Physica D 374–375 (2018) 45–57

Contents lists available at ScienceDirect

Physica D

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

Simultaneous estimation of deterministic and fractal stochastic components in non-stationary time series

Constantino A. García [a,*], Abraham Otero [b], Paulo Félix [a], Jesús Presedo [a], David G. Márquez [a]

[a] Centro Singular de Investigación en Tecnoloxías da Información (CiTIUS), Universidade de Santiago de Compostela, 15782, Santiago de Compostela, Spain
[b] Department of Information and Communications Systems Engineering, Universidad San Pablo CEU, 28668, Madrid, Spain

HIGHLIGHTS

- The superimposition of deterministic and fractal stochastic components is studied.
- Deviations from the expected energy by scale describe the deterministic features.
- A Bayesian technique able to characterize both components is proposed.
- The method is applied to economic and physiological data.

ARTICLE INFO

Article history:
Received 4 August 2016
Received in revised form 14 March 2018
Accepted 7 April 2018
Available online 24 April 2018
Communicated by T. Sauer

Keywords:
$1/f$ noise
Fractional Brownian motion
Bayesian estimation
Wavelet transform

ABSTRACT

In the past few decades, it has been recognized that $1/f$ fluctuations are ubiquitous in nature. The most widely used mathematical models to capture the long-term memory properties of $1/f$ fluctuations have been stochastic fractal models. However, physical systems do not usually consist of just stochastic fractal dynamics, but they often also show some degree of deterministic behavior. The present paper proposes a model based on fractal stochastic and deterministic components that can provide a valuable basis for the study of complex systems with long-term correlations. The fractal stochastic component is assumed to be a fractional Brownian motion process and the deterministic component is assumed to be a band-limited signal. We also provide a method that, under the assumptions of this model, is able to characterize the fractal stochastic component and to provide an estimate of the deterministic components present in a given time series. The method is based on a Bayesian wavelet shrinkage procedure that exploits the self-similar properties of the fractal processes in the wavelet domain. This method has been validated over simulated signals and over real signals with economical and biological origin. Real examples illustrate how our model may be useful for exploring the deterministic–stochastic duality of complex systems, and uncovering interesting patterns present in time series.

1. Introduction

In many physical systems when measuring a physical quantity along time, we often obtain a time series which seems to fluctuate in a non-periodic, apparently random manner, with complex correlations that extend over all measured time scales. These long-term dependencies are specially evident in the frequency domain and they regularly appear over wide ranges of frequencies as a $1/f$ process, that is, a random process with a power spectral density $S(f) = \text{constant}/|f|^{\gamma}$, where γ receives the name of spectral index [1]. A large number of physical and informational systems from biology, chemistry, geology, meteorology, economics or engineering, among others, exhibit such behavior [2]. A $1/f$ process has a long and dynamic memory; the effects of an event persists in time, and the influence of recent events is added to and gradually supersedes the influence of distant events [3]. This long-memory property accounts for the behavior of informational systems, which exhibit an evolving increase in structure and complexity by accumulating information, combining the strong influence of past events with the influence of current events. Another important feature of $1/f$ processes is the lack of a characteristic time scale, exhibiting statistical self-similar properties. This permits the explanation of the thermodynamic behavior of a large number of physical systems, based on the concurrence of multiple processes with relaxation times comparable to all time scales of interest. Moreover, this feature is particularly expected in biological systems, since the

* Corresponding author.
E-mail address: constantinoantonio.garcia@usc.es (C.A. García).

https://doi.org/10.1016/j.physd.2018.04.002
0167-2789/© 2018 Elsevier B.V. All rights reserved.

existence of a preferred frequency of operation would seriously limit the reaction capability of a living system [4,5].

Despite the existence of effective mathematical tools for studying such $1/f$ processes, modeling systems with long-term dependencies remains a challenge. This is mainly because real systems do not usually consist of merely stochastic or deterministic components, but they often manifest both random and predictable features. The need to integrate stochastic and deterministic approaches has already been considered by means of a wide variety of techniques. The Langevin formalism [6–8] (equivalent to the Fokker–Planck formulation) describes the time evolution of a system in terms of a deterministic and a stochastic driving force. In most practical applications, the free coefficients of the dynamical equations are directly estimated from the time series. Langevin equations provide a powerful method for characterizing systems in which noise is an integral part of the dynamics, but they usually require two strong properties: stationarity and Markovianity. Recent advances have enabled the application of Langevin equations to time series that do not meet some of these requirements. For example, Langevin-like modeling has been successfully applied to non-stationary conditions by means of additional measurable properties [9,10] or by assuming a parametric model in which the parameters themselves evolve according to stationary stochastic processes [11]. On the other hand and concerning Markovianity, consider for example [12], in which Langevin-like equations have been properly estimated in the presence of observational noise spoiling the Markov properties of the experimental time series.

The Coarse Graining Spectral Analysis technique (CGSA) [13] assumes that the physical quantity under study is a sum of a fractal stochastic model (a fractional Brownian motion) and a sum of stationary sinusoids, and aims to isolate the fractal stochastic spectrum in order to calculate more precisely the parameters that characterize it. Thus, it does not provide any temporal estimate of the harmonic or of the fractal stochastic signals. Furthermore, real data often behave as non-stationary. As it will become apparent below, our method may be seen as a generalization of the CGSA that provides temporal estimates and addresses the issue of non-stationarity.

Explicit modeling and non-stationary approaches like [14–16] are only possible when the mechanisms that underlie the generation of the signal are, at least to some extent, understood. This makes possible building precise mathematical models that account for the behavior of the observed time series. In our opinion, it is through explicit model building that we gain a truly deeper understanding about how a system behaves. However, model building is a complicated task which usually requires many iterations. An accurate characterization of the main stochastic and deterministic features can help in taking modeling decisions.

The main purpose of this paper is to expedite model building by providing reliable simultaneous estimates of the deterministic and stochastic components present in a system where the deterministic–stochastic interactions may be ignored. We expect these estimates to be useful for exploring the deterministic–stochastic duality in complex systems with long-term correlations, providing a valuable starting point towards the development of interpretable models. To this end, we first propose a realistic, yet simple model, that is able to capture key features of these sort of complex systems. From this model we have developed a method that enables us to characterize the fractal stochastic component and to provide an estimate of the deterministic components of the system.

The remainder of the paper is organized as follows. In Section 2, we describe the specific model in which we are going to focus and its statistical properties. We also propose a method that exploits these statistical properties to estimate the deterministic and fractal stochastic parts of a signal. In Sections 3–5, we apply our method on simulated data and real data with economical and biological origin, respectively. Finally, the results of this paper are discussed and some conclusions are given in Section 6.

2. Time series with stochastic and deterministic contributions

2.1. The stochastic contribution

Let us consider a complex physical system characterized by the temporal series Y. We assume that Y is the result of the superposition of a stochastic series B and some band-limited deterministic signal x:

$$Y[n] = x[n] + B[n] \quad n = 1, \ldots, N. \tag{1}$$

For the moment, we will focus on the stochastic part and we will assume that our signal $Y[n]$ can be approximated by $Y[n] = B[n]$, where $B[n]$ is an evolutionary random process with long-term dependencies. Therefore, its present behavior is strongly influenced by its entire history. According to a convenient modeling approach [1], long-memory signals are treated as realizations of one of two processes: either a fractional Gaussian noise (fGn), or a fractional Brownian motion (fBm). Since the increments of a non-stationary fBm signal yield a stationary fGn signal, we will focus on the fBm model without any loss of generality.

The fBm is a stochastic non-stationary process with zero mean that is fully characterized by its variance σ^2 and the so-called Hurst exponent $0 < H < 1$. Due to the relationship between fGn and fBm, fGn models are also fully specified by these two parameters. We denote a fBm model using $B(t)$ and a discrete fBm (dfBm) by $B[k] = B(k \cdot T_s)$, being T_s the sampling period of the signal [17]. The non-stationary behavior of the dfBm signals is apparent when considering its covariance [18]:

$$\mathrm{Cov}(B[k], B[l]) = \frac{\sigma^2 T_s^{2H}}{2}(|k|^{2H} + |l|^{2H} - |k-l|^{2H}). \tag{2}$$

Although non-stationary, dfBm does have stationary increments $B[k+d]-B[k]$, and its probability properties only depend on the lag d, the exponent H and σ^2. This increment process follows a discrete fGn (dfGn) distribution [17], and it is self-similar in the sense that, for any $a > 0$, B fulfills

$$(B[k + a \cdot d] - B[k]) \sim a^H (B[k+d] - B[k]), \tag{3}$$

where $\sim$ means "distributed as".

In summary, two important features appear as relevant when analyzing dfBm signals: non-stationarity, which demands for a time-dependent analysis; and self-similarity, which demands for a scale-dependent analysis. Wavelet analysis, which provides a time and scale dependent method, seems to be a proper framework for studying this kind of time series. The next section summarizes the most important statistical properties of the dfBm wavelet coefficients.

2.2. Exploring self-similarity with wavelets

Let us assume that B is an infinite signal, so that we can ignore border effects on the wavelet transform. The wavelet transform of B decomposes the signal in terms of a set of oscillating functions $\{\psi_{j,l}\}$, each of which is well localized in time around l and it is related with a resolution level j. In the notation of this paper, the jth resolution level is associated with the scale 2^{-j} and thus, larger j correspond to finer scales. Since the set of functions $\{\psi_{j,l}\}$ are generated by parameterizing the function ψ, the latter is called mother wavelet. The wavelet decomposition of B is

$$B[n] = \sum_{j=J}^{\infty} \sum_{l \in \mathbb{Z}} wB_{j,l}[n]\psi_{j,l}[n] + \sum_{l \in \mathbb{Z}} aB_J[n]\phi_{J,l}[n],$$

where $\phi_{J,l}[n]$ is the so-called scaling function (closely related to the mother wavelet), wB_j are the wavelet coefficients of the jth resolution level, and aB_J are the approximation coefficients associated with some Jth resolution level taken as reference level.