MULTI-SCALE TEST PROCEDURE FOR NON-STATIONARITY IN SHORT AND LONG MEMORY TIME SERIES

Olaf Kouamo *Cédric Gouy-Pailler*

CEA, LIST, 91191 Gif-sur-Yvette CEDEX, FRANCE

ABSTRACT

In this paper, we develop a test procedure for non-stationarity for possibly long-memory processes. Contrary to most of the proposed methods, the test procedure has the same distribution for short-range and long-range dependence stationary processes.
Such tests have been already proposed in [1], but these authors do not have taken into account the dependence of the wavelet coefficients within scales and between scales. We also propose an application to electric power consumption monitoring.

Index Terms— Long memory process, Wavelet, Change points.

1. INTRODUCTION

The danger of confusing long-range dependence with non-stationarity has been pointed out by many authors. Finding an answer to this difficult question is of importance to model time-series showing trend-like behavior, in particular the river run-off in hydrology, the historical temperature in the study of climates changes, asset evolution in financial econometrics or packet counts in network traffic engineering. There has been a long lasting controversy to decide whether the deviations to short memory stationarity should be attributed to long-range dependence or are related to the presence of breakpoints in the mean, the variance, the covariance function or other types of more sophisticated structural changes. The links between non-stationarity and long-range dependence (LRD) have been pointed out by many authors ([2], [3] [4], [5]) and the references therein.
The main goal of this paper is to develop a test procedure for non-stationarity for possibly long-memory processes. Contrary to most of the proposed methods, the test procedure has the same distribution for short-range and long-range dependence covariance stationary processes, which means that this test is able to detect the presence of non-stationarity for processes showing long-range dependence; in addition, the test procedure is shown to be robust to the presence of slowly varying trends. The procedure described in this paper deals with the problem of detecting changes which may occur in the spectral content of a process. We will consider a process X which, before and after the change, is not necessary stationary but whose difference of at least a given order is stationary, so that polynomial trends up to that order can be discarded. Denote by $\mathbf{\Delta}X$ the first order difference of X,

$$[\mathbf{\Delta}X]_n \stackrel{\text{def}}{=} X_n - X_{n-1}, \quad n \in \mathbb{Z} ,$$

and define, for an integer $K \geq 1$, the K-th order difference recursively as follows: $\mathbf{\Delta}^K = \mathbf{\Delta} \circ \mathbf{\Delta}^{K-1}$. As defined in [6], a process X is said to be K-th order difference stationary if $\mathbf{\Delta}^K X$ is covariance stationary. Let f be a non-negative 2π-periodic symmetric function such that there exists an integer K satisfying, $\int_{-\pi}^{\pi} |1 - \mathrm{e}^{-\mathrm{i}\lambda}|^{2K} f(\lambda) \mathrm{d}\lambda < \infty$. We say that the process X admits *generalized spectral density* f if $\mathbf{\Delta}^K X$ is weakly stationary and with spectral density function

$$f_K(\lambda) = |1 - \mathrm{e}^{-\mathrm{i}\lambda}|^{2K} f(\lambda) . \tag{1}$$

This class of process include both short-range dependent and long-range dependent processes, but also unit-root and fractional unit-root processes.
In this paper, we consider the so-called *a posteriori* or *retrospective* method (see [7, Chapter 3]). The proposed test is formulated in the wavelet domain, where a change in the generalized spectral density results in a change in the covariance structure of the wavelet coefficients across scales. Such tests have been already proposed in [1], but these authors do not have taken into account the dependence of the wavelet coefficients within scales and across scales. Therefore, the asymptotic distribution of the test they have proposed only holds in case i.i.d which is not the case in general settings.
The paper is organised as follows. In Section 2, we introduce the wavelet setting and the relation ship between the generalized spectral density and the variance of wavelet coefficients. Section 3 is consecrated to our main assumptions and results. In Section 4.1, finite sample performance of the test is studied based on Monte Carlo simulations. To end, this test is applied to river run-off.

2. WAVELET SETTING

The wavelet setting involves two functions ϕ and ψ and their Fourier transforms
$\widehat{\phi}(\xi) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} \phi(t)\mathrm{e}^{-\mathrm{i}\xi t}\, \mathrm{d}t$ and $\widehat{\psi}(\xi) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} \psi(t)\mathrm{e}^{-\mathrm{i}\xi t}\, \mathrm{d}t$, and assume the following:

- (W-1) ϕ and ψ are compactly-supported, integrable, and $\widehat{\phi}(0) = \int_{-\infty}^{\infty} \phi(t)\, \mathrm{d}t = 1$ and $\int_{-\infty}^{\infty} \psi^2(t)\, \mathrm{d}t = 1$.
- (W-2) There exists $\alpha > 1$ such that $\sup_{\xi\in\mathbb{R}} |\widehat{\psi}(\xi)|\, (1 + |\xi|)^\alpha < \infty$.
- (W-3) The function ψ has M vanishing moments, *i.e.* $\int_{-\infty}^{\infty} t^m \psi(t)\, dt = 0$ for all $m = 0, \ldots, M-1$
- (W-4) The function $\sum_{k\in\mathbb{Z}} k^m \phi(\cdot - k)$ is a polynomial of degree m for all $m = 0, \ldots, M-1$.

Condition (W-2) ensures that this Fourier transform decreases quickly to zero. Condition (W-3) is an important characteristic of wavelets: it ensures that they oscillate and that their scalar product with continuous-time polynomials up to degree $M-1$ vanishes. Daubechies wavelets and Coiflets having at least two vanishing moments satisfy these conditions. Viewing the wavelet $\psi(t)$ as a basic template, define the family $\{\psi_{j,k}, j \in \mathbb{Z}, k \in \mathbb{Z}\}$ of translated and dilated functions $\psi_{j,k}(t) = 2^{-j/2}\, \psi(2^{-j}t - k), \quad j \in \mathbb{Z},\, k \in \mathbb{Z}$.

2.1. Discrete Wavelet Transform (DWT) in discrete time.

We now describe how the wavelet coefficients are defined in discrete time, that is for a real-valued sequence $\{x_k,\ k \in \mathbb{Z}\}$ and for a finite sample $\{x_k,\ k = 1, \ldots, n\}$. Using the scaling function ϕ, we first interpolate these discrete values to construct the following continuous-time functions

$$\mathbf{x}_n(t) \stackrel{\text{def}}{=} \sum_{k=1}^{n} x_k\, \phi(t-k) \quad \text{and}\ \mathbf{x}(t) \stackrel{\text{def}}{=} \sum_{k\in\mathbb{Z}} x_k\, \phi(t-k),\ t \in \mathbb{R}\ .$$

Without loss of generality we may suppose that the support of the scaling function ϕ is included in $[-\mathrm{T}, 0]$ for some integer $\mathrm{T} \geq 1$. Then $\mathbf{x}_n(t) = \mathbf{x}(t) \quad \text{for all} \quad t \in [0, n-\mathrm{T}+1]$. We may also suppose that the support of the wavelet function ψ is included in $[0, \mathrm{T}]$. With these conventions, the support of $\psi_{j,k}$ is included in the interval $[2^j k, 2^j(k+\mathrm{T})]$. The wavelet coefficient $W_{j,k}$ at scale $j \geq 0$ and location $k \in \mathbb{Z}$ is formally defined as the scalar product in $\mathrm{L}^2(\mathbb{R})$ of the function $t \mapsto \mathbf{x}(t)$ and the wavelet $t \mapsto \psi_{j,k}(t)$:

$$W_{j,k} \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} \mathbf{x}(t)\psi_{j,k}(t)\,\mathrm{d}t = \int_{-\infty}^{\infty} \mathbf{x}_n(t)\psi_{j,k}(t)\,\mathrm{d}t, \qquad j \geq 0, k \in \mathbb{Z}\ , \quad (2)$$

when $[2^j k, 2^j k + \mathrm{T})] \subseteq [0, n-\mathrm{T}+1]$, that is, for all $(j,k) \in \mathcal{I}_n$, where

$$\mathcal{I}_n \stackrel{\text{def}}{=} \{(j,k) :\ j \geq 0, 1 \leq k \leq n_j\} \quad \text{with} \quad n_j = 2^{-j}(n-\mathrm{T}+1) - \mathrm{T} + 1\ . \quad (3)$$

If $\mathbf{\Delta}^M X$ is stationary, then for [8, Eq 17] the process $\{W_{j,k}\}_{k\in\mathbb{Z}}$ of wavelet coefficients is stationary but the two-dimensional process $\{[W_{j,k}\,, W_{j',k}]^T\}_{k\in\mathbb{Z}}$ with $j \geq j'$ is not stationary. This is why we consider instead the stationary between scale process $\{[\mathbf{W}^T_{j,k}(j-j'),\, W_{j,k}]^T\}_{k\in\mathbb{Z}}$. where the superscript ^{T} denotes the transpose and $\mathbf{W}_{j,k}(u)$, $u = 0, 1, \ldots, j$, is defined as follows: $\mathbf{W}_{j,k}(u) \stackrel{\text{def}}{=} [W_{j-u,2^u k},\, W_{j-u,2^u k+1},\, \ldots, W_{j-u,2^u k+2^u-1}]^T$. For all $j, j' \geq 1$, the covariance function of the between scale process is given by

$$\operatorname{Cov}\left(W_{j,0}, \mathbf{W}_{j,k}(u)\right) = \int_{-\pi}^{\pi} \mathrm{e}^{\mathrm{i}\lambda k}\, \mathbf{D}_{j,u}(\lambda; f)\,\mathrm{d}\lambda\ , \quad (4)$$

where $\mathbf{D}_{j,u}(\lambda; f)$ is the cross-spectral density function of the between-scale process. For further details, we refer the reader to [8, Corollary 1]. The case $j = j'$ corresponds to the spectral density of the with-scale process $\{W_{j,k}\}_{k\in\mathbb{Z}}$.

2.2. The wavelet spectrum and the scalogram.

Let $X = \{X_t,\ t \in \mathbb{Z}\}$ be a real-valued process with wavelet coefficients $\{W_{j,k},\ k \in \mathbb{Z}\}$ and define $\sigma^2_{j,k} = \operatorname{Var}(W_{j,k})$. If $\mathbf{\Delta}^M X$ is stationary, then $\{W_{j,k},\ k \in \mathbb{Z}\}$, is also stationary. Then, the wavelet variance $\sigma^2_{j,k}$ does not depend on k, $\sigma^2_{j,k} = \sigma^2_j$. The sequence $(\sigma^2_j)_{j\geq 0}$ is called the *wavelet spectrum* of the process X. If moreover $\mathbf{\Delta}^M X$ is centered, the wavelet spectrum can be estimated by using the scalogram, defined as the empirical mean of the squared wavelet coefficients computed from the sample $X_1, \ldots, X_n$: $\widehat{\sigma}^2_j = \frac{1}{n_j}\sum_{k=1}^{n_j} W^2_{j,k}$. By [8, Proposition 1], if $K \leq M$, then the scalogram of X can be expressed using the generalized spectral density f appearing in (1) and the filters H_j defining in [8, Eq 14] as follows:

$$\sigma^2_j = \int_{-\pi}^{\pi} |H_j(\lambda)|^2\ f(\lambda)\,\mathrm{d}\lambda, \quad j \geq 0\ . \quad (5)$$

3. ASYMPTOTIC DISTRIBUTION OF THE TEST STATISTIC

Let $X_1, \ldots, X_n$ be the n observations of a time series, and denote by $W_{j,k}$ for $(j,k) \in \mathcal{I}_n$ with $\mathcal{I}_n$ defined in (3) the associated wavelet coefficients. In view of (5), if $X_1, \ldots, X_n$ are a n successive observations of a K-th order difference stationary process, then the wavelet variance at each given scale j should be constant. If the process X is not K-th order stationary, then it can be expected that the wavelet variance will change either gradually or abruptly (if there is a shock in the original time-series). This thus suggests to investigate the consistency of the variance of the wavelet coefficients. We will adapt the approach developed in [9], which uses cumulative sum (CUSUM) of squares to detect change points in the variance. Under the null hypothesis that X is K-th order stationary, a *multiple scale* procedure aims at testing that the scalogram in a range satisfies

$\mathcal{H}_0 : \sigma^2_{j,1} = \cdots = \sigma^2_{j,n_j}, \forall\ \ j \in \{J_1, J_1+1, \ldots, J_2\}$ where J_1 and J_2 are the *finest* and the *coarsest* scales included in the procedure, respectively. Consider the following process $Y_{J_1,J_2}[i] = \left(W^2_{J_2,i}, \sum_{u=0}^{1} W^2_{J_2-1,2i+u}, \ldots, \sum_{u=0}^{2^{(J_2-J_1)}-1} W^2_{J_1,2^{(J_2-J_1)}i+u}\right)^T$ and denote the corresponding partial sum process by

$$S_{J_1,J_2}(t) = \frac{1}{\sqrt{n_{J_2}}} \sum_{i=1}^{\lfloor n_{J_2} t\rfloor} Y_{J_1,J_2}[i]\ . \quad (6)$$

Because wavelet coefficients at different scales are not uncorrelated, both the *within-scale* and the *between scale* covariances need to be taken into account. We use The Bartlett estimator of the covariance matrix of the square wavelet's coefficients for scales $\{J_1, \ldots, J_2\}$, $\widehat{\Gamma}_{J_1,J_2}$ given by :

$$\widehat{\Gamma}_{J_1,J_2} = \sum_{\tau=-q(n_{J_2})}^{q(n_{J_2})} w_\tau[q(n_{J_2})]\widehat{\gamma}_{J_1,J_2}(\tau)\ , \quad (7)$$

where $\widehat{\gamma}_{J_1,J_2}$ is the usual empirical covariance sequence based on observations $Y_{J_1,J_2}[1], \ldots, Y_{J_1,J_2}[n_{J_2}]$ and, for a given integer q, the so-called Bartlett weights are defined by

$$w_l(q) = 1 - \frac{|l|}{1+q}\ , l \in \{-q, \ldots, q\}\ . \quad (8)$$

Theorem 1 *Suppose that X is a Gaussian process with generalized spectral density f. Let (ϕ, ψ) be a scaling and a wavelet function satisfying (W-1)-(W-4). Assume that $\Delta^M X$ is non-deterministic and centered, and that $\lambda^{2M} f(\lambda)$ is two times differentiable in λ with bounded second order derivative. Assume more that $q(n_{J_2}) \to \infty$ and $q(n_{J_2})/n_{J_2} \to 0$ as $n_{J_2} \to \infty$. Then as $n \to \infty$,*

$$\widehat{\Gamma}_{J_1,J_2} = \Gamma_{J_1,J_2} + O_P\left(\frac{q(n_{J_2})}{n_{J_2}}\right) + O_P(q^{-1}(n_{J_2})), \quad (9)$$

where Γ_{J_1,J_2}, denoted the asymptotic covariance matrix of $S_{J_1,J_2}(1)$ and,

$$\widehat{\Gamma}^{-1/2}_{J_1,J_2}\left(S_{J_1,J_2}(t) - \mathbb{E}\left[S_{J_1,J_2}(t)\right]\right) \xrightarrow{\mathcal{L}} B(t), \quad (10)$$

where the weak convergence holds in the space $D^{J_2-J_1+1}[0,1]$ *of cádl'ag functions defined on* $[0,1]$ *and valued in* $\mathbb{R}^{J_2-J_1+1}$, *and* $B(t) = (B_{J_1}(t), \ldots, B_{J_2}(t))$ *are independent* Brownian motions. *As classical in convergence results for cádl'ag functions,* $D^{J_2-J_1+1}[0,1]$ *is equipped with the Skorokhod metric, see [10, Chapter 3].*

The proof of Theorem is omitted here for reasons of space.

The test is based on the statistics $T_{J_1,J_2} : [0,1] \to \mathbb{R}_+$ defined by

$$T_{J_1,J_2}(t) \stackrel{\text{def}}{=} (S_{J_1,J_2}(t) - tS_{J_1,J_2}(1))^T \, \widehat{\Gamma}^{-1}_{J_1,J_2} \\ (S_{J_1,J_2}(t) - tS_{J_1,J_2}(1)) , \quad t \in [0,1] \; . \quad (11)$$

As a consequence of (10), $T_{J_1,J_2}(t)$ also converges weakly in the Skorokhod space $D([0,1])$, as $n_{J_2} \to \infty$,

$$T_{J_1,J_2}(t) \xrightarrow{\mathcal{L}} \sum_{\ell=J_1}^{J_2} \left(B^0_\ell(t)\right)^2 \quad (12)$$

where $\{B^0_j(t), t \geq 0\}_{j=J_1,\ldots,J_2}$ are independent Brownian bridges. Using continuous mapping theorem (see [10, Theorem 2.7]), we can then define a Cramer Von-Mises test statistic as :

$$\mathrm{CVM}(J_1, J_2) \stackrel{\text{def}}{=} \int_0^1 T_{J_1,J_2}(t) \mathrm{d}t \; , \quad (13)$$

which converges to $C(J_2 - J_1 + 1)$ where for any integer d,

$$C(d) \stackrel{\text{def}}{=} \int_0^1 \sum_{\ell=1}^{d} \left[B^0_\ell(t)\right]^2 \mathrm{d}t \; . \quad (14)$$

The test rejects the null hypothesis when $\mathrm{CVM}(J_1, J_2) \geq c(J_2 - J_1 + 1, \alpha)$, where $c(d, \alpha)$ is the $1 - \alpha$th quantile of the distribution of $C(d)$. The distribution of the random variable $C(d)$ has been derived by [11] (see also [12] for more recent references).

4. APPLICATIONS

4.1. Application on simulated data

In this section, we report the results of a limited Monte-Carlo experiment to assess the finite sample property of the test procedure. Recall that the test rejects the null if $\mathrm{CVM}(J_1, J_2)$ defined in (13) exceeds the $(1-\alpha)$-th quantile of the distributions $C(J_2 - J_1 + 1)$ specified in (14). To study the influence on the test procedure of the strength of the dependency, we consider different classes of Gaussian processes, including white noise, autoregressive moving average (ARMA) processes as well as fractionally integrated ARMA (ARFIMA(p, d, q)) processes which are known to be long range dependent. In all the simulations we set the lowest scale to $J_1 = 1$ and vary the coarsest scale $J_2 = J$. We used a wide range of values of sample size n, of the number of scales J and of the parameters of the ARMA and ARFIMA processes but, to conserve space, we present the results only for $n = 1024$, $J = 3, 4, 5$ and four different models: a white noise (WN), an AR(1) process with parameter $\phi = 0.9$ and two ARFIMA(1,d,0) processes with memory parameter $d = 0.3$, $d = 0.4$ respectively and $\phi = 0.9$. In our simulations, we have used the Newey-West estimate of the bandwidth $q(n_j)$ for the covariance estimator (as implemented in the R-package *sandwich*). In the AR and the ARFIMA cases, the test rejects the null much too often when the number of scales is large compared to the sample size (the difficult problem being in that case to estimate the covariance matrix of the test). However, when $J_2 = 3$, the target rejection rate is obtained (cf: Table 1).

	WN	AR	ARFIMA	
			$d = 0.3$	$d = 0.4$
$J = 3$	0.050	0.045	0.033	0.053
$J = 4$	0.041	0.200	0.160	0.13
$J = 5$	0.086	0.340	0.400	0.556

Table 1. Empirical level of CVM for some gaussian processes.

4.2. Application on real data

We present an application of the proposed procedure in the context of electric power consumption monitoring in a residential building. Motivations to perform such a monitoring are various. As mentioned in [13], small and large utility companies are interested in many aspects of such a monitoring. In the long-term understanding the influence of weather or external conditions on the consumption behaviour of consumers would be helpful to help forecasting and coping with load peaks. It would also help defining more accurate energy prices. In the shorter term the advent of sustainable electricity production methods have raised complex control issues for integrating these systems into existing ones. They usually entail a continuous monitoring of power consumption in individual households to detect energy load changes and adapt to them.

To illustrate our change-point detection approach in this context, we consider the "Individual household electric power consumption" dataset freely available from the widely used UCI repository [14]. Data consist of electric power consumption measurements in one residential building. Notably the global active power as well as three sub-metering values corresponding to particular rooms were sampled every minute from December 2006 to November 2010. In order to focus on comparable daily energy consumptions, we consider a 24-hour integrated version of the global active power, from which sub-metering values are subtracted (event-based consumption due to washing machine, tumble-drier, . . .). Data imputation of missing values has been made using a k-nearest neighbour approach [15]. The time series has also been seasonally adjusted to remove the 12-month cycle. The approach describes above assumes at most

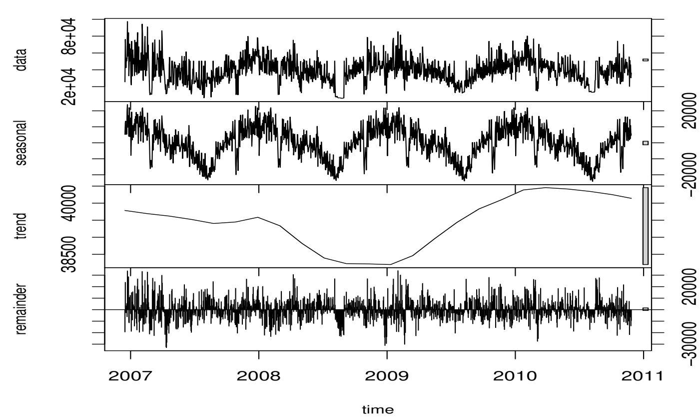

Fig. 1. Seasonal decomposition of the daily power electric consumption from 16-Dec-2006 to 30-Nov-2010. From top to bottom, 1) daily energy consumption in the household (24-hour integrated version of the raw data), 2) the seasonal component based on the 12-month natural cycle obtained by averaging over the four available years, 3) the background trend and 4) the remainder, which corresponds to the noise. Units are in watts per hour.

one possible change point in the time series. If one is interested in *multiple change points detection*, the Iterated Cumulative Sums of Squares (ICSS) Algorithm proposed by [9] is easily adapted to our context.

Let us denote by $\widetilde{T}^{k_1:k_2}_{J_1,J_2}(k)$ the statistic $T_{J_1,J_2}(t)$ based on the observations $Y_{J_1,J_2}[k_1 : k_2) = \{Y_{J_1,J_2}(k),\ k_1 \leq k < k_2\}$ in absolute time $k = k_1, \ldots, k_2$. In other words, we set, for $k_1 \leq k \leq k_2$,

$$\widetilde{T}^{k_1:k_2}_{J_1,J_2}(k) = \left(\widetilde{S}_{J_1,J_2}(k) - \frac{k-k_1}{k_2-k_1}\widetilde{S}_{J_1,J_2}(k_2)\right)^T \left(\widehat{\Gamma}^{k_1:k_2}_{J_1,J_2}\right)^{-1}\left(\widetilde{S}_{J_1,J_2}(k) - \frac{k-k_1}{k_2-k_1}\widetilde{S}_{J_1,J_2}(k_2)\right) ,$$

where $\widehat{\Gamma}^{k_1:k_2}_{J_1,J_2}$ is the estimator $\widehat{\Gamma}^{\mathrm{L}}_{J_1,J_2}$ based on $Y_{J_1,J_2}[k_1 : k_2)$ and

$$\widetilde{S}_{J_1,J_2}(k) = \frac{1}{\sqrt{k_2-k_1}} \sum_{i=k_1}^{k} Y_{J_1,J_2}[i] .$$

Finally we denote by $\mathrm{CVM}^{k_1:k_2}_{J_1,J_2}$ the corresponding test statistic,

$$\mathrm{CVM}^{k_1:k_2}_{J_1,J_2} = \frac{1}{k_2-k_1} \sum_{k=k_1}^{k_2-1} \widetilde{T}^{k_1:k_2}_{J_1,J_2}(k) .$$

The *multiple change points detections algorithm* of [9, Section 3] can be readily applied to $\widetilde{T}^{k_1:k_2}_{J_1,J_2}(k)$ with the appropriate quantiles. To this end we denote by $\mathbf{c}$ the asymptotic quantile associated to probability 0.95. We quote this algorithm hereafter for convenience.

Step 0 Let $k_1 = 1$

Step 1 Calculate $\widetilde{T}^{k_1:n_{J_2}}_{J_1,J_2}(k)$ for $k_1 \leq k < n_{J_2}$. Let $k^*_{k_1:n_{J_2}}$ be the maximizing point of this sequence. If $\mathrm{CVM}^{k_1:n_{J_2}}_{J_1,J_2} > \mathbf{c}$, consider that there is change point at $k^*_{k_1:n_{J_2}}$ and proceed to Step 2a. Otherwise, there is no evidence of variance changes in the series. The algorithm stops.

Step 2a Let $k_2 = k^*_{k_1:n_{J_2}}$. Evaluate $\widetilde{T}^{k_1:k_2}_{J_1,J_2}(k)$ for $k_1 \leq k < k_2$. If $\mathrm{CVM}^{k_1:k_2}_{J_1,J_2} > \mathbf{c}$, then we have a new point of change and should repeat Step 2a until $\mathrm{CVM}^{k_1:k_2}_{J_1,J_2} \leq \mathbf{c}$. When this occurs, we can say that there is no evidence change in $k = k_1, \ldots, k_2$ and, therefore, the first change point is then $k_{\mathrm{first}} = k_2$.

Step 2b Now do a similar search starting from the first change point found in step 1 toward the end of the series. Define a new value for k_1: let $k_1 = k^*_{k_1:n_{J_2}}$. Evaluate $\widetilde{T}^{k_1:k_2}_{J_1,J_2}(k)$ for $k_1 \leq k < k_2$ and repeat step 2b until $\mathrm{CVM}^{k_1:k_2}_{J_1,J_2} \leq \mathbf{c}$. Then let $k_{\mathrm{last}} = k_1 - 1$

Step 2c If $k_{\mathrm{first}} = k_{\mathrm{last}}$, there is just one change point. The algorithm stops there. If $k_{\mathrm{first}} < k_{\mathrm{last}}$, keep both values as possible change points and repeat Step 1 and Step 2 on the middle part of the series; that is, $k_1 = k_{\mathrm{first}}$ and $k_2 = k_{\mathrm{last}}$. Each time that Steps 2a and 2b are repeated, the result can be one or two more points. Call N the number of change points found so far.

Step 3 If there are two or more possible change points, make sure they are in increasing order. Let *cp* be the vector of all the possible change points found so far. Define the two extreme values $cp_0 = 0$ and $cp_{N+1} = n_{J_2}$. Check each possible change point by calculating $\widetilde{T}^{cp_{j-1}:cp_{j+1}}_{J_1,J_2}$ $j = 1, \ldots, N$. If $\mathrm{CVM}^{cp_{j-1}:cp_{j+1}}_{J_1,J_2} > \mathbf{c}$, then cp_j will be replaced by the corresponding change point; otherwise eliminate it. Repeat Step 3 until the number of change points does not change and the points found in each new pass are "close" to those on the previous pass. We consider that if each cp_j has not increase nor decrease by more than one from the previous iteration, then the algorithm has converged.

Applied to $\widetilde{T}^{k_1:k_2}_{J_1,J_2}(k)$, the output of this algorithm is a (possibly empty) set of change points $cp_1, \ldots, cp_N$ corresponding to indices at scale J_2; hence the corresponding time indices are $cp_1 2^{J_2}, \ldots, cp_N 2^{J_2}$. We applied this method to detect multiple changes in the spectral content of the absolute remainder observations (after removal of trend and seasonal component). We set $J_1 = 1$ and $J_2 = 3$ in our application. The interpretation of

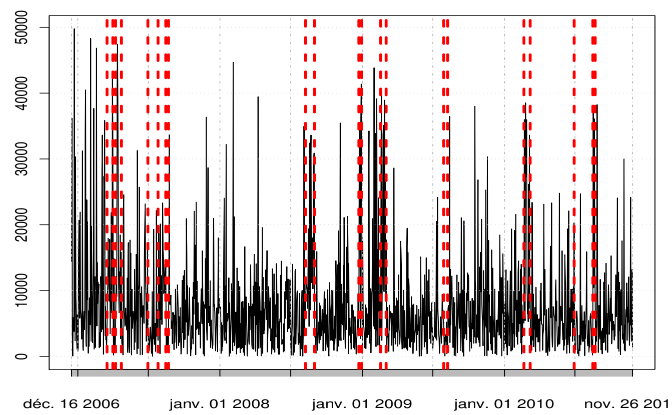

Fig. 2. Detection of change points in the absolute remainder time series. Red dotted lines correspond to change points

the obtained change points has been made on the raw series. It should be noted that two main kinds of change points are obtained. First a group of "summer" points clearly correspond to periods of vacations. This is confirmed by the fact that the electric power consumption clearly decreases below the 10-day moving average of the series of daily consumptions during 7 to fourteen consecutive days. This observation confirms the ability of the proposed algorithm to detect periods of unusual low power consumption. Second "winter" change points have been detected on the series. Nevertheless it should be noted that periods (defined as a few consecutive days) of high electrical load are main causes of detected change points, while isolated consumption peaks, involving only one day, are usually not detected. This last point clearly illustrates a normal behaviour of the algorithm. Change point detection is indeed simultaneously considered over different scales of the wavelet decomposition of the series. It is thus designed for identifying concurrent changes over time scales, which is clearly not the case when only one day is involved.

While the application demonstrates the ability of the algorithm to identify changes in the distribution of the series, further interpretation of the results would need complementary information about the behaviour of the inhabitants. This constitutes a research direction currently at stake for this particular application.

5. REFERENCES

[1] B. Whitcher, S. D. Byers, P. Guttorp, and Percival D., "Testing for homogeneity of variance in time series : Long memory, wavelets and the nile river," *Water Resources Research*, vol. 38, no. 5, 2002.

[2] R. N. Bhattacharya, Vijay K. Gupta, and Ed Waymire, "The hurst effect under trends," *Journal of Applied Probability*, vol. 20, no. 3, pp. 649–662, 1983.

[3] T. Mikosch and C. Stărică, "Changes of structure in financial time series and the Garch model," *REVSTAT*, vol. 2, no. 1, pp. 41–73, 2004.

[4] I. Berkes, L. Horvatz, Kokoszka P., and Qi-Man Shao, "On discriminating between long-range dependence and change in mean,," *The annals of statistics*, vol. 34, no. 3, pp. 1140–1165, 2006.

[5] C. M. Hurvich, G. Lang, and P. Soulier, "Estimation of long memory in the presence of a smooth nonparametric trend," *J. Amer. Statist. Assoc.*, vol. 100, no. 471, pp. 853–871, 2005.

[6] Jan Beran, *Statistics for long-memory processes*, vol. 61 of *Monographs on Statistics and Applied Probability*, Chapman and Hall, New York, 1994.

[7] B. E. Brodsky and B. S. Darkhovsky, *Non-parametric statistical diagnosis*, vol. 509 of *Mathematics and its Applications*, Kluwer Academic Publishers, Dordrecht, 2000, Problems and methods.

[8] E. Moulines, F. Roueff, and M. S. Taqqu, "On the spectral density of the wavelet coefficients of long memory time series with application to the log-regression estimation of the memory parameter," *J. Time Ser. Anal.*, vol. 28, no. 2, pp. 157–187, 2007.

[9] C. Inclan and G. C. Tiao, "Use of cumulative sums of squares for retrospective detection of changes of variance," *American Statistics*, vol. 89, no. 427, pp. 913–923, 1994.

[10] P. Billingsley, *Convergence of probability measures*, Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition, 1999, A Wiley-Interscience Publication.

[11] J. Kiefer, "K-sample analogues of the Kolmogorov-Smirnov and Cramér-V. Mises tests," *Ann. Math. Statist.*, vol. 30, pp. 420–447, 1959.

[12] P. Carmona, F. Petit, J. Pitman, and M. Yor, "On the laws of homogeneous functionals of the Brownian bridge," *Studia Sci. Math. Hungar.*, vol. 35, no. 3-4, pp. 445–455, 1999.

[13] Georges Hébrail, Bernard Hugueney, Yves Lechevallier, and Fabrice Rossi, "Exploratory analysis of functional data via clustering and optimal segmentation," *Neurocomputing*, vol. 73, no. 79, pp. 1125 – 1141, 2010, ¡ce:title¿Advances in Computational Intelligence and Learning¡/ce:title¿ ¡ce:subtitle¿17th European Symposium on Artificial Neural Networks 2009¡/ce:subtitle¿ ¡xocs:full-name¿17th European Symposium on Artificial Neural Networks 2009¡/xocs:full-name¿.

[14] A. Frank and A. Asuncion, "UCI machine learning repository," 2010.

[15] T. Hastie, R Tibshirani, and J. Friedman, *The Elements of Statistical Learning : Data Mining, Inference, and Prediction.*, Second Edition, Springer Series in Statistics, Frebruary 2009.