Signal Processing 143 (2018) 241–252

Contents lists available at ScienceDirect

Signal Processing

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

A nonparametric test for slowly-varying nonstationarities

Douglas Baptista de Souza[a,*], Jocelyn Chanussot[b], Anne-Catherine Favre[c], Pierre Borgnat[d]

[a] GE Global Research Center, Rua Trinta e Seis, Ilha do Bom Jesus, 21941-593, Rio de Janeiro, RJ, Brazil
[b] GIPSA-Lab, 11 rue des Mathématiques, BP46 - 38402, Saint-Martin d'Hères, France
[c] LTHE, 70 rue de la physique, BP53 - 38400, Saint-Martin d'Hères, France
[d] Lab PHYS, École normale supérieure de Lyon, CNRS, 46, allée d'Italie 69364 Lyon cedex 07 France

ARTICLE INFO

Article history:
Received 16 March 2017
Revised 19 July 2017
Accepted 31 August 2017
Available online 1 September 2017

Keywords:
Stationarity test
Slowly-varying nonstationarities
Time marginals
Trend detection
Empirical mode decomposition
Real-world signals
Block bootstrapping

ABSTRACT

This paper develops a new nonparametric method that is suitable for detecting slowly-varying nonstationarities that can be seen as trends in the time marginal of the time-varying spectrum of the signal. The rationale behind the proposed method is to measure the importance of the trend in the time marginal by using a proper test statistic, and further compare this measurement with the ones that are likely to be found in stationary references. It is shown that the distribution of the test statistic under stationarity can be modeled fairly well by a Generalized Extreme Value (GEV) pdf, from which a threshold can be derived for testing stationarity by means of a hypothesis test. Finally, the new method is compared with other ones found in the literature.

1. Introduction

Stationarity is a crucial assumption for many statistical models, but many real-world signals turn out to be nonstationary. For instance, rainfall [1] and sunspot [2] signals are real-world processes that commonly exhibit nonstationary behaviors. Assessment of stationarity is thus an important task in signal processing and time series analysis. The goal of this paper is to propose a method suitable for testing slow nonstationary evolutions. As a matter of fact, detecting such nonstationarities is especially challenging and most of traditional tests fail.

Several stationarity tests have been proposed in the last decades, some being rooted in time series modeling [3–5], spectral analysis [6], or detection of abrupt changes [7]. The emerging alternatives in the literature can be categorized into parametric and nonparametric approaches. The definition of nonparametric technique adopted in this work is that of a method which does assume any *a priori* functional form or parametric model for the input signal [8]. Therefore, even if the technique makes use of a particular window function to analyze the signal (which could be considered *a priori* a kind of model), if the methods requires no assumption regarding the distribution (Gaussian, gamma, etc.) or process model (TVAR, ARMA, etc.) of the input signal, then we consider this method as nonparametric. Nonparametric methods may be preferable in real-world applications, as the performance of a parametric method depends on the accuracy of the chosen model, which is hardly assessed for real-world data.

Methods for detecting slow nonstationary evolutions, however, are more common in the class of parametric techniques. Some parametric methods that have been proposed in the past years assume the underlying signal can be modeled by a time-varying autoregressive process (TVAR) [3,4,9,10]. Among the nonparametric methods, the work in [11] has presented a technique for detecting changes in high-order statistics, whereas [12] has proposed a test for second-order stationarity in the TF domain. A common drawback of TF-based methods, however, is the computational load required to estimate full TF representations [13]. Moreover, traditional TF representations often estimate poorly the spectral content at very low frequencies [14]. The latter is an important issue if the goal is to detect slowly-varying nonstationarities. In this regard, the method of [12] has been modified in [15] so as to improve the detection of nonstationary signals with spectral content more concentrated at low frequencies. Some other methods have also been proposed in the literature to improve the resolution of TF transforms at low frequency bands [16,17]. Unfortunately, these modifications end up increasing even more the computational complexity of the TF-based approaches.

* Corresponding author.
E-mail addresses: dougdabaso@gmail.com (D. Baptista de Souza), jocelyn.chanussot@gipsa-lab.grenoble-inp.fr (J. Chanussot), Anne-Catherine.Favre-Pugin@ense3.grenoble-inp.fr (A.-C. Favre), Pierre.Borgnat@ens-lyon.fr (P. Borgnat).

http://dx.doi.org/10.1016/j.sigpro.2017.08.026
0165-1684/

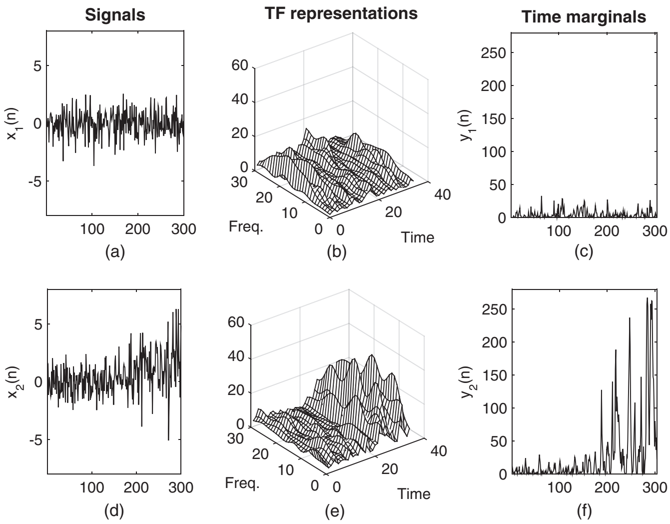

Fig. 1. (a) A stationary Gaussian signal with $\mu = 0$, $\sigma^2 = 1$ and length $N = 300$. (d) A Gaussian signal with time-varying mean and variance of the same length. The TF representations of $x_1(n)$ and $x_2(n)$ are shown in (b) and (e). Their time marginal series $y_1(n)$ and $y_2(n)$ are shown in (c) and (f).

In this paper, we show how a nonparametric test for slow nonstationary evolutions can be built in time domain, by developing a hypothesis test for the presence of trends in the marginal of the time-varying spectrum. A crucial point of the proposed technique is that we compute the time marginal directly in time domain, therefore avoiding the problems mentioned above involving TF representations. We remark that this paper is a modified and extended version of the work presented in [18]. Different from the present paper, the work in [18] has been built in TF domain and has not been developed as a proper hypothesis test. To build the new stationarity test, concepts like bootstrapping and GEV modeling are introduced. Furthermore, this paper develops the mathematics to describe the link between slowly-varying nonstationarities and trend-like structures in the time marginals. Also, the experimental study is extended significantly, by testing more nonstationary processes against a longer list of alternative methods.

This paper is organized as follows. In Section 2, we show how nonstationarities that vary slow in time can appear as trends in the time marginal. We define how to approximate these trends and assess their significance by means of the trend importance estimator. In Section 3, we describe the framework and the resampling method needed to build the hypothesis test. In Section 4, we analyze the behavior of the trend importance estimator in stationary and nonstationary situations. In Section 5, we propose a model for the distribution of this estimator under stationarity. This model allows for characterizing a hypothesis test for stationarity in Section 6. The experimental study and the conclusions are shown in Sections 7 and 8, respectively.

2. The rationale behind the framework

This paper proposes a method for detecting slowly-varying nonstationarities that can be seen as a trend in the time marginal $y(n)$ of the time-varying spectrum of the signal. As illustration, Fig. 1 shows the time-varying spectra and the time marginals of a stationary Gaussian signal ($x_1(n)$) and a nonstationary one ($x_2(n)$), (Fig. 1a and d, respectively). The mean and variance of the nonstationary process start to increase slowly at $n = N/2$. Notice the difference between the two TF representations (Fig. 1b and e) and the trend-like behavior that can be seen in the time marginal of the nonstationary process (Fig. 1f).

For testing nonstationary behaviors that appear as temporal structures in the time marginal $y(n)$, the tasks of computing $y(n)$ should be performed. A straightforward approach to compute $y(n)$ is to intregrate over the frequency axis the TF representation of $x(n)$ [19]. However, such approach would require the estimation of full TF representations, for the signal $x(n)$ and any possible stationarized reference used by the method [12]. Unfortunately, resorting to TF representations to compute $y(n)$ would lead to the problems mentioned in Section 1 regarding computational complexity [13] and poor estimation of the spectral content at very low frequencies [14]. Thanks to the marginal properties of TF distributions [19], the time marginal $y(n)$ of the TF representation of a given real, discrete-time signal $x(n)$ can be easily computed by squaring the result of the numerical convolution of $x(n)$ and a given short-time window $h(n)$:

$$y(n) = \Big[x(n) * h(n)\Big]^2. \tag{1}$$

Doing so, we can compute $y(n)$ without having to deal with the estimation of TF representations. According to the properties of quadratic TF distributions, (1) is an approximation of the instantaneous power ($|x(n)|^2$) of the signal [19]. Here, the study of stationarity is reduced to the study of the time marginal computed by (1). Therefore, the proposed method is developed in the time domain, as it uses only information coded in $y(n)$. In following section, we show how the computation of $y(n)$ via (1) is affected if stationarity does not hold.

2.1. The influence of a nonstationary behavior on the time marginal computation

Let us assume the short-time window $h(n)$ in (1) contains L samples, or weights, given by $h(0), h(1), \dots, h(L-1)$, where $L \ll N$ and N is the number of samples of the signal $x(n)$. Then, (1) can be written as follows:

$$y(n) = \left[\sum_{l=0}^{L-1} h(l)x(n-l)\right]^2, \tag{2}$$

The choice of the (deterministic) window function $h(n)$ and its size L will be discussed in Section 2.6. After some algebra, one can rewrite (2) as follows:

$$y(n) = \sum_{l=0}^{L-1} h^2(l)x^2(n-l) + 2\sum_{l=1}^{L-1}\sum_{j=0}^{L-1-l} h(j)h(j+l)x(n-j)x(n-j-l) \tag{3}$$

from where an expression for the expected value of $y(n)$ can be obtained

$$\mathbb{E}[y(n)] = \sum_{l=0}^{L-1} h^2(l)\mathbb{E}[x^2(n-l)] + 2\sum_{l=1}^{L-1}\sum_{j=0}^{L-1-l} h(j)h(j+l)\mathbb{E}[x(n-j)x(n-j-l)]. \tag{4}$$

Now, we define the time-varying autocorrelation function $R_x(n, l)$ of $x(n)$ [20]:

$$R_x(n, l) = \mathbb{E}[x(n)x^*(n-l)], \tag{5}$$

which is function of both time n and lag l. Given (5), one can then express (4) as function of $R_x(n, l)$:

$$\mathbb{E}[y(n)] = \sum_{l=0}^{L-1} h^2(l)R_x(n-l, 0)$$