal address: Facultad de Ingeniería – UNER, C.C. 47 – Suc. 3 – 3100 Paraná, Entre Ríos, Argentina.
Tel.: +54 343 4975100x122.
E-mail addresses: rleonarduzzi@bioingenieria.edu.ar (R.F. Leonarduzzi), metorres@santafe-conicet.gov.ar (M.E. Torres), patrice.abry@ens-lyon.fr (P. Abry).

[1] Work supported by the French-Argentine CNRS-CONICET Grant # 112432 2011-2012, the ANPCyT-UNER PICT Grant # 433 2008, ANPCyT-CONICET PICT Grant # 2954 2012, and the ANR AMATIS Grant # 112432, 2010-2014.

http://dx.doi.org/10.1016/j.sigpro.2014.06.002
0165-1684/

Classical choices are increments, oscillations, or the nowadays commonly used wavelet coefficients. In practice, scale invariance can be evidenced and measured via a power-law (or algebraic) behavior of the time averages of $T_X(a,t)$, with respect to analysis scales a, over a large range of scales,

$$\frac{1}{n_a}\sum_k |T_X(a,k)|^q \simeq C_q a^{\zeta(q)}, \quad a_m \le a \le a_M, \ \frac{a_M}{a_m} \gg 1, \tag{1}$$

where n_a denote the number of such $T_X(a,k)$ available at scale a and the $\zeta(q)$ are usually termed the scaling exponents. These scaling exponents are classically used to analyze, characterize and classify signals or images (cf. e.g., [8,9]). More recently, multifractal analysis (cf. e.g., [10,11]), that aims at quantifying the fluctuations of regularity along time (or space) via the so-called multifractal spectrum, has received considerable interest in signal and image processing and is now considered a standard analysis tool. A recent and powerful formulation of multifractal analysis relies on a choice of specific multiresolution quantities, referred to as wavelet leaders, and is largely used in the sequel.

1.2. Scaling range selection

The practical use of the concept of scale invariance essentially amounts to estimating the scaling exponents $\zeta(q)$. Whatever the estimation procedure, Eq. (1) above clearly shows that practical estimation strongly relies on the choice of the range of scales, from now on referred to as the *scaling range*, where the power-law behavior holds. While the theoretical assumption that data X are exactly self-similar would imply an infinite scaling range (i.e., $a_m \to 0$ and $a_M \to +\infty$), in practice, the scaling range must often be considered limited, which may stem from many different causes. At the theoretical level, models used to describe data often yield asymptotic only power law behaviors, as in Eq. (1). For instance, multiplicative construction (underlying multifractal models) implies $a_m \to 0$ [10], while Long Memory models implies $a_M \to +\infty$ [12]. At the practical level, finite scaling range may result from physical (or physiological) mechanisms whose dynamics involve a large yet bounded range of scales, while other competing mechanisms may become dominant at finer or coarser scales (e.g., dissipation in turbulence [5], beat-to-beat nature in heart rate variability [13]). Also, data are digitalized at a given rate, with sampling devices necessarily destroying the power law behaviors at fine scales. Along the same line, the necessarily finite duration of recordings is likely to introduce power-law cut-off at coarse scales. Additionally, noise can be superimposed to truly scaling data, often leading to a substantial narrowing of the range of scales where scale invariance can actually be observed. These different mechanisms thus imply that scale invariance in practice necessarily holds only within a potentially large but finite range of scales, bounded below and above by lower and upper cutoffs. Further, purely from a performance perspective, estimation of the scaling exponents $\zeta(q)$ requires a careful selection of the range of scales, where estimation should be performed. Essentially, that selection is driven by a classical bias-variance trade-off: A large scaling range yields a low variance at the risk of bias, due to the often asymptotic nature of scale invariance; a narrow range, centered over scales where power law holds, reduces bias but at the price of an increased variance.

1.3. Related works

Though most practitioners are perfectly aware both of the crucial impact of scaling range selection on estimation performance and of the difficulties an objective and automated selection raise, this issue remains barely addressed. Essentially, scaling range can be selected either from *fundamental* arguments related to the physics (the physiology, etc.) underlying the data at hand (e.g., Kolmogorov dissipation scale in hydrodynamic turbulence [5], sympathetic–parasympathetic frequency band split in heart rate variability [14]), or from empirical data analysis. In this latter perspective, visual inspection and empirical experience remain the dominant practice amongst practitioners. This is however obviously a tedious and error prone procedure, notably for large databases, where each signal needs to be inspected and where noise level corruption is likely to vary individually from one signal to another, a very common observation in biomedical data notably. Amongst the rare attempts to address the issue in an automated way, χ^2-statistics and *F*-statistics based procedures were devised and studied in [15] and [16] respectively, to obtain the lower cutoff scale in the analysis of Gaussian long memory processes. In [17], attempts were made to relax the Gaussian assumption, relying on the use of Theil's inequality coefficient. Yet, those approaches do not straightforwardly nor relevantly extend to multifractal analysis, toward which we concentrate in the present contribution, where Gaussian assumptions do not a priori constitute valid approximations.

1.4. Goals, contributions and outline

In this context, the present contribution aims at proposing and assessing a practically effective procedure for the automated selection of the scaling range in the general wavelet leader framework for empirical multifractal analysis, thus not assuming a priori Gaussianity of the multiresolution quantities actually used [8,11]. To that end, a short introduction for empirical and practical multifractal analysis is recalled in Section 2. Inspired from the use of the χ^2-statistics in [15] and of the bootstrap framework developed in [8], a bootstrap-based procedure is motivated and constructed in Section 3 that permits an automated scaling range selection. Its performance is assessed by means of Monte Carlo simulations based on synthetic multifractal processes according to the protocol detailed in Section 4.1. Performance is reported and discussed with respect to optimal mean square error, data length and trade-off between estimation and analysis, for perfect multifractal processes in Section 4.2. Performance and robustness are further evaluated against noise corrupted multifractal processes (cf. Section 4.3), or multifractal processes with upper cutoff (cf. Section 4.4), or multifractal processes suffering from both corruptions at fine and