

Statistical hypothesis testing with time-frequency surrogates to check signal stationarity

Cédric Richard, André Ferrari, Hassan Amoud, Paul Honeine, Patrick Flandrin, Pierre Borgnat

► To cite this version:

Cédric Richard, André Ferrari, Hassan Amoud, Paul Honeine, Patrick Flandrin, et al.. Statistical hypothesis testing with time-frequency surrogates to check signal stationarity. IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP-10, Mar 2010, Dallas, United States. IEEE, 2010. <ensl-00476017>

HAL Id: ensl-00476017

<https://hal-ens-lyon.archives-ouvertes.fr/ensl-00476017>

Submitted on 23 Apr 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

STATISTICAL HYPOTHESIS TESTING WITH TIME-FREQUENCY SURROGATES TO CHECK SIGNAL STATIONARITY

Cédric Richard⁽¹⁾, *André Ferrari*⁽¹⁾, *Hassan Amoud*⁽²⁾, *Paul Honeine*⁽²⁾
Patrick Flandrin⁽³⁾, *Pierre Borgnat*⁽³⁾

⁽¹⁾ Laboratoire Fizeau (UMR CNRS 6525), Observatoire de la Côte d'Azur, Université de Nice Sophia-Antipolis, France

⁽²⁾ Institut Charles Delaunay (FRE CNRS 2848), Université de Technologie de Troyes, France

⁽³⁾ Laboratoire de Physique (UMR CNRS 5672), École Normale Supérieure de Lyon, France

ABSTRACT

An operational framework is developed for testing stationarity relatively to an observation scale. The proposed method makes use of a family of stationary surrogates for defining the null hypothesis of stationarity. As a further contribution to the field, we demonstrate the strict-sense stationarity of surrogate signals and we exploit this property to derive the asymptotic distributions of their spectrogram and power spectral density. A statistical hypothesis testing framework is then proposed to check signal stationarity. Finally, some results are shown on a typical model of signals that can be thought of as stationary or nonstationary, depending on the observation scale used.

Index Terms— Time-frequency analysis, stationarity test, surrogate, spectrogram, probability density function

1. INTRODUCTION

Time-frequency representations provide a powerful tool for nonstationary signal analysis and classification, and cover a wide range of applications [1]. Considering stationarity is central in many signal processing applications, which raises the operationally important issue of how testing stationarity. Recently, the authors have made use of a family of stationary surrogate signals for defining the null hypothesis of stationarity and, based upon this information, to derive tests operating in the time-frequency domain. Two classes of approaches have been considered in [2, 3]. The first one uses suitably chosen distances between local and global spectra. The second one is implemented as a one-class classifier, where time-frequency features are extracted from the surrogates to generate a learning set for stationarity. In [4], time-frequency learning machines have been used to test stationarity, based on one-class support vector machine and the set of surrogates. This approach takes full advantage of the use of the whole time-frequency representations of surrogates, compared with the arbitrary time-frequency features considered previously. Unfortunately, all these methods are often hampered by the large

number of surrogates required to analyze and test stationarity, resulting in an increased computation time and memory space. This drawback comes mainly from the relative lack of knowledge about the statistical properties of surrogates and of their time-frequency distributions.

In the spirit of [5], where the authors studied the probability density function (pdf) of the spectrogram of correlated Gaussian signals, we derive here the asymptotic pdf of the spectrogram of surrogates. It allows us to propose a statistical test for detecting nonstationarity without any need to generate surrogates. This work does not only provide important new insights in time-frequency analysis of the surrogate signals, but it also offers a means to understand the theoretical background. The remainder of the paper is organized as follows. In Sect. 2, the general framework of the proposed approach is outlined, detailing the time-frequency rationale of the method and motivating the use of surrogate data for characterizing the null hypothesis of stationarity. The strict-sense stationarity of surrogate signals is also demonstrated here. This property is exploited in Sect. 3 to derive an asymptotic statistical model for their spectrogram. A statistical hypothesis testing framework is then proposed to check signal stationarity. Some simulation results are shown in Sect. 4 on a typical model of signals that can be thought of as stationary or nonstationary, depending on the observation scale.

2. STATIONARIZATION VIA SURROGATES

Stationarity refers to a strict invariance of statistical properties with respect to time shifts. This theoretical definition can be loosely relaxed so as to encompass stationarity over some limited interval of observation. In order to test this property, it has been proposed in [2, 3] that a reference of stationarity be defined directly from the signal itself. The procedure consists of generating a family of stationarized signals which have the same psd as the initial signal. For an identical marginal spectrum over the same observation interval, nonstationary processes are expected to differ from stationary ones by some

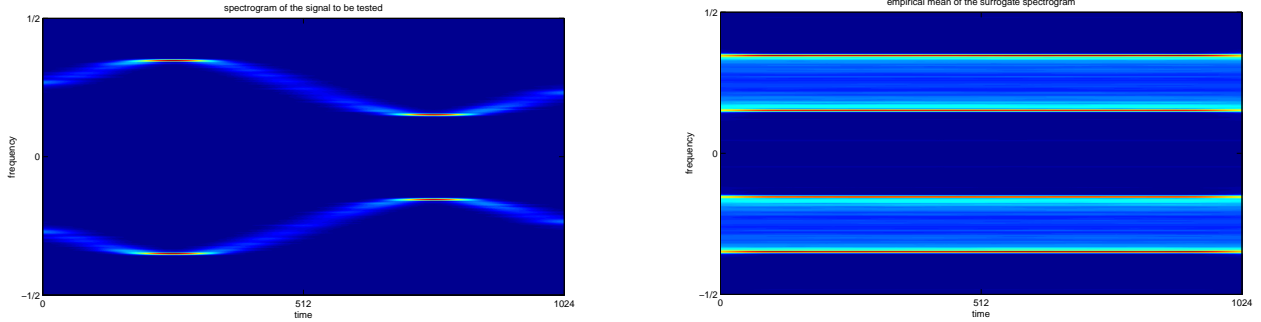


Fig. 1. Spectrogram of the FM signal (12) in the case where $T = T_0$ (left), and empirical mean of the spectrogram of its surrogates (right).

structured organization in time, hence in their time-frequency distribution. Surrogate data technique [6] is an appropriate solution to generate a family of stationarized signals, since it destroys the time-varying structures in the signal phase while keeping its power spectral density (psd) unchanged. In practice, this is achieved by keeping unchanged the magnitude of its Fourier transform, and replacing its phase by a i.i.d. one. More formally, let us consider the continuous-time signal $x(t)$ with Fourier transform $X(f)$ such that¹

$$X(f) = \int x(t) e^{-j2\pi ft} dt. \quad (1)$$

The surrogate signals $s(t)$ of $x(t)$ are constructed from the magnitude $A(f) = |X(f)|$ as follows:

$$s(t) = \int A(f) e^{j\Psi(f)} e^{j2\pi ft} df \quad (2)$$

with $\Psi(f)$ an i.i.d. phase. See illustration in Fig. 1. Let $\Phi_\Psi(u) = \mathbb{E}[e^{j\Psi u}]$ be the characteristic function of Ψ . We will assume in the sequel that

$$\Phi_\Psi(k) = 0, \forall k \in \mathbb{Z}^*. \quad (3)$$

Simple examples are random variables uniformly distributed over $[-\pi, \pi]$, $\Phi_\Psi(u) = \text{sinc}(\pi u)$, or the sum of M independent such random variables for which $\Phi_\Psi(u) = \text{sinc}^M(\pi u)$. Finally, it is noteworthy that the sum of two independent random variables where at least one verifies (3) also verifies (3).

2.1. Strict-sense stationarity

Recently in [3], the authors have demonstrated that surrogates are wide-sense stationary signals, that is, their first and second order moments are time-shift invariant. We shall now establish the strict-sense stationarity of surrogates, which is one of

¹Since a stochastic process cannot be represented as a standard Fourier integral, note that the Cramer representation should be considered. However, the use of Fourier transform will be preferred for notational simplicity.

the main contributions of this study. Let us derive the time-shift invariance of the $L + 1$ order cumulant

$$c(t; t_1, \dots, t_L) = \text{cum}(s(t)^{\epsilon_0}, s(t + t_1)^{\epsilon_1}, \dots, s(t + t_L)^{\epsilon_L})$$

where $\epsilon_i = \pm 1$ and $x^{\epsilon_i} = x^*$ when $\epsilon_i = -1$. We suggest the reader to refer, e.g., [7], for a detailed description of the tools related to high-order analysis of complex random processes. Using the multilinearity of the cumulants, we have

$$c(t; t_1, \dots, t_L) = \int A(f_0) \cdots A(f_L) \kappa(\mathbf{f}) e^{j2\pi t \sum_{i=0}^L \epsilon_i f_i} e^{j2\pi \sum_{i=1}^L \epsilon_i t_i f_i} d\mathbf{f}$$

with $\mathbf{f} = (f_0, \dots, f_L)$ and

$$\kappa(\mathbf{f}) = \text{cum}(e^{j\epsilon_0 \Psi(f_0)}, \dots, e^{j\epsilon_L \Psi(f_L)}).$$

Note that if one variable f_i in \mathbf{f} is different from the others, the corresponding random variable $e^{j\epsilon_0 \Psi(f_i)}$ is independent from the others and $\kappa(\mathbf{f}) = 0$. Consequently the joint cumulant of the surrogate simplifies to

$$c(t; t_1, \dots, t_L) = \kappa_{L+1} \int A(f)^{L+1} e^{j2\pi f t \sum_{i=0}^L \epsilon_i} e^{j2\pi f \sum_{i=1}^L \epsilon_i t_i} df$$

where $\kappa_{L+1} = \text{cum}(e^{j\epsilon_0 \Psi}, \dots, e^{j\epsilon_L \Psi})$. Application of the Leonov-Shiryaev formula to this cumulant leads to

$$\kappa_{L+1} = \sum_{\pi} (|\pi| - 1)! (-1)^{|\pi|-1} \prod_{B \in \pi} \Phi_\Psi(\sum_{i \in B} \epsilon_i) \quad (4)$$

where π runs through the list of all partitions of $\{0, \dots, L\}$ and B runs through the list of all blocks of the partition π . This expression can be simplified using assumption (3) and noting that $\sum_{i \in B} \epsilon_i \in \mathbb{Z}$. Consequently $\Phi_\Psi(\sum_{i \in B} \epsilon_i)$ is non zero, and necessary equal to 1, only if $\sum_{i \in B} \epsilon_i = 0$.

- In the case where L is even, whatever π , at least one block $B \in \pi$ has an odd cardinal. For this block, we have $\sum_{i \in B} \epsilon_i \in \mathbb{Z}^*$ and, consequently, $\kappa_{L+1} = 0$.

- If L is odd, the product in (4) is non zero, and necessarily equals 1, only when all the blocks B of the partition verifies $\sum_{i \in B} \epsilon_i = 0$. Since $\sum_B \sum_{i \in B} \epsilon_i = \sum_{i=0}^L \epsilon_i$, this product is non zero if, and only if, $\sum_{i=0}^L \epsilon_i = 0$.

As a conclusion, high-order cumulants of the surrogate signal $s(t)$ are non-zero only if $\sum_{i=0}^L \epsilon_i = 0$. This implies that $s(t)$ is a circular complex random signal. Moreover, substitution of the constraint in (4) leads to

$$c(t; t_1, \dots, t_L) = \kappa_{L+1} \int A(f)^{L+1} e^{j2\pi f \sum_{i=1}^L \epsilon_i t_i} df \quad (5)$$

which proves that surrogates are strict-sense stationary.

2.2. Polyspectra

The previous expression of the cumulant makes it possible to compute the L order polyspectra of surrogate signals. The polyspectra is defined as the L -dimension Fourier transform of the cumulants, namely,

$$\begin{aligned} S(f_1, \dots, f_L) &= \iint c(t; t_1, \dots, t_L) e^{-j2\pi \sum_{i=1}^L f_i t_i} dt_1 \dots dt_L \\ &= \kappa_{L+1} \int A(f)^{L+1} \left(\prod_{i=1}^L \int e^{-j2\pi(f_i - \epsilon_i f) t_i} dt_i \right) df \\ &= \kappa_{L+1} \int A(f)^{L+1} \left(\prod_{i=1}^L \delta(f_i - \epsilon_i f) \right) df \end{aligned}$$

Hence, the only non-zero values of the polyspectra are located over the line $\{(\epsilon_1 f, \dots, \epsilon_L f), f \in \mathbb{R}\}$ with

$$S(\epsilon_1 f, \dots, \epsilon_L f) = \kappa_{L+1} A(f)^{L+1} \quad (6)$$

For $L = 1$, note that the above equation leads to the surrogate psd $S(f) = A(f)^2$, which is obviously equal to the psd of the original signal. This result also shows that, among stationary signals, surrogates are only specific via their second-order characteristics. This justifies the use of second-order statistics, in the next section, to test stationarity.

The above properties have been derived in the continuous time case. They could have been considered in the discrete time case, which justifies their use as described below.

3. TESTING STATIONARITY WITH SURROGATES

The purpose of this section is to derive a test statistics to evaluate the stationarity of any discrete time signal $x(n)$. This composite test is based on the comparison of the second-order characteristics of the spectrogram of $x(n)$ with the spectrogram of its surrogates.

3.1. Asymptotic distribution of the spectrogram

We define the spectrogram $S(n, k)$ of the N -length surrogate signal $s(n)$ as

$$S(n, k) = \left| \sum_{\ell} s(\ell) w\left(\frac{\ell - n}{K}\right) e^{-j2\pi \frac{\ell k}{K}} \right|^2 \quad (7)$$

where $w(u)$ vanishes for $|u| > 1$, and $K < N$ the length of the discrete Fourier transform. The signal $s(n)$ being strictly stationary, the statistical properties of $S(n, k)$ are independent of n . For this reason, we will focus in the sequel on

$$S(0, k) = \left| \sum_{\ell} s(\ell) w\left(\frac{\ell}{K}\right) e^{-j2\pi \frac{\ell k}{K}} \right|^2 \quad (8)$$

The above expression coincides with the modified periodogram of $s(n)$, whose asymptotic distribution has been extensively studied in the literature. In [8], Theorem 5.2.7, the asymptotic distribution of $S(0, k)$ as K tends to infinity is derived under the assumption that $s(n)$ is strictly stationary with absolutely summable cumulants of all orders.

The strict-sense stationarity of $s(n)$ has been proved above. The absolute summability of the cumulants is essentially required as a sufficient condition for the existence of the polyspectra which, as seen in the previous section, are perfectly defined for surrogates. Consequently, we will assume for sufficiently large K that the distribution of $S(n, k)$ can be approximated by the asymptotic distribution of $S(0, k)$. In particular, given n , the $S(n, k)$ are (asymptotically) independent for $k \pm l \neq 0 [K]$ and $k \neq 0 [K]$. Moreover, we have

$$S(n, k) \sim \eta_w^2 A(k)^2 \chi_2^2 \quad (9)$$

where $A(k)^2$ is the psd of $s(n)$ and $\eta_w^2 = \sum_{\ell} w(\ell/K)^2$.

3.2. Test statistics

Let us now define a “normalized” instantaneous power $P_n(s)$ as follows

$$P_n(s) = \sum_k \frac{S(n, k)}{\eta_w^2 A(k)^2} \quad (10)$$

Independence with respect to k and (9) implies that the marginal distribution of P_n is χ_{2K}^2 . Choosing parameter K sufficiently large, we can use the standard approximation of a chi-square distribution

$$T_n(s) = \frac{P_n(s) - 2K}{\sqrt{4K}} \sim \mathcal{N}(0, 1). \quad (11)$$

As a consequence, we propose to reject the hypothesis of stationarity for $x(n)$ if the normal distribution hypothesis of $T_n(x)$ is rejected. This can be implemented via the Kolmogorov-Smirnov test, here applied to undersampled values of $T_n(x)$ with respect to n in order to ensure their approximate independence. The correlation time delay of $x(n)$ and the length of the window $w(n)$ should be considered to perform this downsampling efficiently.

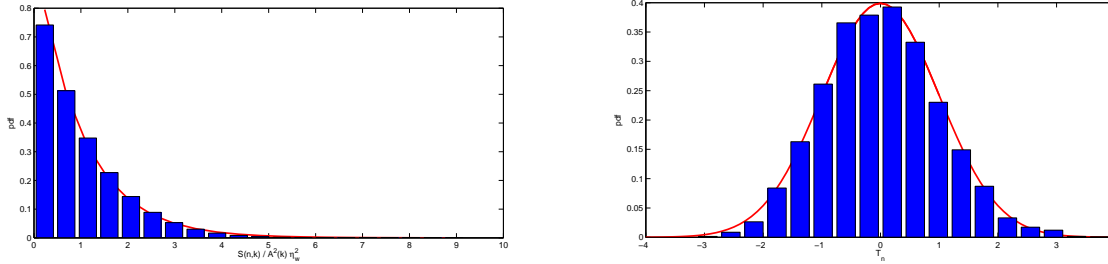


Fig. 2. Histograms of $S(n, k)/\eta_w^2 A(k)^2$ (left) and T_n (right) constructed from 5000 surrogate signals. Both are superimposed to the statistical models (9) and (11), respectively.

4. ILLUSTRATION

To test our method, we used the same FM signal as in [2]. While not covering all the situations of nonstationarity, this signal gives meaningful examples. It is modeled by

$$x(n) = \sin 2\pi(f_0 n + m \sin(2\pi n/N_0)) + e(n), \quad n \in N \quad (12)$$

with $m \leq 1$, f_0 the central frequency of the FM, N_0 its period, and $e(n)$ a zero-mean white Gaussian noise. Based on the relative values of N_0 and the signal duration N , three cases can be distinguished, see [2] for more details:

- $N \gg N_0$: The signal contains a great number of oscillations. This periodicity indicates a stationary regime.
- $N \approx N_0$: Only one oscillation is available. The signal can be considered as nonstationary.
- $N \ll N_0$: With a small portion of a period, there is no significative change in the frequency of the signal. It can be considered as stationary.

In our experiment, the signal duration N was set to 1024. The central frequency f_0 and the parameter m were fixed to 0.25 and 0.1, respectively. Signal-to-noise ration was set to 10 dB. Spectrograms were computed with a Hamming window of duration 256 samples. The relevance of the statistical modeling (9) is illustrated in Fig. 2 for $N = N_0$, where the χ^2_2 fit is superimposed to the histogram of $S(n, k)/\eta_w^2 A(k)^2$ constructed from 5000 surrogate signals. Fig. 2 also supports the claim that the pdf of T_n defined in (11) is reasonably well approached by the Gaussian pdf $\mathcal{N}(0, 1)$.

One hundred realizations of the FM signal (12) were finally tested in the following cases: $N_0 = N/10$, $N_0 = N$ and $N_0 = 10N$. The Kolmogorov-Smirnov test always rejected the hypothesis of stationarity in the second case, at significance level 0.05, while accepting it in the other two cases.

5. CONCLUSION

A new statistical framework was proposed for characterizing stationarity from a time-frequency viewpoint. A key point of the method is that the hypothesis of stationarity is defined

statistically by a class of surrogate signals which all share the same average spectrum as the analyzed signal. We demonstrated the strict-sense stationarity of surrogates and we exploited this property to derive the asymptotic distributions of their spectrogram and power spectral density. A statistical hypothesis test was finally presented to check signal stationarity.

6. REFERENCES

- [1] P. Flandrin, *Time-Frequency/Time-Scale Analysis*, Academic Press Inc, 1998.
- [2] J. Xiao, P. Borgnat, P. Flandrin, and C. Richard, "Testing stationarity with surrogates - a one-class SVM approach," in *Proc. IEEE Statistical Signal Processing*, Madison, WI, USA, August 2007, pp. 720–724.
- [3] P. Borgnat, P. Flandrin, P. Honeine, C. Richard, and J. Xiao, "Testing stationarity with surrogates: A time-frequency approach," *submitted to IEEE Trans. Signal Processing*, 2009.
- [4] H. Amoud, P. Honeine, C. Richard, P. Borgnat, and P. Flandrin, "Time-frequency learning machines for nonstationarity detection using surrogates," in *Proc. IEEE Statistical Signal Processing (SSP)*, Cardiff, Wales, UK, August 2009.
- [5] Julien Huillery, Fabien Millioz, and Nadine Martin, "On the Description of Spectrogram Probabilities with a Chi-Squared Law," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2249–2258, 06 2008.
- [6] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. Doynne Farmer, "Testing for nonlinearity in time series: the method of surrogate data," *Physica D: Nonlinear Phenomena*, vol. 58, no. 1-4, pp. 77–94, 1992.
- [7] P. O. Amblard, M. Gaeta, and J. L. Lacoume, "Statistics for complex variables and signals – part I and II," *Signal Processing*, vol. 53, no. 1, pp. 1 – 25, 1996.
- [8] D.R. Brillinger, *Time Series: Data Analysis and Theory*, McGraw-Hill, 1981.