Testing Stationarity with Time-Frequency Surrogates

Jun Xiao, Pierre Borgnat, Patrick Flandrin

TESTING STATIONARITY WITH TIME-FREQUENCY SURROGATES

Jun Xiao, Pierre Borgnat, and Patrick Flandrin

Physics Department, École Normale Supérieure de Lyon
46 allée d'Italie, 69364, Lyon Cedex 07, France
phone: + 33 (0) 4 72 72 81 60, fax: + 33 (0) 4 72 72 80 80
email: {Jun.Xiao, pborgnat, flandrin}@ens-lyon.fr

ABSTRACT

A method is proposed for testing stationarity in an operational sense, i.e., by both including explicitly an observation scale in the definition and elaborating a stationarized reference so as to reject the null hypothesis of stationarity with a controlled level of statistical significance. While the approach is classically based on comparing local vs. global features in the time-frequency plane, the test operates with a family of stationarized surrogates whose analysis allows for a characterization of the null hypothesis. The general principle of the method is outlined, practical issues related to its actual implementation are discussed and a typical example is provided for illustrating the approach and supporting its effectiveness.

1. INTRODUCTION

Testing stationarity is an important issue in several respects. As far as methods are concerned, signal processing is equipped with many powerful algorithms devoted to stationary processes, whose applicability should therefore be first assessed prior using them. Turning to interpretation, rejecting stationarity (and measuring some degree of nonstationarity) is of primary importance in numerous applications ranging from exploratory data analysis to diagnosis or surveillance.

However, whereas the concept of (weak, or second-order) stationarity is well-defined in theory, it turns out that it generally cannot be used as such in practice. Indeed, stationarity refers to a strict invariance of statistical properties over time, but common practice generally considers this invariance in a looser sense, relatively to some (explicit or implicit) observation scale. This certainly agrees with the physical intuition according to which a signal might appear, e.g., as both "short-term stationary" and "long-term nonstationary" (think of speech), but this also calls for a well-defined framework encompassing the observation scale as part of the definition as well as the possibility of testing such a stationarity with a controlled level of statistical significance.

Several attempts in this direction can be found in the literature, mostly based on concepts such as *local stationarity* [12]. Most of them however share the common philosophy of comparing statistics of adjacent segments, with the objective of detecting changes in the data [4, 8] and/or segmenting it over homogeneous domains [9] rather than addressing the aforementioned issue. Early attempts have nevertheless be made in this direction too by contrasting local properties with global ones [10, 11], but not necessarily properly phrased in terms of hypothesis testing.

It is therefore the purpose of this contribution to propose and describe an approach aimed at deciding whether an observed signal can be considered as stationary, *relatively to a given observation scale*, and, if not, to give an *index* as well as a typical *scale* of nonstationarity.

The paper is organized as follows. In Sect. 2, the general framework of the proposed approach is outlined. Sect. 2.1 details the time-frequency rationale of the method, whereas Sect. 2.2 is concerned with the introduction of *surrogate data* for characterizing the null hypothesis of stationarity and constructing the test. The test itself is discussed in Sect. 2.3, from which both an index and a scale of nonstationarity are defined in Sect. 2.4. The actual implementation of the method is discussed in Sect. 3, with elements regarding test signals in Sect. 3.1, distance measures in Sect. 3.2 and thresholds in Sect. 3.3. An illustration is given in Sect. 3.4 for supporting the efficiency of the method and, finally, some of the many possible variations and extensions are briefly outlined in Sect. 4.

2. GENERAL FRAMEWORK

Second order stationary processes are a special case of the more general class of (nonstationary) harmonizable processes, for which time-varying spectra can be properly defined [5]. When the analyzed process happens to be stationary, those time-varying spectra may reduce to the classical (stationary, time-independent) Power Spectrum Density (PSD) when suitably chosen: this holds true, e.g., for the Wigner-Ville Spectrum (WVS) [5]. In the case of more general definitions that can be considered as estimators of the WVS (e.g., spectrograms), the key point is that stationarity still implies time-independence: the time-varying spectra identify, at each time instant, to some frequency smoothed version of the PSD. The basic idea underlying the approach proposed here is therefore that, when considered over a given duration, a process will be referred to as *stationary relatively to this observation scale* if its time-varying spectrum undergoes no evolution or, in other words, if the local spectra at all different time instants are statistically similar to the global spectrum obtained by marginalization. This idea has already been pushed forward [10, 11], but the novelty is to address the significance of the difference "local vs. global" by elaborating from the data itself a *stationarized* reference serving as the null hypothesis for the test.

2.1 The Time-Frequency Approach

Given a signal $x(t)$, we compute a multitaper spectrogram estimate of its WVS according to [2]

$$S_{x,K}(t,f) = \frac{1}{K}\sum_{k=1}^{K} S_x^{(h_k)}(t,f),$$

where the $\{S_x^{(h_k)}(t,f), k=1,\dots K\}$ stand for the K spectrograms computed with the K first Hermite functions as short-time windows $h_k(t)$:

$$S_x^{(h_k)}(t,f) = \left| \int x(s)\, h_k(s-t)\, e^{-i2\pi fs}\, ds \right|^2 .$$

The multitaper approach is preferred to a classical spectrogram in order to reduce the level of statistical fluctuations of the nonstationary spectrum without recoursing to a time-averaging step which would be unsatisfactory in the context of (non-)stationarity estimation. The multitaper spectrogram is in fact evaluated only at N time positions $\{t_n, n=1,\dots N\}$, with a spacing $t_{n+1}-t_n$ which is an adjustable fraction of the temporal width T_h of the K windows $h_k(t)$. Given therefore the set of spectral "slices" $\{S_{x,K}(t_n,f), n=1,\dots N\}$, we can compute an average spectrum via the marginalization in time:

$$\langle S_{x,K}(t_n,f)\rangle_{n=1,\dots N} = \frac{1}{N}\sum_{n=1}^{N} S_{x,K}(t_n,f)$$

and compare it to each spectral slice according to some dissimilarity measure $\kappa(.,.)$ (possible choices for this measure will be discussed in Sect. 3.2), thus leading to the series of "distances" over the time interval from t_1 to t_N:

$$\{c_n^{(x)} := \kappa\left(S_{x,K}(t_n,.), \langle S_{x,K}(t_n,.)\rangle_{n=1,\dots N}\right), n=1,\dots N\}. \quad (1)$$

In the idealized case where $x(t)$ would be stationary and the estimation perfect, all those coefficients would be zero. In practice however, they of course fluctuate and the issue is to determine whether the observed fluctuation is significant or not: this is where *surrogates* enter into the play.

2.2 Surrogates

The time-frequency interpretation of stationarity described above amounts to say that, for a same marginal spectrum over the same observation interval, *nonstationary* processes differ from *stationary* ones by some structured organization in time. Distinguishing between stationarity and nonstationarity would therefore be made easier if, besides the analyzed signal itself, we also had at our disposal some reference having the same marginal spectrum while being stationary. Since such a reference is generally not available, one possibility is to *create it from the data*: this is the rationale behind the idea of "surrogate data", a technique which has been introduced and widely used in the physics literature, mostly for testing nonlinearity [14, 16]. Up to some related proposal reported in [7], it seems to have never been used as proposed here for testing nonstationarity, i.e., in the following sense: given the observed signal $x(t)$, a set of J "surrogates" is computed in such a way that each of them has the same PSD as the original signal while being "stationarized".

As mentioned before, this can be achieved by destroying the organized phase structure controlling the nonstationarity of $x(t)$, if any. In practice, $x(t)$ is first Fourier transformed to $X(f)$, and the modulus of $X(f)$ is then kept unchanged while its phase is replaced by a random one, uniformly distributed over $[-\pi,\pi]$. This modified spectrum is finally (inverse) Fourier transformed, leading to as many (stationary) surrogate signals as phase randomizations are operated.

2.3 Stationarity Test

Let us label $\{s_j(t), j=1,\dots J\}$ the J surrogate signals obtained as just described. When they are analyzed as explained above for the original signal $x(t)$, we finally end up with a new collection of distances

$$\{c_n^{(s_j)} := \kappa\left(S_{s_j,K}(t_n,.), \langle S_{s_j,K}(t_n,.)\rangle_n\right), n=1,\dots N, j=1,\dots J\},$$

depending on both time indexes and randomizations.

In order to measure the fluctuations in time of the divergences $c_n^{(.)}$ between local spectra and global ones, we can use the l_2-distance defined by

$$L(g,h) := \frac{1}{N}\sum_{n=1}^{N}(g_n-h_n)^2, \quad (2)$$

for any pair of sequences $\{(g_n,h_n), n=1,\dots N\}$.

As far as the intrinsic variability of surrogate data is concerned, the dispersion of divergences under the null hypothesis of stationarity can be measured by the distribution of the J empirical variances

$$\left\{\Theta_0(j) = L\left(c^{(s_j)}, \langle c^{(s_j)}\rangle_{n=1,\dots N}\right), j=1,\dots J\right\}. \quad (3)$$

This distribution allows for the determination of a threshold γ above which the null hypothesis is rejected. The effective test is therefore based on the statistics

$$\Theta_1 = L\left(c^{(x)}, \langle c^{(x)}\rangle_{n=1,\dots N}\right) \quad (4)$$

and takes on the form of the one-sided test:

$$\begin{cases} \Theta_1 > \gamma & : \text{"}\textit{nonstationarity}\text{"}; \\ \Theta_1 < \gamma & : \text{"}\textit{stationarity}\text{"}. \end{cases} \quad (5)$$

2.4 Index and Scale of Nonstationarity

Assuming that the null hypothesis of stationarity is rejected, an *index of nonstationarity* can be introduced as a function of the ratio between the test statistics (4) and the mean value (or the median) of its stationarized counterparts (3):

$$\mathrm{INS} := \sqrt{\frac{\Theta_1}{\frac{1}{J}\sum_{j=1}^{J}\Theta_0(j)}}. \quad (6)$$

If the signal happens to be stationary, INS is expected to take a value close to unity whereas, the more nonstationary the signal, the larger the index.

Finally, it has to be remarked that, whereas the tested stationarity is globally relative to the time interval T over which the signal is chosen to be observed, the analysis still depends on the window length T_h of the spectrogram. Given T, the index INS will therefore be a function of T_h and, repeating the test with different window lengths, we can end up with a typical *scale of nonstationarity* SNS defined as:

$$\mathrm{SNS} := \frac{1}{T}\,\arg\max_{T_h}\{\mathrm{INS}(T_h)\}, \quad (7)$$

with T_h in the range of window lengths such that the prescribed threshold is exceeded in (5).

3. THE TEST IN ACTION

The principle of the test having been outlined, its actual implementation depends on a number of choices that have to be made and justified, regarding distances, surrogates, thresholds, etc.: addressing such issues is the purpose of this Section. Many options are however offered, that are moreover intertwined. A complete investigation of all possibilities and their combinations is clearly beyond the scope of this paper. Nevertheless, key features that are important for the test to be used in practice will be highlighted here.

3.1 Test signals

Setting specific parameters in the implementation is likely to end up with performance depending on the type of nonstationarity of the signal under test. Whereas no general framework can be given for encompassing all possible situations, two main classes of nonstationarities can be distinguished, which both give rise to a clear picture in the time-frequency plane: amplitude modulation (AM) and frequency modulation (FM). We will base the following discussions on such classes. In the first case (AM), a basic, stochastic representative of the class can be modelled as:

$$x(t) = (1+\alpha \sin 2\pi t/T_0)\, e(t), t \in T, \tag{8}$$

with $\alpha \leq 1$ and where $e(t)$ is white Gaussian noise, T_0 is the period of the AM and T the observation duration. In the second case (FM), a deterministic model can be defined as:

$$x(t) = \sin 2\pi (f_0 t + \alpha \sin 2\pi t/T_0), t \in T, \tag{9}$$

with $\alpha \leq 1$ and where f_0 is the central frequency of the FM, T_0 its period and T the observation duration.

3.2 Distances

Within the chosen time-frequency perspective, the proposed test (5) amounts to compare local spectra with their average over time thanks to some "distance" (1), and to decide that stationarity is to be rejected if the fluctuation of such descriptors (as given by (4)) is significantly larger than what would be obtained in a stationary case with a similar global spectrum. The choice of a distance (or dissimilarity) measure is therefore instrumental for contrasting local vs. global features.

Many approaches are available in the literature [1] that, without entering into too much details, can be broadly classified in two groups. In the first one, the underlying interpretation is that of a probability density function, one of the most efficient candidate being the well-known Kullback-Leibler (KL) divergence defined as

$$\kappa_{\mathrm{KL}}(G,H) := \int_\Omega (G(f)-H(f)) \log \frac{G(f)}{H(f)}\, df, \tag{10}$$

where, by assumption, the two distributions $G(.)$ and $H(.)$ to be compared are positive and normalized to unity over the domain Ω. In our context, such a measure can be envisioned for (always positive) spectrograms thanks to the probabilistic interpretation that can be attached to distributions of time and frequency [5].

A second group of approaches, which is more of a spectral nature, is aimed at comparing distributions in both shape and level. One of the simplest examples in this respect is the log-spectral deviation (LSD) defined as

$$\kappa_{\mathrm{LSD}}(G,H) := \int_\Omega \left| \log \frac{G(f)}{H(f)} \right| df. \tag{11}$$

Intuitively, the KL measure (10) should perform poorer than the LSD one (11) in the AM case (8), because of normalization. It should however behave better in the FM case (9), because of its recognized ability at discriminating distribution shapes. In order to take advantage of both measures, it is therefore proposed to combine them in some ad hoc way as

$$\kappa(G,H) := \kappa_{\mathrm{KL}}(\tilde{G},\tilde{H}).(1+\lambda\, \kappa_{\mathrm{LSD}}(G,H)), \tag{12}$$

with $\tilde{G}$ and $\tilde{H}$ the normalized versions of G and H, and where λ is a trade-off parameter to be adjusted. In practice, the choice $\lambda = 1$ ends up with a good performance, as justified in Fig. 1 (the performance measure used in this Figure is the inverse of the maximum value (over T_h) of the index of nonstationarity INS defined in (6), i.e., an inverse measure of contrast).

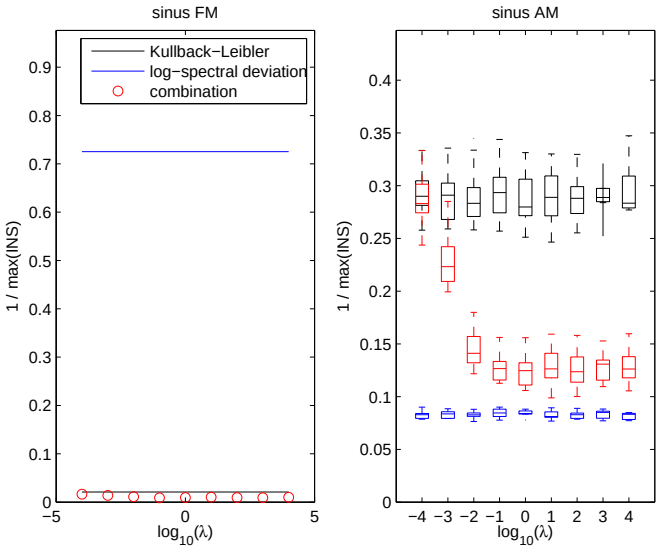

Figure 1: Choosing a distance. The inverse of the maximum value (over T_h) of the index of nonstationarity INS defined in (6) is used as a performance measure. Comparing the Kullback-Leibler (KL) divergence with the log-spectral deviation (LSD), a better result (i.e., a lower value) is obtained with KL (black) in the FM case (left, with $\alpha = 0.03$), and with LSD (blue) in the AM case (right, with $\alpha = 0.5$). A better balanced performance is obtained when using the combined distance (red) defined in (12): in the FM case, this measure performs best, and in the AM case it achieves a good contrast when $\lambda \geq 1$. In the AM case, the boxplots resulting from 10 realizations of the process are displayed.

3.3 Distribution of Surrogates and Threshold

The basic ingredient (and originality) of the approach is the use of surrogate data for creating signals whose spectrum is

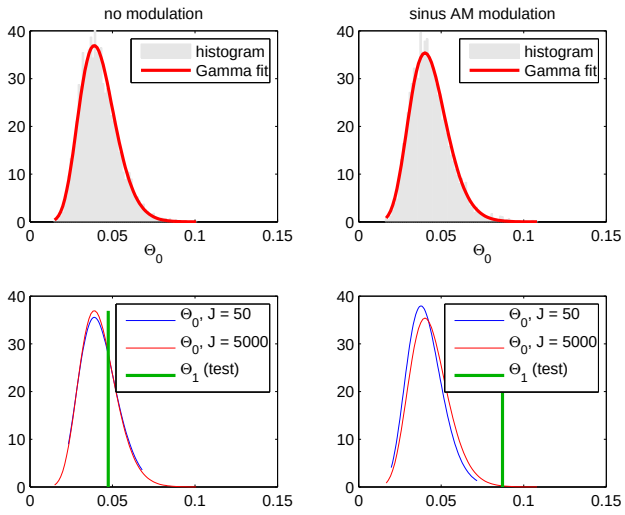

Figure 2: Distribution based on surrogates. The top row superimposes empirical histograms of the variances (3) based on $J = 5000$ surrogates (grey) and their Gamma fits (red), in the case of a white Gaussian noise without (left) and with (right) a sinusoidal AM (with $\alpha = 0.5$). The bottom row compares the corresponding probability density functions, as parameterized by using $J = 50$ (blue) and 5000 (red) surrogates. The values of the test statistics (4) computed on the analyzed signal are pictured in both cases as green lines.

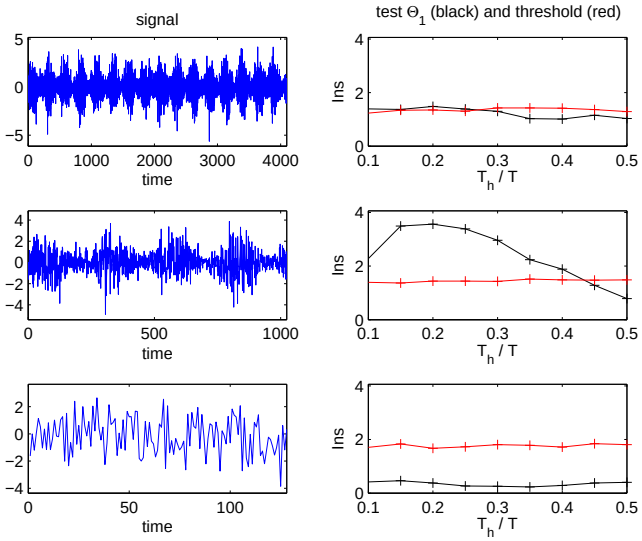

Figure 3: AM example ($\alpha = 0.5$). In the case of the same signal (8) observed over different time intervals (left column), the indexes of nonstationarity INS (right column, black) are consistent with the physical interpretation according to which the observation can be considered as stationary at macroscale (top row), nonstationary at mesoscale (middle row) and stationary again at microscale (bottom row). The threshold (red) of the stationarity test (5) is calculated with a confidence level of 95% and represented in term of INS as $\sqrt{\gamma/(\sum_j \Theta_0(j)/J)}$, with $J = 50$. In the nonstationary case, the position of the maximum of INS also gives an indication of a typical scale of nonstationarity.

identical to that of the original one while being stationarized by getting rid of a well-defined structuration in time. Since those surrogates can be viewed as distinct, independent realizations of the stationary counterpart of the analyzed signal, the central part of the test is based on the statistical distribution of the J variances given in (3).

When using the combined distance suggested above in Sect. 3.2, an empirical study (on both AM and FM signals) has shown that such a distribution can be fairly well approximated by a Gamma distribution. This makes sense since, according to (2), the test statistics basically sums up squared, possibly dependent quantities which themselves result from a strong mixing likely to act as a Gaussianizer. An illustration of the relevance of this modeling is given in Fig. 2, where Gamma fits are superimposed to actual histograms in the asymptotic regime ($J = 5000$ surrogates). Assuming the Gamma model to hold, it is possible to estimate its 2 parameters directly from the J surrogates, e.g., in a maximum likelihood sense. In this respect, Fig. 2 also supports the claim that the "theoretical" probability density function (more precisely, its estimate in the asymptotic regime) can be reasonably well approached with a reduced number of surrogates (typically, $J \approx 50$). Finally, the value of the test (4), computed on the actual signals under study, is also plotted and shown to stand in the middle of the distribution in the stationary case while clearly appearing as an outlier in the considered nonstationary situation.

Given the Gamma model for the distribution based on surrogates, it becomes straightforward to derive a threshold above which the null hypothesis of stationarity is rejected with a given statistical significance.

3.4 Illustration

In order to illustrate the proposed approach and to support its effectiveness, a simple example is given in Fig. 3. The analyzed signal consists of one realization of an AM process of the form (8). Depending on the relative values of T_0 and T, three regimes can be intuitively distinguished:

1. if $T \gg T_0$ (macroscale), many oscillations are present in the observation, creating a sustained, well-established quasi-periodicity that corresponds to a form of stationarity;
2. if $T \approx T_0$ (mesoscale), emphasis is put on the local evolutions due to the AM, suggesting to rather consider the signal as nonstationary, with some typical scale;
3. if $T \ll T_0$ (microscale), no significant difference in amplitude is perceived, turning back to stationarity.

What is shown in Fig. 3 is that such interpretations of *relative stationarity* are precisely evidenced by the proposed test. They are moreover quantified in the sense that, when the null hypothesis of stationarity is rejected (middle diagram), both an *index* and a *scale* of nonstationarity can be defined according to (6) and (7). In the present case, the maximum value of INS is obtained for SNS $= T_h/T \approx 0.2$, in qualitative accordance with the 4 AM periods entering the observation window.

In this specific example, the data could have been be referred to as *cyclostationary* and analyzed by tools dedicated

to such processes [15]. However, it has to be stressed that no such a priori modeling is assumed in the proposed methodology, and that the existence of a typical scale of stationarity (related to the periodic correlation) naturally emerges from the analysis.

4. CONCLUSION

A new approach has been proposed for testing stationarity from a time-frequency viewpoint, relatively to a given observation scale. A key point of the method is that the null hypothesis of stationarity (which corresponds to time-invariance in the time-frequency spectrum) is statistically characterized on the basis of a set of surrogates which all share the same average spectrum as the analyzed signal while being stationarized.

The basic principles of the method have been outlined, with a number of considerations related to its implementation, but it is clear that the proposed framework still leaves room for more thorough investigations as well as variations and/or extensions. In terms of time-frequency distributions for instance, one could imagine to go beyond spectrograms and take advantage of more recent advances [17, 19]. Turning to the test itself, the way the estimated time-frequency spectrum fluctuates in time has been considered here by comparing local features (frequency "slices") to a global one (the average spectrum resulting from marginalization) thanks to some distance measure. This is a classical approach, but it has the drawback of calling for the choice of a (more or less arbitrary) distance and the evaluation of associated distributions. A different possibility would be to look at the statistical decision problem from a learning perspective [3, 6, 13], and to consider surrogates as defining a learning set. Extraction of suitable descriptors from this learning set may therefore allows to make use of the powerful machinery of kernel methods, and especially of one-class support vector machines. This is under current investigation and will be reported elsewhere [18].

REFERENCES

[1] M. Basseville, "Distance measures for signal processing and pattern recognition," *Signal Proc.*, vol. 18, no. 4, pp. 349-369, 1989.

[2] M. Bayram and R.G. Baraniuk, "Multiple window time-varying spectrum estimation,"in *Nonlinear and Nonstationary Signal Processing* (W.J. Fitzgerald et al., eds.), pp. 292–316, Cambridge Univ. Press, 2000.

[3] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge Univ. Press, 2000.

[4] M. Davy and S. Godsill, "Detection of abrupt signal changes using Support Vector Machines: An application to audio signal segmentation," in *Proc. IEEE ICASSP-02*, Orlando (FL), 2002.

[5] P. Flandrin, *Time-Frequency/Time-Scale Analysis*. San Diego: Academic Press, 1999.

[6] P. Honeiné, C. Richard and P. Flandrin, "Time-frequency learning machines," *IEEE Trans. on Signal Proc.*, to appear 2007. `http://prunel.ccsd.cnrs.fr/ensl-00118896`.

[7] C.J. Keylock, "Constrained surrogate time series with preservation of the mean and variance structure," *Phys. Rev. E*, vol. 73, pp. 030767.1-030767.4, 2006. 85, pp.

[8] H. Laurent and C. Doncarli, "Stationarity index for abrupt changes detection in the time-frequency plane," *IEEE Signal Proc. Lett.*, vol. 5, no. 2, pp. 43–45, 1998.

[9] S. Mallat, G. Papanicolaou and Z. Zhang, "Adaptive covariance estimation of locally stationary processes," *Ann. of Stat.*, vol. 24, no. 1, pp. 1–47, 1998.

[10] W. Martin, "Measuring the degree of non-stationarity by using the Wigner-Ville spectrum," in *Proc. IEEE ICASSP-84*, San Diego (CA), pp. 41B.3.1–41B.3.4, 1984.

[11] W. Martin and P. Flandrin, "Detection of changes of signal structure by using the Wigner-Ville spectrum," *Signal Proc.*, vol. 8, pp. 215–233, 1985.

[12] R.A. Silverman, "Locally stationary random processes," *IRE Trans. on Info. Theory*, vol. 3, pp. 182–187, 1957.

[13] B. Schölkopf and A.J. Smola, *Learning With Kernels: Support Vector Machines, Regularization, Optimization and Beyond*. MIT Press, 2001.

[14] T. Schreiber and A. Schmitz, "Improved surrogate data for nonlinearity tests," *Phys. Rev. Lett.*, vol. 77, no. 4, pp. 635–638, 1996.

[15] E. Serpedin, F. Panduru, I. Sari and G.B. Giannakis, "Bibliography on cyclostationarity," *Signal Proc.*, vol. 85, no. 12, pp. 2233-2303, 2005.

[16] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian and J.D. Farmer, "Testing for nonlinearity in time series: The method of surrogate data," *Physica D*, vol. 58, no. 1, pp. 77–94, 1992.

[17] J. Xiao and P. Flandrin, "Multitaper time-frequency reassignment," in *Proc. EUSIPCO-06*, Firenze (I), 2006.

[18] J. Xiao, P. Borgnat, P. Flandrin and C. Richard, "Testing stationarity with surrogates — A one-class SVM approach," in *Proc. IEEE Stat. Sig. Proc. Workshop SSP-07*, Madison (WI), to appear 2007.

[19] J. Xiao and P. Flandrin, "Multitaper time-frequency reassignment for nonstationary spectrum estimation and chirp enhancement," *IEEE Trans. on Signal Proc.*, to appear 2007. `http://prunel.ccsd.cnrs.fr/ensl-00111665`.