**Control and Cybernetics**

vol. **34** (2005) No. 4

# Wavelets for time series analysis – a survey and new results

by

**Jan Mielniczuk and Piotr Wojdyłło**[1]

Institute of Computer Science, Polish Academy of Sciences
Ordona 21, 01-237 Warsaw, Poland
e-mail: miel,pwoj@ipipan.waw.pl

**Abstract:** In the paper we review stochastic properties of wavelet coefficients for time series indexed by continuous or discrete time. The main emphasis is on decorrelation property and its implications for data analysis. Some new properties are developed as the rates of correlation decay for the wavelet coefficients in the case of long-range dependent processes such as the fractional Gaussian noise and the fractional autoregressive integrated moving average processes. It is proved that for such processes the within-scale covariance of the wavelet coefficients at lag $k$ is $\mathcal{O}(k^{2(H-N)-2})$, where $H$ is the Hurst exponent and $N$ is the number of vanishing moments of the wavelet employed. Some applications of decorrelation property are briefly discussed.

**Keywords:** decorrelation property, spectral density, time series, wavelets, fractional Gaussian noise (FGN), fractional autoregressive integrated moving average (FARIMA), Hurst exponent, long-range dependence

## 1. Introduction

Let $(X(t))_{t\in\mathbb{R}}$ be a real valued stochastic process such that $\mathbb{E}X(t) = 0$ and $\mathbb{E}X(t)^2 < \infty$ for any $t \in \mathbb{R}$. Throughout, the index $t$ will have connotation of actual time. We refer to such a process as the time series. In Section 4 we also discuss the case of the discrete uniform sampling when $t \in \mathbb{Z}$. Consider a function $\psi(\cdot) \in \mathcal{L}^2(\mathbb{R})$ such that $\int \psi(s)\, ds = 0$ and let

$$\psi_{j,k}(t) = 2^{-j/2}\psi(2^{-j}t - k) \qquad \text{for} \quad j,k \in \mathbb{Z}$$

be its rescaled and translated version. The function $\psi(\cdot)$ is called a wavelet when the family $\{\psi_{j,k}(\cdot)\}_{j,k\in\mathbb{Z}}$ forms an orthonormal basis in $\mathcal{L}^2(\mathbb{R})$. Note

[1] The second author is currently with Institute of Mathematics, Polish Academy of Sciences, Warsaw, Poland.

in particular that the $\mathcal{L}^2$-norm of $\psi$ denoted by $\| \psi \|_2$ equals 1. The name "wavelet" corresponds to the oscillating nature of $\psi(\cdot)$ expressed by its moment of order 0 equal to 0 and its compact support or quickly diminishing tails. Observe that if $\psi(\cdot)$ has a compact support $[a, b]$ then the support of $\psi_{j,k}(\cdot)$ is $[2^j(a+k), 2^j(b+k)]$. $2^j$ is called a scale of $\psi_{j,k}(\cdot)$ and $j$ its resolution or octave. Negative $j$s correspond to finer resolutions than nonnegative ones.

For a general introduction to wavelets with emphasis on applications to statistics we refer to Vidakovic (1999) and to Percival and Walden (2000) for a careful exposition of the discrete time series case. Abry et al. (2002) provide an excellent review of the subject devoted mainly to continuous time processes. We also refer the reader to Nason and von Sachs (1999) and Gençay et al. (2002). Here, our aim is to provide a self-contained and up-to-date exposition of the stochastic properties of the wavelet coefficients for both continuous and discrete time and to point out analogies between both cases. In particular, we discuss a decorrelation property of wavelet coefficients for stationary processes with special emphasis on the case of long-range dependence. Although this phenomenon is widely known and frequently used in applications, formal results concerning decorrelation exist only in few cases of specific processes including a fractional Brownian motion (Tewfik and Kim, 1992; Flandrin, 1992) and some short-range dependent processes as in Dijkerman and Mazumdar (1994). In Sections 3 and 4 we establish two new results on the rates of decorrelation for the two most popular models of long-range dependent processes: fractional Gaussian noise (FGN) (Theorem 3.1) and fractionally differenced ARMA processes (FARIMA) (Theorem 4.1). Namely, we give a formal proof of the fact that in both cases when a wavelet with $N$ vanishing moments is used the within-scale covariance of wavelet coefficients at lag $k$ is $\mathcal{O}(k^{2(H-N)-2})$. In Proposition 3.8 we state the assumptions under which such results can be obtained without imposing a specific parametric structure on the process. In Section 2 we discuss four important examples of wavelets, for which decorrelation property may be established via the results proved in the paper. In Section 4 we study the case of discrete time series. In the applications section (Section 5) we shortly discuss how the effect of decorrelation is used to study the properties of wavelet-based estimators of the Hurst exponent of long-range dependence and to simulate long-range dependent processes.

Let $\phi(\cdot)$ be a scaling function pertaining to $\psi(\cdot)$ (see Vidakovic, 1999, Section 3.3) and $u_k$ the Fourier coefficients of $\phi_{1,0}(\cdot) := 2^{-1/2}\phi(\cdot/2)$ with respect to an orthonormal sequence $\{\phi_{0,k}(\cdot)\}_{k\in\mathbb{Z}} = \{\phi(\cdot - k)\}_{k\in\mathbb{Z}}$:

$$u_k = 2^{-1/2} \int \phi(t/2)\phi(t-k)\,dt, \qquad k \in \mathbb{Z},$$

$(v_k)_{k\in\mathbb{Z}}$ is an analogously defined sequence of the Fourier coefficients of $\psi_{1,0}(\cdot) = 2^{-1/2}\psi(\cdot/2)$:

$$v_k = 2^{-1/2} \int \psi(t/2)\phi(t-k)\,dt.$$

Unless specified otherwise, $\int$ stands for an integral over $\mathbb{R}$. We have $v_k = (-1)^k u_{L-1-k}$ for the so-called quadrature mirror filters, where $L$ is the length of the filter $(v_k)$ i.e. the minimal $l \in \mathbb{N}$ such that $v_j = 0$ for $j \notin \{0, 1, \ldots, L-1\}$.

In the paper we focus on stochastic properties of wavelet coefficients (details) based on a trace $(X(t))_{t\in\mathbb{R}}$:

$$d_{j,k} = \int X(t)\psi_{j,k}(t)\,dt \tag{1}$$

and respective approximation (scaling) coefficients

$$a_{j,k} = \int X(t)\phi_{j,k}(t)\,dt, \tag{2}$$

where $\phi_{j,k}(t) = 2^{-j/2}\phi(2^{-j}t - k)$. Mapping $X(\cdot) \to (d_{j,k})_{j,k\in\mathbb{Z}}$ is called a Discrete Wavelet Transform (DWT). Note that as the typical wavelet is centered around 0 and quickly decaying for large $t$, the wavelet coefficients $d_{j,k}$ can be viewed, due to $\int \psi = 0$, as the differences of the weighted averages of $X(t)$ at the scale $2^j$ in a vicinity of $2^j k$. On the other hand, the approximation coefficients correspond to aggregation of the process at the scale $2^j$. The shape of the trace $X(\cdot)$ should resemble that of $\psi_{j,k}(\cdot)$ in order for large values of $d_{j,k}$ to occur. Observe also that $d_{j,k} = D(2^j, 2^j k)$, where $D(a,\tau)$ is a Continuous Wavelet Transform (CWT) defined as

$$D(a,\tau) = \frac{1}{\sqrt{a}} \int X(t)\psi\Big(\frac{t-\tau}{a}\Big)\,dt.$$

A crucial property of those sets of coefficients is that for a given resolution $j$ both $(d_{j,\cdot})$ and $(a_{j,\cdot})$ can be recursively computed from approximation coefficients at finer resolutions. Namely (see, e.g. Vidakovic, 1999, Section 4.2),

$$d_{j,k} = \sum_{n\in\mathbb{Z}} v_{-n} a_{j-1,2k-n} = \sum_{n\in\mathbb{Z}} v_n^{\vee} a_{j-1,2k-n} = v_{\cdot}^{\vee} \star a_{j-1,\cdot}(2k), \tag{3}$$

where $v_n^{\vee} = v_{-n}$ denotes the sequence $v_n$ with reversed time and $\star$ denotes the convolution in $l^2(\mathbb{Z})$.

Analogously,

$$a_{j,k} = \sum_{n\in\mathbb{Z}} u_n^{\vee} a_{j-1,2k-n} = u_{\cdot}^{\vee} \star a_{j-1,\cdot}(2k). \tag{4}$$

Observe that the coefficients $d_{j,k}$ are obtained by two consecutive operations: first the sequence $(a_{j-1,\cdot})$ is filtered using the sequence $(v_{\cdot}^{\vee})$ (a filtering stage) and then all elements of the resulting sequence with odd indices are discarded (a decimation stage). The filters $(v_{-n})$ and $(u_{-n})$ preserve high and low frequencies of the process, respectively. The two equalities above justify the so called pyramidal algorithm (Burt and Adelson, 1983). Here, the approximation coefficients $a_{0,k}$ at resolution 0 are calculated directly from the sample path of the process $X(t)$ using equation (2). Then the coefficients $d_{j,k}$ and $a_{j,k}$ at a resolution $j \geq 1$ are calculated recursively from approximation coefficients $a_{j-1,k}$ at the finer resolution $j-1$. The coefficients $(a_{j-1,\cdot})$ can be reconstructed using the sequences $(a_{j,\cdot})$ and $(d_{j,\cdot})$ pertaining to a coarser scale. Thus, for any $J \in \mathbb{N}$ the initial approximation coefficients $(a_{0,\cdot})$ may be recovered from a family of detail sequences $\{(d_{j,\cdot}), j = 1, \ldots, J\}$ and the coarsest approximation coefficients $(a_{J,\cdot})$. Let $\mathcal{P}_0 X(t)$ be the projection of the sample path $X(t)$ on the closure of a subspace of $\mathcal{L}^2(\mathbb{R})$ spanned by $(\phi_{0,k}(\cdot))_{k \in \mathbb{Z}}$. It follows that sequences $(a_{J,\cdot})$ and $(d_{j,\cdot})_{j=1,\ldots,J}$ contain full information about $\mathcal{P}_0 X(t)$. Namely,

$$\mathcal{P}_0 X(t) = \sum_{k \in \mathbb{Z}} a_{J,k} \phi_{J,k}(t) + \sum_{j=1}^{J} \sum_{k \in \mathbb{Z}} d_{j,k} \psi_{j,k}(t) \tag{5}$$

in $\mathcal{L}^2(\mathbb{R})$. Let us note that if the decimation step is replaced by a uniform sampling in $k$ one arrives at non-decimated DWT (NDWT; see, e.g. Nason and Sachs, 1999, or Chapter 5 in Percival and Walden, 2000), also called maximal overlap DWT, which can be used in the problems studied here as well.

## 2. Examples

In this section and throughout the paper we denote by $U(\lambda) = |u(\lambda)|^2 = |\sum_{k \in \mathbb{Z}} u_k e^{ik\lambda}|^2$ a squared gain (power) function of filter u $= (u_k)$ and by $V(\lambda)$ the analogously defined function for filter v $= (v_k)$. The material contained in this section can be found e.g. in Vidakovic (1999), Section 3.4. For more detailed expositions we refer the reader to Daubechies (1992), Meyer (1994), and Wojtaszczyk (1997).

EXAMPLE 2.1 Let $\phi(x) = I_{[0,1]}(x)$ and $\psi(x) = I_{[0,1/2]}(x) - 1_{[1/2,1]}(x)$ with $I_A$ denoting indicator of a set $A$. Then it is easy to see that the system of functions $(\psi_{jk}(x))_{j,k \in \mathbb{Z}}$ is an orthonormal basis in $L^2(\mathbb{R})$. The obtained system is the well-known Haar basis. Note that $\phi$ satisfies the scaling equation

$$\phi_{1,0}(x) = \frac{1}{\sqrt{2}} \phi_{0,0}(x) + \frac{1}{\sqrt{2}} \phi_{0,1}(x)$$

and for $\psi$ we have

$$\psi_{1,0}(x) = \frac{1}{\sqrt{2}} \phi_{0,0}(x) - \frac{1}{\sqrt{2}} \phi_{0,1}(x).$$

Thus, the coefficients of the filter u are $u_0 = u_1 = \frac{1}{\sqrt{2}}$, and the squared gain function $U(\lambda) = 2\cos^2(\lambda/2)$. The coefficients of the filter v are $v_0 = \frac{1}{\sqrt{2}}$, $v_1 = -\frac{1}{\sqrt{2}}$, and the squared gain function $V(\lambda) = 2\sin^2(\lambda/2)$. The Haar wavelet has one vanishing moment: $\int_{-\infty}^{\infty} \psi(x)dx = 0$. It is compactly supported but discontinuous, and its Fourier transform being $\widehat{\psi}(\lambda) = -\frac{(e^{i\lambda/2}-1)^2}{2\pi i\lambda}$ decays slowly. The Haar scaling function enjoys the rare property of being symmetric in time domain. The use of the filter v in this case is equivalent to the differencing procedure.

Example 2.2 Let $\phi(x) = sinc(x) = \frac{sin(\pi x)}{\pi x}$. Since $\widehat{\phi}(\lambda) = I_{[-\pi,-\pi]}(\lambda)$, the system $(\phi(x-k))_{k\in\mathbb{Z}}$ is orthonormal. The function $\phi$ satisfies the scaling equation $\phi_{1,0}(x) = \sum_{k\in\mathbb{Z}} u_k\phi_{0,k}(x)$, where the coefficients of the filter u are given by $u_k = \frac{1}{\sqrt{2}} sinc(\frac{\pi}{2k})$. The pertaining wavelet $\psi(x) = sinc\,\pi(x-1/2) - 2sinc\,2\pi(x-1/2)$ is called the Shannon wavelet. Its Fourier transform has a simple form $\widehat{\psi}(\lambda) = -e^{-i\lambda/2} I_{[-2\pi,-\pi]\cup[\pi,2\pi]}(\lambda)$. Note that the Shannon filter u is the ideal low-pass and v an ideal high-pass filter. The squared gain functions of the Shannon filters satisfy

$$U(\lambda) = 2I_{[-\pi/2,-\pi/2]}(\lambda), \qquad V(\lambda) = 2I_{[-\pi,-\pi/2]\cup[\pi/2,\pi]}(\lambda).$$

Example 2.3 Let $\nu$ be a smooth function satisfying

$$\begin{aligned} &\nu(x) + \nu(1-x) = 1 \\ &\nu(x) = 0, \qquad x \leq 0 \\ &\nu(x) = 1, \qquad x \geq 1. \end{aligned}$$

Then define $\Phi$ as

$$\Phi(\lambda) = \begin{cases} 1 & |\lambda| \leq \frac{2\pi}{3} \\ \cos(\frac{\pi}{2}\nu(\frac{3|\lambda|}{2\pi} - 1)) & \frac{2\pi}{3} \leq |\lambda| \leq \frac{4\pi}{3} \\ 0 & |\lambda| \geq \frac{4\pi}{3} \end{cases} .$$

Denote the inverse Fourier transform of $\Phi$ by $\phi$. The system of translates $(\phi(x-k))_{k\in\mathbb{Z}}$ is orthonormal, which follows from the properties of $\nu$. Since $\Phi$ is supported on a compact interval, $\phi$ is infinitely differentiable but has an infinite support. The Fourier transform of the pertaining wavelet $\psi$ is

$$\widehat{\psi}(\lambda) = -e^{-i\lambda/2}[\Phi(\lambda - 2\pi) + \Phi(\lambda + 2\pi)]\Phi(\lambda/2).$$

The function $\widehat{\psi}$ is supported on $[-\frac{8\pi}{3}, \frac{8\pi}{3}]$. This construction was introduced by Y. Meyer in 1988, and the pertaining family of wavelets indexed by function $\nu$ bears his name.

EXAMPLE 2.4 Assume that the squared gain function of a filter u is of the following form

$$U(\lambda) = 2\left(1 - c\int_0^\lambda \sin^{2N-1} x\, dx\right)$$

and the transfer function

$$u(\lambda) = \sqrt{2}\left(\frac{1+e^{i\lambda}}{2}\right)^N \cdot L(\lambda),$$

where $u(\lambda) = \sum_{k\in\mathbb{Z}} u_k e^{ik\lambda}$ and the constant $c$ is such that $U(\pi) = 0$.

If the trigonometric polynomial $L$ is suitably chosen, the function $\phi^N$ satisfying the scaling equation is compactly supported on $[-N, N-1]$ and belongs at least to the class $C^{N/5}(\mathbb{R})$. This construction is due to I. Daubechies. It was the first proposal of a compactly supported differentiable wavelet. The wavelets of this family shall be referred to in the paper as $D(N)$. The squared gain function of the Daubechies filter v equals

$$V(\lambda) = 2c\int_0^\lambda \sin^{2N-1} x dx\,. \tag{6}$$

Let us note that the wavelet $\psi^N$ has $N$ vanishing moments of order $0, 1, \ldots N-1$ (see, e.g. Gençay et al., 2002, p. 114, see also remark following Theorem 4.2).

## 3. Stochastic properties of the wavelet coefficients

Observe that since $d_{j,k}$ and $a_{j,k}$ depend on underlying process $X(t)$ they are random and therefore it is of interest to study their stochastic properties. We begin with a property stating that stationarity of $X(t)$ is inherited by the wavelet coefficients at each resolution level.

PROPOSITION 3.1 *Let $(X(t))_{t\in\mathbb{R}}$ be a strongly stationary time series. Then $(d_{j,k})_{k\in\mathbb{Z}}$ is strongly stationary sequence for each $j \in \mathbb{Z}$.*

*Proof.* We will prove that $(d_{j,k_1+h}, d_{j,k_2+h}) \stackrel{\mathcal{D}}{=} (d_{j,k_1}, d_{j,k_2})$, where $\stackrel{\mathcal{D}}{=}$ means equality of distributions. The extension to $m$-dimensional distributions for $m > 2$ is straightforward. We have

$$(d_{j,k_1+h}, d_{j,k_2+h}) =$$
$$2^{-j/2}\left(\int X(t_1)\psi(2^{-j}t_1 - k_1 - h)\, dt_1, \int X(t_2)\psi(2^{-j}t_2 - k_2 - h)\, dt_2\right)$$

$$= 2^{-j/2}\left(\int X(t_1+2^jh)\psi(2^{-j}t_1-k_1)\,dt_1, \int X(t_2+2^jh)\psi(2^{-j}t_1-k_2)\,dt_2\right)$$
$$\stackrel{\mathcal{D}}{=} 2^{-j/2}\left(\int X(t_1)\psi(2^{-j}t_2-k_1)\,dt_1, \int X(t_2)\psi(2^{-j}t_2-k_2)\,dt_2\right)$$
$$= (d_{j,k_1}, d_{j,k_2}),$$

where the penultimate equality follows from the stationarity of $X(t)$. The same property holds when $X(t)$ has stationary increments i.e. becomes stationary after differencing with a step $h$ for any $h \in \mathbb{N}$. In particular, this is the case of the fractional Brownian motion described in Example 3.1 of Section 3. ■

Proposition 3.2 *Let* $(X(t))_{t\in\mathbb{R}}$ *be time series such that* $Y_h(t) := X(t+h) - X(t)$ *is strongly stationary for any* $h \in \mathbb{N}$*. Then for any* $j \in \mathbb{Z}$ *the coefficients* $(d_{j,k})_{k\in\mathbb{Z}}$ *form a strongly stationary sequence.*

The last proposition is proved analogously to Proposition 3.1 by using multivariate extension of the equalities

$$\int X(t+h)\psi(t)\,dt = \int (X(t+h) - X(h))\psi(t)\,dt \stackrel{\mathcal{D}}{=}$$
$$\int (X(t) - X(0))\psi(t)\,dt = \int X(t)\psi(t)\,dt,$$

where the first and the last equality follows from $\int \psi(t)\,dt = 0$ and the second one from stationarity of increments. Observe that Proposition 3.1, but not necessarily Proposition 3.2, is true for the approximation coefficients $a_{j,k}$.
From the very definition of wavelet and approximation coefficients it follows that for a zero-mean process $X(t)$ we have $\mathbb{E}a_{jk} = \mathbb{E}d_{j,k} = 0$. Consider now a zero mean weakly stationary process $X(t)$ with a covariance function $r(t) := \mathbb{E}(X(t+s)-\mathbb{E}X(t+s))(X(s)-\mathbb{E}X(s)) = \mathbb{E}(X(t+s)X(s))$. Assume additionally that a spectral distribution of $X(t)$ is absolutely continuous i.e. there exists $f \in \mathcal{L}^1(\mathbb{R})$ such that

$$r(t) = \int_{\mathbb{R}} e^{it\lambda} f(\lambda)\,d\lambda$$

for $t \in \mathbb{R}$. The function $f$ is called the spectral density of $X(t)$ and is defined uniquely up to a set of Lebesgue measure 0. Thus, denoting by $\hat{g}(t) = \int e^{it\lambda} g(\lambda)\,d\lambda$ the Fourier transform of a function $g$, we have that $r = \hat{f}$. We stress that for a continuous time process the spectral distribution might be supported on the whole real line in contrast to discrete time series when its support is confined to $[-\pi, \pi)$.

It follows from the definition of the wavelet coefficients that their covariance can be written as

$$\begin{aligned}\mathrm{Cov}(d_{j,k}, d_{j',k'}) &= \mathbb{E}(d_{j,k}d_{j',k'}) = \int \mathbb{E}(X(t)X(s))\psi_{j,k}(t)\psi_{j',k'}(s)\,dt\,ds \\ &= \int r(t-s)\psi_{j,k}(t)\psi_{j',k'}(s)\,dt\,ds.\end{aligned}$$

The next proposition provides a convenient representation of the covariance in terms of the spectral density function of the process.

PROPOSITION 3.3 *Assume that the spectral density of* $(X(t))_{t\in\mathbb{R}}$ *exists. Then*

$$\mathrm{Cov}(d_{j,k}, d_{j',k'}) = 2^{(j+j')/2}\int f(\lambda)\hat{\psi}(2^j\lambda)\hat{\psi}^*(2^{j'}\lambda)e^{i\lambda(2^jk-2^{j'}k')}\,d\lambda, \tag{7}$$

*where* $^*$ *denotes complex conjugation.*

This is stated, for instance, as equation (13.8) in Walter (1994).

*Proof.* The proposition follows from the above expression for covariance and definition of a spectral density by changing the order of integration

$$\begin{aligned}&\mathrm{Cov}(d_{j,k}, d_{j',k'}) = \int r(u)\psi_{j,k}(s+u)\psi_{j',k'}(s)\,du\,ds \\ &= \int e^{i\lambda u}f(\lambda)\psi_{j,k}(s+u)\psi_{j',k'}(s)\,d\lambda\,du\,ds \\ &= \int f(\lambda)\int e^{i\lambda(s+u)}\psi_{j,k}(s+u)\,du\,\psi_{j',k'}(s)e^{-i\lambda s}\,ds\,d\lambda \\ &= \int f(\lambda)\hat{\psi}_{j,k}(\lambda)\hat{\psi}^*_{j',k'}(\lambda)\,d\lambda\end{aligned}$$

after noting that $\hat{\psi}_{j,k}(\lambda) = 2^{j/2}\hat{\psi}(2^j\lambda)e^{i2^jk\lambda}$. ■

Consider first the behavior of second order moments of $d_{j,k}$ when $(X(t))_{t\in\mathbb{R}}$ is weakly dependent.

PROPOSITION 3.4 *Assume that the spectral density* $f$ *of* $(X(t))_{t\in\mathbb{R}}$ *exists, is bounded on* $\mathbb{R}$ *and continuous at 0. Then* $\mathbb{E}d^2_{j,k} \to 2\pi f(0)$ *when* $j\to\infty$ *for any* $k\in\mathbb{Z}$.

*Proof.* Let $f_j(\lambda) := f(\lambda/2^j)|\hat{\psi}(\lambda)|^2$. The proof follows from (7), $\mathbb{E}d_{j,k} = 0$ and the decomposition

$$\begin{aligned}\mathbb{E}d^2_{j,k} &= \int f(\lambda/2^j)|\hat{\psi}(\lambda)|^2\,d\lambda = \int f_j(\lambda)\,d\lambda \\ &= \int_{|\lambda|\le j} f_j(\lambda)\,d\lambda + \int_{|\lambda|>j} f_j(\lambda)\,d\lambda.\end{aligned} \tag{8}$$

As the spectral density is bounded and $\hat{\psi} \in \mathcal{L}^2(\mathbb{R})$, the second integral tends to 0 when $j \to \infty$ whereas the first is asymptotically equivalent to $2\pi f(0)$ in view of Plancherel equality $\| \hat{\psi} \|_2^2 = 2\pi \| \psi \|_2^2 = 2\pi$ and continuity of $f$ at 0. ■

The proposition also holds when spectral density has a limit at 0 with $f(0)$ replaced with the value of the limit.

Remark 3.1 Observe that the second moment of $d_{j,k}$ for the Haar wavelet $\psi$ corresponds to a well-known statistical quantity. Namely, in this case

$$d_{j,k} = 2^{-j/2}\Big(\int_{2^j k}^{2^j k+2^{j-1}} X(s)\,ds - \int_{2^j k+2^{j-1}}^{2^j (k+1)} X(s)\,ds\Big),$$

and thus $\mathbb{E}d_{j,k}^2 = 2^{j-1}\sigma_X^2(2^{j-1})$, where for a stationary $X(t)$

$$\sigma_X^2(\tau) = 2^{-1}\,\mathbb{E}\Big(\frac{1}{\tau}\int_0^{\tau} X(s)\,ds - \frac{1}{\tau}\int_{\tau}^{2\tau} X(s)\,ds\Big)^2$$

is the Allan's variance (see, e.g. Percival and Walden, 2000, Section 8.6) measuring the variability of adjacent averages of size $\tau$ of the process. At the same time for the Haar wavelet $a_{jk} = \int_{2^j k}^{2^j (k+1)} X(s)\,ds$ and its variance corresponds to the marginal variance of an aggregated process at the scale $2^j$.

We deal now with covariance of $d_{j,k}$ and $d_{j,k'}$ for a fixed $j$. From Proposition 3.3 we get

$$\mathrm{Cov}(d_{j,k}, d_{j,k'}) = \int f(\lambda/2^j)|\hat{\psi}(\lambda)|^2 e^{i\lambda(k-k')}\,d\lambda = \hat{f}_j(k-k').$$

Thus in view of the Lebesgue lemma we have

Proposition 3.5 *If* $f_j \in \mathcal{L}^1(\mathbb{R})$ *then* $\mathrm{Cov}(d_{j,k}, d_{j,k'}) \to 0$ *when* $|k-k'| \to \infty$.

In particular, $f_j \in \mathcal{L}^1(\mathbb{R})$ when $f$ is bounded or $\psi \in \mathcal{L}^1(\mathbb{R})$, which implies that its Fourier transform is bounded. By imposing stronger conditions on $f_j$ the decay rates of the covariance of $(d_{j,\cdot})$ are obtained.

Proposition 3.6 *(a) If* $f_j$ *is* $p$ *times differentiable,* $f_j^{(p)} \in \mathcal{L}^1(\mathbb{R})$ *and* $f_j^{(s)}(\lambda) \to 0$ *when* $\lambda \to \infty$ *for* $s = 0, 1, \ldots, p-1$, *then* $|\mathrm{Cov}(d_{j,k}, d_{j,k'})| = o(|k-k'|^{-p})$.
*(b) If* $\hat{\psi}(\cdot)$ *is compactly supported and* $f_j \in \mathcal{C}^p(\mathbb{R})$ *then (a) holds.*

Part (b) is stated in Walter (1994) in Proposition 13.1(iii) for the Meyer type wavelets.
*Proof.* Integration by parts $p$ times yields

$$\mathrm{Cov}(d_{j,k}, d_{j',k'}) = \frac{(-1)^p}{(i(k-k'))^p}\int f_j^{(p)}(\lambda)e^{i\lambda(k-k')}\,d\lambda = o(|k-k'|^{-p})$$

by the Lebesgue lemma and noting that the boundary terms disappear due to $f_j^{(s)}(\lambda) \to 0$ for $s = 0, 1, \ldots, p-1$ when $\lambda \to \infty$. By the same token we get the proof of (b). ■

REMARK 3.2 Note that if $r$ and $\psi$ are such that integrals $\int |r(t)||t|^p\, dt$ and $\int |\psi(t)||t|^p\, dt$ are finite then $f_j \in \mathcal{C}^p(\mathbb{R})$ and $f_j^{(k)}(\lambda) \to 0$ for $k = 0, 1, \ldots, p$. This follows from the observation that by the inversion formula for the Fourier transform $f \in \mathcal{C}^p(\mathbb{R})$ in view of $\int |r(t)||t|^p\, dt < \infty$. Moreover, its derivatives tend to 0 for $\lambda \to \infty$ in view of the Lebesgue lemma. The same properties hold for $|\hat{\psi}|^2$. Thus, in this case only the integrability of $f_j^{(p)}$ needs to be checked in order to satisfy the assumptions of Proposition 3.6(a).

Observe that the wavelet coefficients $d_{j,k}$ and $d_{j',k'}$ at possibly different levels $j$ and $j'$ become asymptotically uncorrelated when $|2^j k - 2^{j'} k'| \to \infty$ and analogous conditions to those imposed in Proposition 3.6 are assumed for a function $f_{j,j'}(\lambda) = f(\lambda)\hat{\psi}(2^j\lambda)\hat{\psi}^*(2^{j'}\lambda)$. Moreover, if the support of $\hat{\psi}$ is bounded and does not contain a neighborhood of 0 (as in the case of the Shannon wavelets) supports of $\hat{\psi}(2^j\cdot)$ and $\hat{\psi}(2^{j'}\cdot)$ become disjoint for sufficiently large $|j - j'|$ and in this case $\mathrm{Cov}(d_{j,k}, d_{j',k'}) = 0$ for arbitrary $k, k' \in \mathbb{Z}$.

Consider now the case when the spectral density $f \in \mathcal{L}^1(\mathbb{R})$ has a pole at 0, more specifically

$$f(\lambda) \sim c_f|\lambda|^{-\gamma} \qquad \text{for} \quad \lambda \to 0, \tag{9}$$

where $0 < \gamma < 1, c_f > 0$ and $\sim$ denotes asymptotic equivalence i.e. $f(\lambda)/c_f\lambda^{-\gamma} \to 1$ when $\lambda \to 0$. It follows from (9) that $\int |r(t)|\, dt = \infty$ (compare discussion in Section 4.1); this case of slowly decaying correlations is often described as long-range dependence or long-memory. The occurrence of the pole of $f$ at 0 explains the often used name $1/f$-type processes. Beran (1994) is a nice introduction to statistical problems for discrete long-memory processes (defined also by (9)).

EXAMPLE 3.1 Let $(Y(t))_{t\in\mathbb{R}}$ be the fractional Brownian motion (FBM) with the Hurst coefficient $1 > H \geq 0$ i.e. Gaussian process with stationary increments which is $H$-self-similar i.e. such that for each $a > 0$ $(Y(at))_{t\in\mathbb{R}} \overset{\mathcal{D}}{=} (a^H Y(t))_{t\in\mathbb{R}}$. Consider its first order difference $(X(t))_{t\in\mathbb{R}}$ defined as $X(t) := Y(t+1) - Y(t)$. Process $(X(t))_{t\in\mathbb{R}}$ is called the fractional Gaussian noise (FGN) and it is easy to see that it is stationary and its covariance

$$r_X(t) = \frac{\sigma^2}{2}\Big(|t+1|^{2H} + |t-1|^{2H} - 2|t|^{2H}\Big),$$

where $\sigma^2 = \mathbb{E}Y^2(1)$. When $t \to \infty$, $r_X(t) \sim H(2H-1)\sigma^2 t^{2H-2}$ for $H \neq 1/2$, thus the fractional Gaussian noise is long-range dependent when $1/2 < H < 1$.

Its spectral density is (Samorodnitsky, Taqqu, 1994, formula (7.2.25))

$$f(\lambda) = \frac{\sigma^2}{C(H)}\Big|\frac{e^{i\lambda}-1}{i\lambda}\Big|^2 |\lambda|^{-2(H-1/2)}$$

with $C(H)$ a positive constant depending only on $H$. Thus for $H > 1/2$ the spectral density of the FGN satisfies (9) with $\gamma = 2(H - 1/2)$.

EXAMPLE 3.2 Let $(Y(t))_{t\in\mathbb{R}}$ be a stationary Gaussian long-range dependent process such that $r_Y(t) \sim ct^{-\alpha}$ for $0 < \alpha < 1$ and $G \in \mathcal{L}^2(\mathbb{R}, \phi)$. Consider the subordinated Gaussian process $X(t) := G(Y(t))$. It turns out (see, e.g. Beran, 1994, Section 3.2) that the covariance function of $X(t)$ satisfies $r_X(t) \sim m!c^m t^{-m\alpha}$ provided $m\alpha < 1$, where $m$ is Hermite rank of $G$ defined as the smallest integer $n \geq 1$ such that $\mathbb{E}(G(Z)H_n(Z)) \neq 0$, where $H_n(\cdot)$ denotes $n^{th}$ Hermite polynomial and $Z$ is the standard normal random variable. Thus for $m$ such that $m\alpha < 1$ the subordinated Gaussian process $((X(t))_{t\in\mathbb{R}}$ is long-range dependent.

EXAMPLE 3.3 Let $(Y(t))_{t\in\mathbb{R}}$ and $(Z(t))_{t\in\mathbb{R}}$ be independent copies of a Gaussian stationary long-range dependent process such that $r_Y(t) = r_Z(t) \sim ct^{-\alpha/2}$ with $0 < \alpha < 1$. Then $X(t) = (Y^2(t) + Z^2(t))/2$ is a long-range dependent process with $r_X(t) = r_Y^2(t) \sim c^2 t^{-\alpha}$ having exponential marginals. Long-range dependence of the processes subordinated to the process $X(t)$ can be characterized analogously to the Gaussian case described in Example 3.2 with the role of the Hermite polynomials taken over by the Laguerre polynomials (see Gajek and Mielniczuk, 1999).

It turns out that the type of strong dependence defined in (9) implies that the wavelet coefficients behave differently than in the weak dependent case.

PROPOSITION 3.7 *Assume that (i)* $\psi \in \mathcal{L}^1(\mathbb{R})$*, (ii) condition (9) is satisfied and (iii)* $\sup_{|\lambda|\geq\varepsilon_n} |f(\lambda)| = \mathcal{O}(\varepsilon_n^{-\gamma})$ *for any* $\varepsilon_n \to 0$*. Then*

$$\mathbb{E}d_{j,k}^2 \sim 2^{j\gamma} c_f \int |\lambda|^{-\gamma} |\hat{\psi}(\lambda)|^2 \, d\lambda, \tag{10}$$

*when* $j \to \infty$*.*

*Proof.* Observe that the integral on the RHS of (10) exists since

$$\int |\lambda|^{-\gamma}|\hat{\psi}(\lambda)|^2 \, d\lambda \leq \sup |\hat{\psi}|^2 \int_{|\lambda|\leq 1} |\lambda|^{-\gamma} \, d\lambda + \int |\hat{\psi}(\lambda)|^2 \, d\lambda < \infty$$

as $\hat{\psi} \in \mathcal{L}^2(\mathbb{R})$ and is bounded in view of $\psi \in \mathcal{L}^1$. Consider decomposition (8) and observe that the second integral on its right hand side is bounded by $\mathcal{O}(2^j/j)^\gamma) \int_{|\lambda|>j} |\hat{\psi}(\lambda)|^2 \, d\lambda = o(2^{j\gamma})$. Since condition (iii) is equivalent to $\sup_{|\lambda|\le\varepsilon_n} |f(\lambda)/c_f|\lambda|^{-\gamma} - 1| \to 0$ for $\varepsilon_n \to 0$ the first integral can be written as

$$2^{j\gamma} \int_{|\lambda|\le j} \Big(c_f|\lambda|^{-\gamma} + o(|\lambda|^{-\gamma})\Big)|\hat{\psi}(\lambda)|^2 \, d\lambda.$$

Since $\int_{|\lambda|\le j} |\lambda|^{-\gamma}|\hat{\psi}(\lambda)|^2 \, d\lambda \to \int |\lambda|^{-\gamma}|\hat{\psi}(\lambda)|^2 \, d\lambda$ as $j \to \infty$ the last expression is equivalent to $2^{j\gamma} c_f \int |\lambda|^{-\gamma}|\hat{\psi}(\lambda)|^2 \, d\lambda$ in this case. ■

REMARK 3.3 Note that as the definition of long-range dependence specifies only the behaviour of the spectral density at 0, it is easy to construct LRD processes such that the assumption (iii) of Proposition 3.7 is *not* satisfied. Namely, this happens when $f$ has another singularity at frequency $\omega \neq 0$ as in the case $f(\lambda) \sim C|1 - e^{i\lambda}|^{-\gamma}|1 - 2\cos\omega e^{i\lambda} + e^{2i\lambda}|^{1/2-\eta}$, where $\eta > 1/2$, considered in Gray et al. (1989).

Let us note that if the conditions of Proposition 3.7 are satisfied with condition (i) replaced by $\phi \in \mathcal{L}^1(\mathbb{R})$ then we have in view of $\mathbb{E}a_{j,k} = 0$ that

$$\mathbb{E}a_{j,k}^2 \sim 2^{j\gamma} c_f \int |\lambda|^{-\gamma}|\hat{\phi}(\lambda)|^2 \, d\lambda, \tag{11}$$

when $j \to \infty$. Thus, in both cases of the wavelet and the approximation coefficients we have a power law dependence of their variance at the octave $j$ on the scale $2^j$. Observe that the equivalencies (10) and (11) yield a straightforward method of estimating $\gamma$. Namely, they imply that $\log_2 \mathbb{E}d_{j,k}^2$ and $\log_2 \mathbb{E}a_{j,k}^2$ regressed on $j$ should be approximately linear with a slope $\gamma$. We will return to this problem in Section 5.

REMARK 3.4 When $(X(t))_{t\in\mathbb{R}}$ is the fractional Brownian motion it is easy to see that

$$\begin{aligned}
\mathbb{E}d_{j,k}^2 &= \int \mathrm{Cov}(X(t), X(s))\psi_{j,k}(s)\psi_{j,k}(t) \, ds \, dt \\
&= \frac{\sigma^2}{2} \int (|t|^{2H} + |s|^{2H} - |t-s|^{2H})\psi_{j,k}(s)\psi_{j,k}(t) \, ds \, dt \\
&= \frac{-\sigma^2}{2} \int |t-s|^{2H}\psi_{j,k}(s)\psi_{j,k}(t) \, ds \, dt \\
&= \frac{-\sigma^2}{2} 2^{2j(H+1/2)} \int |t-s|^{2H}\psi(s)\psi(t) \, ds \, dt.
\end{aligned}$$

Thus, in this case a counterpart of (10) is exact scaling of variance for any $j, k \in \mathbb{Z}$. By a similar argument one can show that actually $d_{j,k} \stackrel{\mathcal{D}}{=} 2^{j(H+1/2)} d_{0,k}$.

Hence the translation invariance (stationarity) and scale invariance (self-similarity) is preserved by the wavelet transform for the FBM although the self-similarity coefficient changes in the latter case. Note that the factor of $j$ in the exponent in (10) is $\gamma = 2H - 1$ for the FGN whereas it is $2H + 1$ in the above formula for the FBM. The difference is due to the fact that the FGN is the first order difference of the FBM (see also remark following Proposition 3.9).

Consider now the question of how quickly the covariance of $d_{j,k}$ and $d_{j,k'}$ decays in long-range dependent case when $|k - k'| \to \infty$. The main issue here is that despite long-range dependence of $(X(t))$ the wavelet coefficients $(d_{j,\cdot})$ are actually weakly dependent if the wavelet $\psi(\cdot)$ is appropriately chosen. In order to appreciate why it is plausible, observe that if wavelet $\psi$ is such that its Fourier transform $\hat{\psi}$ vanishes in a neighborhood of 0 the function $f_j$ defined above does not have a pole at 0. One might then expect that the assumptions of Proposition 3.6 are satisfied for certain $p$ if $\hat{\psi}$ is sufficiently smooth. In the more general case consider the wavelet $\psi$ such that its first $N$ moments vanish i.e. $\int x^s \psi(x)\, dx = 0$ for $s = 0, 1, \ldots, N - 1$. $N = 1$ for the Haar wavelet and in general is equal to the order of Daubechies wavelet (see, e.g. Gençay et al., 2002, p. 114). If, moreover, $N$th absolute moment of $\psi$ exists then $\hat{\psi}(\lambda) = \mathcal{O}(|\lambda|^N)$ in a neighborhood of 0 and thus $f_j(\lambda)$ is $\mathcal{O}(|\lambda|^{2N-\gamma})$ there. Then it is reasonable to expect that in this case assumptions of Proposition 3.6 are satisfied with $p = 2N - 1$ yielding $\mathrm{Cov}(d_{j,k}, d_{j,k'}) = \mathcal{O}(|k - k'|^{1-2N})$.

To get some more insight into the decorrelation property of the wavelet coefficients consider also the following reasoning. Suppose that a localization property holds for a characteristic function of $f_j$, namely

$$\hat{f}_j(k) \propto \int_{|\lambda| \le \beta_k} f_j(\lambda) e^{i\lambda k}\, d\lambda \tag{12}$$

when $k \to \infty$ for some positive $\beta_k$ such that $\limsup_{k\to\infty} \beta_k < \infty$ and $a_k \propto b_k$ means that the limit of $a_k/b_k$ is finite and non-zero.

Note that (12) is trivially satisfied with $\beta_k = \beta$ if the support of $\hat{\psi}$ is contained in $[-\beta, \beta]$ or if $f$ is band-limited. Suppose that $f_j(\lambda) = |\lambda|^\delta I_{[-\pi,\pi]}$ for $0 < \delta < 1$. Then (12) holds true with $\beta_k = \pi/k$. Namely, integration by parts and change of variables yields $\int_0^\pi \lambda^\delta \cos k\lambda = -\delta k^{-\delta-1} \int_0^{k\pi} \lambda^{\delta-1} \sin \lambda$ and the last integral converges for $k \to \infty$. On the other hand, $\int_0^{\pi/k} \lambda^\delta \cos k\lambda = -\delta k^{-\delta-1} \int_0^\pi \lambda^{\delta-1} \sin \lambda$. Moreover, (12) holds true for $\delta = 1$ due to equality $\int_0^\pi \lambda \cos k\lambda = \int_0^{e_k\pi/k} \lambda \cos k\lambda$, where $e_k$ equals 2 or 1 depending on whether $k$ is even or odd, respectively. By the same token $\beta_k = \mathcal{O}(k^{-1/\delta})$ can be chosen for $\delta > 1$. Observe also that Theorem 4.1 implies that the localization property with $\beta_k \sim k^{-1}$ holds for $f_j(\cdot)$ pertaining to the FARIMA$(0, d, 0)$ process and the Daubechies wavelet.

In the following $C$ stands for a generic positive constant.

PROPOSITION 3.8 *Suppose that $N$ first moments of $\psi$ vanish and $\int |x|^N |\psi(x)|\, dx < \infty$, $f$ is bounded outside some neighborhood of 0 and conditions (9) and (12)* *are satisfied. Then*

$$\mathrm{Cov}(d_{j,k}, d_{j,k'}) \le C|k-k'|^{-2N-1+\gamma} \int_{|\lambda| \le \eta_{k-k'}} |\lambda|^{2N-\gamma}\, d\lambda$$

*with* $\eta_k = k\beta_k$. *In particular, for* $\beta_k = \mathcal{O}(k^{-\alpha})$, $\mathrm{Cov}(d_{j,k}, d_{j',k'}) = \mathcal{O}((|k-k'|)^{(-2N-1+\gamma)\alpha})$.

For $\alpha = 1$ the last equality corresponds to conjecture based on heuristic reasoning in Abry et al. (2003) stating that an exponent of the covariance decay is at least $-2N-1+\gamma = 2(H-N)-2$. It seems that some kind of property analogous to (12) is needed in order to prove such claim.

*Proof.* In view of (12) it is sufficient to bound

$$\int_{|\lambda| \le \eta_{k-k'}} \frac{1}{k-k'} f_j\Big(\frac{\lambda}{k-k'}\Big) e^{i\lambda}\, d\lambda$$

for $|k-k'| \to \infty$. Assumptions imply that $\hat{\psi}^{(i)}(0) = 0$ for $i = 0, 1, \ldots, N-1$ and $\hat{\psi}^{(N)}(\lambda)$ is continuous and hence $|\hat{\psi}(\lambda)| \le C|\lambda|^N$. Moreover, $|f(\lambda)| \le C|\lambda|^{-\gamma}$ on a bounded neighborhood of 0. Thus, the last integral is bounded by $C|k-k'|^{-2N-1+\gamma} \int_{|\lambda| \le \eta_{k-k'}} |\lambda|^{2N-\gamma}\, d\lambda$. ■

Consider now a decorrelation property for the special case when $X(t)$ is the fractional Gaussian noise. Then we have

Theorem 3.1 *Assume that* $(X(t))_{t\in\mathbb{R}}$ *is the fractional Gaussian noise with the Hurst coefficient* $H$ *and the wavelet* $\psi$ *has a compact support and has* $N$ *vanishing moments. Then* $\mathrm{Cov}(d_{j,k}, d_{j',k'}) = \mathcal{O}(|k-k'|^{2(H-N)-2})$ *when* $|k-k'| \to \infty$.

*Proof.* Observe that in view of the basic representation of the covariance of $d_{j,k}$ and $d_{j,k'}$, the form of the covariance for the FGN and its symmetry we have

$$\begin{aligned}
\mathrm{Cov}(d_{j,k}, d_{j,k'}) &= \int r(t-s)\psi_{j,k}(t)\psi_{j,k'}(s)\, dt\, ds \\
&= 2^j \int r(2^j(s-t+k'-k))\psi(t)\psi(s)\, dt\, ds \\
&= 2^j \int r(2^j(t+k'-k))\Lambda(t)\, dt \\
&= 2^{j-1}\sigma^2 \int (|a_{t,l}+1|^{2H} + |a_{t,l}-1|^{2H} - 2|a_{t,l}|^{2H})\Lambda(t)\, dt,
\end{aligned}$$

where $\Lambda(t) := \int \psi(s-t)\psi(s)\, ds$, $l := k'-k$ and $a_{t,l} := 2^j(t+l)$. It was proved by Tewfik and Kim (1992) that for any compactly supported function $g$ having $k$ vanishing moments

$$\int |a_{t,l}|^{2H} g(t)\, dt = \mathcal{O}(|l|^{2H-k}) \tag{13}$$

when $l \to \infty$. Moreover, from the assumptions it easily follows that for $\Lambda$ defined above we have that it is compactly supported and has $2N$ vanishing moments. Moreover, the formula for $\mathrm{Cov}(d_{j,k}, d_{j,k'})$ implies that

$$\mathrm{Cov}(d_{j,k}, d_{j,k'}) = 2^{j-1}\sigma^2 \int |a_{t,l}|^{2H} \bar{\Lambda}(t)\, dt,$$

where $\bar{\Lambda}(t) := \Lambda(t + 2^{-j}) + \Lambda(t - 2^{-j}) - 2\Lambda(t)$. Note that in view of properties of $\Lambda$, $\bar{\Lambda}$ is compactly supported and has $2N + 2$ vanishing moments since for any $n \in \mathbb{N}$ we have

$$\int t^n \bar{\Lambda}(t)\, dt = \int (t + 2^{-j})^n + (t - 2^{-j})^n - 2t^n)\Lambda(t)\, dt = \int w_{n-2}(t)\Lambda(t)\, dt,$$

where $w_{n-2}$ is a polynomial of degree $n - 2$. Thus the proposition follows from (13) for $g = \bar{\Lambda}$. ■

Remark 3.5 As indicated in the proof above for the FBM process we have $\mathrm{Cov}(d_{j,k}, d_{j,k'}) = \mathcal{O}(|k' - k|^{2(H-N)})$ (Tewfik and Kim, 1992; Flandrin, 1992). The improvement for the FGN in Theorem 3.1 is intuitively related to the fact that the power function pertaining to differencing filter equals $|1 - \exp(-i\lambda)|^2$ and behaves like $\lambda^2$ at 0. Observe also that it follows from the proposition that even for $N = 1$, satisfied e.g. by the Haar wavelet, the covariance function of $(d_{j,\cdot})$ is absolutely summable.

Theorem 3.2 *Assume that $\phi(\cdot)$ has a compact support and $(X(t))_{t\in\mathbb{R}}$ is a strongly stationary process such that $r_X(t) \sim c_r|t|^{-\alpha}$ for $0 < \alpha < 1$. Then*

$$\mathrm{Cov}(a_{j,k}, a_{j,k'}) \sim c_r 2^{j(1-\alpha)}|k' - k|^{-\alpha}\Big(\int \phi(t)\, dt\Big)^2$$

*when $|k' - k| \to \infty$.*

*Proof.* Reasoning as in the proof of Theorem 3.1 we get for $\tilde{\Lambda}(t) = \int \phi(s - t)\phi(s)\, ds$

$$\begin{aligned}\mathrm{Cov}(a_{j,k}, a_{j,k'}) &= 2^j \int r(2^j(s - t + k' - k))\phi(s)\phi(t)\, ds\, dt \\ &\sim c_r 2^j |k' - k|^{-\alpha} \int \frac{\tilde{\Lambda}(t)}{|2^j(\frac{t}{k'-k} + 1)|^{\alpha}}\, dt \\ &\sim c_r 2^{j(1-\alpha)}|k' - k|^{-\alpha} \int \tilde{\Lambda}(t)\, dt,\end{aligned}$$

where the last equivalence follows from the fact that $\tilde{\Lambda}(\cdot)$ has a compact support by expanding $(t/(k' - k) + 1)^{-\alpha}$ for $|t/(k' - k)| < 1$ as in Tewfik and Kim (1992). Noting that $\int \tilde{\Lambda}(t)\, dt = (\int \phi(t)\, dt)^2 \neq 0$ (Wojtaszczyk, 1997, Proposition 3.16) we obtain the result. ■

It follows from the last theorem that in contrast to the wavelet coefficients the decorrelation property does not hold for the approximation coefficients $a_{j,k}$ of the LRD process and their strength of dependence within fixed scale matches that of process $X(t)$. For example, for the Haar wavelet $a_{j,k}$ form a discrete aggregated process for each $j$ and a sequence of normalized aggregated processes tends to the FGN with the same $\gamma$ as in (9).

Let us consider a parallel approach to study dependence of wavelet coefficients and investigate how their spectral densities on different levels relate. The spectral density of $(a_{j,\cdot})$ and $(d_{j,\cdot})$ will be denoted by $f_j^a(\cdot)$ and $f_j^d(\cdot)$, respectively. We have

PROPOSITION 3.9 *Assume that the spectral density $f$ is bounded outside $[-2^{-j}\pi, 2^{-j}\pi]$ for some $j \geq 0$ and $\phi, \psi \in \mathcal{L}^1(\mathbb{R})$. Then*

$$f_j^a(\lambda) = \sum_{n\in\mathbb{Z}} f\Big(\frac{\lambda + 2n\pi}{2^j}\Big)|\hat{\phi}(\lambda + 2n\pi)|^2 \tag{14}$$

*and*

$$f_j^d(\lambda) = \sum_{n\in\mathbb{Z}} f\Big(\frac{\lambda + 2n\pi}{2^j}\Big)|\hat{\psi}(\lambda + 2n\pi)|^2. \tag{15}$$

*Proof.* In view of $\mathrm{Cov}(a_{j,k}, a_{j,k'}) = \int \tilde{f}_j(\lambda)e^{i\lambda(k'-k)}\, d\lambda$ , where $\tilde{f}_j(\lambda) = f(\lambda/2^j)|\hat{\phi}(\lambda)|^2$, to prove (14) it is enough to show that

$$\int_{\mathbb{R}} f_j(\lambda)e^{i\lambda l}\, d\lambda = \int_{-\pi}^{\pi} f_j^a(\lambda)e^{i\lambda l}\, d\lambda.$$

Observe that the integral on the right hand side exists. This follows from separate consideration of the summand for $n = 0$ and summands for $n \neq 0$ in the definition of $f_j^a(\cdot)$. The pertaining integral is finite due to $f \in \mathcal{L}^1(\mathbb{R})$ and $\sup|\hat{\phi}(\cdot)| < \infty$ in the first case and due to boundedness of $f$ outside a small neighborhood of 0 and $\int_{-\pi}^{\pi}\sum_{n\in\mathbb{Z}}|\hat{\phi}(\lambda + 2n\pi)|^2 = \int_{\mathbb{R}}|\hat{\phi}(\lambda)|^2\, d\lambda < \infty$ in the second. Equality (14) follows from the observation that

$$\Big|\int_{-\pi}^{\pi}\sum_{n\in\mathbb{Z}} f\Big(\frac{\lambda + 2n\pi}{2^j}\Big)|\hat{\phi}(\lambda+2n\pi)|^2 e^{il\lambda}\, d\lambda - \int_{-k\pi}^{k\pi} f\Big(\frac{\lambda}{2^j}\Big)|\hat{\phi}(\lambda)|^2 e^{il\lambda}\, d\lambda\Big| \to 0$$

as $k \to \infty$. The above expression is bounded by $\int_{-\pi}^{\pi}\sum_{|n|>k} f((\lambda+2n\pi)/2^j)|\hat{\phi}(\lambda+2n\pi)|^2\, d\lambda$ which tends to 0 in view of the previous argument. ■

Recall that $v(z) = \sum_{j\in\mathbb{Z}} v_j z^j, V(\lambda) = |v(e^{i\lambda})|^2$, and $u(z)$ and $U(z)$ are analogously defined functions pertaining to the sequence $(u_i)$. Then the spectral density of $v_\cdot^\vee \star a_{0,\cdot}$ equals to $V(\lambda)f_0^a(\lambda)$ (see Brockwell, Davis, 1987, Theorem 4.4.1). In order to derive the spectral density of the decimated process $v_\cdot^\vee \star a_{0,\cdot}(2k)$ observe that the spectral density of $\bar{Z}_k = Z_{2k}$ is $g_1(\lambda) = 2^{-1}(g(\lambda/2) + g(\lambda/2+\pi))$, where $g(\cdot)$ is the spectral density of $(Z_k)$ periodically extended to $\mathbb{R}$. This follows from the observation that

$$r_{\bar{Z}}(k) = \frac{1}{2}\int_{-2\pi}^{2\pi} g(\lambda/2)e^{ik\lambda}\,d\lambda = \int_{-\pi}^{\pi} g_1(\lambda)e^{ik\lambda}\,d\lambda,$$

where periodicity of $g(\cdot)$ was used for the last equality. Thus, in view of recursive relations of the pyramidal algorithm we have

Proposition 3.10 *For any* $i \geq 1$

$$f_i^a(\lambda) = \frac{1}{2}(U(\lambda/2)f_{i-1}^a(\lambda/2) + U(\lambda/2+\pi)f_{i-1}^a(\lambda/2+\pi))$$

*and*

$$f_i^d(\lambda) = \frac{1}{2}(V(\lambda/2)f_{i-1}^a(\lambda) + V(\lambda/2+\pi)f_{i-1}^a(\lambda/2+\pi)).$$

Observe that it follows from the Proposition 3.10 that

$$\begin{aligned} f_1^d(\lambda) &= \frac{1}{2}\Big(\sum_{n\in\mathbb{Z}} f(\lambda/2+2n\pi)|\hat{\phi}(\lambda/2+2n\pi)|^2 V(\lambda/2) \\ &+ \sum_{n\in\mathbb{Z}} f(\lambda/2+(2n+1)\pi)|\hat{\phi}(\lambda/2+(2n+1)\pi)|^2 V(\lambda/2+\pi)\Big) \end{aligned}$$

and the expression on the right hand side equals (15) for $j=1$ due to equality $|\hat{\psi}(\lambda)|^2 = 2^{-1}V(\lambda/2)|\hat{\phi}(\lambda/2)|^2$ (see, e.g. Vidakovic, 1999, equation (3.13)).

## 4. Time series with discrete time

The wavelet transform is defined for a continuous time stochastic process and it is important to understand whether its scope can be extended to discrete time series $(X(n))_{n\in\mathbb{Z}}$. We will discuss two approaches to tackle this problem.

### 4.1. Embedding in a continuous time process (Veitch et al., 2000)

This approach involves extending $(X(n))_{n\in\mathbb{Z}}$ to $(\tilde{X}(t))_{t\in\mathbb{R}}$ in such a way that $X(n) = \tilde{X}(n)$ for $n \in \mathbb{Z}$ and spectral densities of $X(n)$ and $\tilde{X}(n)$ coincide. Consider

$$\tilde{X}(t) = \sum_{n\in\mathbb{Z}} X(n)sinc(t-n) \tag{16}$$

where

$$sinc(x) = \begin{cases} \frac{\sin(\pi x)}{\pi x} & \text{for } x \neq 0; \\ 1 & \text{for } x = 0. \end{cases}$$

It turns out that such $\tilde{X}(t)$ is well defined.

PROPOSITION 4.1 *(Veitch et al., 2000) Let $X(n)$ be a weakly stationary process with mean 0. Then $\tilde{X}(t)$ defined in (16) converges in $\mathcal{L}^2(\mathbb{R})$ for any $t \in \mathbb{R}$. Moreover, the random spectral measure of $(\tilde{X}(t))$ coincides with the random spectral measure $Z$ of $(X(n))_{n\in\mathbb{Z}}$.*

*Proof.* For a fixed $0 \le r < 1$ define $Y(k) = \tilde{X}(k+r)$ for $k \in \mathbb{Z}$ and observe that

$$\begin{aligned} Y(k) &= \sum_{n\in\mathbb{Z}} X(n) sinc(k+r-n) = \sum_{n\in\mathbb{Z}} X(k-n) sinc(n+r) \\ &= \sum_{n\in\mathbb{Z}} X(k-n) h_r(n) = X \star h_r(k), \end{aligned}$$

where $h_r(n) = sinc(n+r)$. The convolution $X \star h_r(k)$ converges in $\mathcal{L}^2(\mathbb{R})$ provided $h_{n,r}(e^{-i\lambda}) := \sum_{|j|\le n} h_r(j) e^{-ij\lambda}$ converges in $\mathcal{L}^2([-\pi,\pi), F)$, where $F$ is the spectral measure of $(X(n))$. Moreover, denoting by $h(e^{-i\lambda})$ the limit of $h_{n,r}(e^{-i\lambda})$ we have $Y(k) = \int_{-\pi}^{\pi} h(e^{-i\lambda}) e^{ik\lambda}\, dZ(\lambda)$. Observe that $sinc(x+r)$ is the Fourier transform of $(2\pi)^{-1} e^{i\lambda r} I_{[-\pi,\pi)}(\lambda)$ at $x$. As $h_{n,r}(e^{-i\lambda}) \to e^{ir\lambda} I_{[-\pi,\pi)}(\lambda)$ for any $\lambda \in [-\pi,\pi)$ and $\sup_{|\lambda|\le\pi} |h_{n,r}(e^{-i\lambda})|$ is uniformly bounded in $n$ (see Brockwell, Davis, 1987, Proposition 4.11.2), convergence in $\mathcal{L}^2([-\pi,\pi), F)$ holds and thus $Y(k)$ is well defined. Moreover,

$$\tilde{X}(k+r) = Y(k) = \int_{-\pi}^{\pi} e^{ir\lambda} e^{ik\lambda}\, dZ(\lambda) = \int_{-\pi}^{\pi} e^{i(r+k)\lambda}\, dZ(\lambda)$$

for arbitrary $t = k + r$. ■

Thus, from the last proposition it follows, in particular, that if the spectral density of $X(n)$ exists then the wavelet coefficients $\tilde{d}_{j,k}$ pertaining to the process $\tilde{X}(t)$ satisfy

$$\mathrm{Cov}(\tilde{d}_{j,k}, \tilde{d}_{j',k'}) = \int_{-\pi}^{\pi} f(\lambda) \hat{\psi}_{j,k} \hat{\psi}^*_{j',k'}(\lambda)\, d\lambda = \int r(t-s) \psi_{j,k}(t) \psi_{j'k'}(s)\, dt ds,$$

where $r(t) := \int_{-\pi}^{\pi} e^{i\lambda t} f(\lambda)\, d\lambda$ for $t \in \mathbb{R}$. Hence the results of Section 3 are valid for $\tilde{d}_{j,k}$ if $f$ is meant as the spectral density of the sequence $X(n)$. Consider in particular a white noise process $(\varepsilon(n))_{n\in\mathbb{Z}}$. Then $\tilde{\varepsilon}(t)$ is correlated with $\tilde{\varepsilon}(t')$ if $t - t' \in \mathbb{R} \setminus \mathbb{Z}$. Indeed, $r_{\tilde{\varepsilon}}(t) = (\sigma^2/2\pi) \int_{-\pi}^{\pi} e^{it\lambda}\, d\lambda = sinc(t)\sigma^2$ and $\int_{-\infty}^{\infty} |r_{\tilde{\varepsilon}}(t)|\, dt = \infty$, thus $(\tilde{\varepsilon}(t))$ is long-range dependent. However, we have

PROPOSITION 4.2 *Assume that the support of $\hat{\psi}$ is contained in $[-2^l\pi, 2^l\pi]$ for some $l \in \mathbb{Z}$. Then for $(j,k) \neq (j',k')$ and $\min(j,j') \geq l$ the wavelet coefficients $\tilde{d}_{j,k}$ and $\tilde{d}_{j',k'}$ are uncorrelated.*

*Proof.* We have

$$\mathrm{Cov}(d_{j,k}, d_{j',k'}) = \frac{\sigma^2}{2\pi} 2^{(j+j')/2} \int_{-\pi}^{\pi} \hat{\psi}(2^j\lambda)\hat{\psi}^*(2^{j'}\lambda)e^{i\lambda(2^jk-2^{j'}k')}\,d\lambda$$

$$= \frac{\sigma^2}{2\pi} 2^{(j+j')/2} \int \hat{\psi}(2^j\lambda)\hat{\psi}^*(2^{j'}\lambda)e^{i\lambda(2^jk-2^{j'}k')}\,d\lambda$$

$$= \sigma^2 \int \psi_{j,k}(t)\psi_{j',k'}(t)\,dt = 0,$$

where the penultimate equality follows from the fact that the support of $\hat{\psi}(2^j\cdot)$ is contained in $[-2^{l-j}\pi, 2^{l-j}\pi] \subset [-\pi,\pi]$ for $j \geq l$. ■

In particular, the proposition holds for the Meyer wavelet with $l = 2$ as the support of $\hat{\psi} \subset [-4\pi, 4\pi]$ and for the Shannon wavelet with $l = 1$.

A simple calculation shows that the approximation coefficients $\tilde{a}_{0,k}$ pertaining to the process $\tilde{X}$ are actually obtained by filtering the process $X(n)$ with the filter $I(\cdot)$, where $I(m) = \int sinc(t+m)\phi(t)\,dt$. Thus there is no need to calculate $\tilde{X}(t)$ for all $t$ to evaluate $(a_{0,k})$.

It is common to consider long-range dependent processes for discrete time. In this case they are either defined by the condition (9) in which $f$ is the spectral density of the process defined on $(-\pi,\pi]$ or by $r(k) \sim Ck^{\gamma-1}$ where $r(\cdot)$ is the pertaining covariance function of $X(n)$. These two conditions are equivalent provided $r(\cdot)$ is of bounded variation and quasi-decreasing i.e. $r(k+1) \leq r(k)(1+C/k)$ for some $C < \infty$ and sufficiently large $k$ (see Yong, 1974, Theorem III-14).

EXAMPLE 4.1
Fractional autoregressive integrated moving average FARIMA$(0,d,0)$ with $0 < d < 1/2$ is defined as a process $X(n)$ such that $(1-B)^dX(n) = \varepsilon(n)$, where $(\varepsilon(n))_{-\infty}^{\infty}$ is i.i.d. Gaussian $N(0,\sigma^2)$ sequence. Moreover, $(1-B)^d$ is the fractional differencing operator defined by $(1-B)^d = \sum_{k=0}^{\infty}\binom{d}{k}(-1)^kB^k$ where $\binom{d}{k} = \Gamma(d+1)/(\Gamma(k+1)\Gamma(d-k+1))$, $\Gamma(\cdot)$ is the gamma function and $B^kX(n) = X(n-k)$. For properties of FARIMA$(0,d,0)$ processes including the proof that they are LRD with $\gamma = 2d$ see, e.g., Beran (1994).

### 4.2. Direct application of the pyramidal algorithm to $(X(n))_{n\in\mathbb{Z}}$

The second approach to deal with a discrete time series consists in formally defining the approximation coefficients at the scale 0 as $a_{0,n} = X(n)$ and using the recursions of the pyramidal algorithm to define the coefficients $d_{j,n}$ and $a_{j,n}$ for $j > 0$. Observe that this can be put in a continuous time framework considered in Section 4.1 by embedding $X(n)$ in a process $X^{\diamond}(t) = \sum_{n\in\mathbb{Z}} X(n)\phi(t-n)$.

Indeed, due to orthogonality of $\{\phi(\cdot - k)\}$ for different $k$ we have that

$$a_{0,n} = \int X^{\diamond}(t)\phi(t-n)\,dt = \int \left( \sum_{k\in\mathbb{Z}} X(k)\phi(t-k) \right) \phi(t-n)\,dt = X(n).$$

Moreover, one can prove by the methods similar to those used for Proposition 4.1 that the spectral density of $X^{\diamond}(\cdot)$ equals $\tilde{f}(\lambda)|\hat{\phi}(\lambda)|^2$, where $\tilde{f}(\cdot)$ is the periodic extension of $f(\cdot)$ to $\mathbb{R}$. Thus, the only difference between the process $\tilde{X}(t)$ in (16) and $X^{\diamond}(t)$ is that the spectral densities of $X(n)$ and $X^{\diamond}(t)$ differ when $\phi(\cdot) \neq sinc(\cdot)$. However, if $X(n)$ is LRD, then $X^{\diamond}(t)$ retains this property with the same $\gamma$ when $\hat{\phi}(\cdot)$ is continuous at 0. This is true in most cases e.g. when the number of nonzero $v_i$ is finite (see Wojtaszczyk, 1997, Theorem 4.1). Note, moreover, that for a scaling function $\phi$ with the support in [0,1] putting $\bar{X}(t) := X(n)$ for $t \in [n, n+1)$ yields $(a_{0,n}) = CX(n)$. It is frequently easier to study the dependence structure of the wavelet coefficients by referring directly to the pyramidal algorithm starting at $X(n)$. The rest of this section is devoted to providing examples of such an approach. In particular, equations (3) and (4) imply that

$$\mathrm{Cov}(d_{j,k}, d_{j',k'}) = \sum_{m,n} v_{m-2k} v_{n-2k'} \mathrm{Cov}(a_{j-1,m}, a_{j'-1,n})$$
$$\mathrm{Cov}(a_{j,k}, a_{j',k'}) = \sum_{m,n} u_{m-2k} v_{n-2k'} \mathrm{Cov}(a_{j-1,m}, a_{j'-1,n})$$

(see, e.g. Vanucci, Corradi, 1999) which can be used to compute recursively the covariance structure of the wavelet coefficients. Moreover, as in this approach the pyramidal algorithm applies, Proposition 3.10 is still true with $f_0^a(\cdot) = f(\cdot)$, where $f(\cdot)$ stands for the spectral density of the sequence $X(n)$. We show now that this recursive relation implies decorrelation property of $d_{jk}$ for a discrete long-range dependent processes when the Daubechies wavelet $D(N)$ is used.

PROPOSITION 4.3 *Let $X(n)$ be a weakly stationary process with mean 0 and $f \in \mathcal{L}^1[-\pi, \pi)$ its spectral density. If $f^{(l)}$ exists almost everywhere, is bounded on the compact subsets of $[-\pi, \pi)\backslash\{0\}$ and $x^l f^{(l)}(x)$ is integrable for $l = 0, 1, \ldots, 2N$, then the process $(d_{j\cdot})$ of Daubechies $D(N)$ wavelet coefficients for $X(n)$ satisfies*

$$|r_j^d(k)| = o(|k|^{-2N}),$$

*where $r_j^d(\cdot)$ is the covariance function of $(d_{j\cdot})$ and $k \to \infty$.*

*Proof.* Consider first $j = 1$. The sequence $(r_1^d(k))_{k\in\mathbb{Z}}$ consists of the Fourier coefficients of $f_1^d$. Observe that it is enough to show that $f_1^d$ has an integrable derivative of order $2N$ and $(f_1^d)^{(l)}(-\pi) = (f_1^d)^{(l)}(\pi)$ for $l = 0, 1, \ldots, 2N-1$, as then we will obtain the assertion via integration by parts as in Proposition 3.6. By Proposition 3.10 $(f_1^d)^{(l)}(x) = \frac{1}{2^{l+1}} \sum_{i+i'=l} P_{ii'}(x)$, where

$$P_{ii'}(x) = f^{(i)}(x/2)V^{(i')}(x/2) + f^{(i)}(x/2+\pi)V^{(i')}(x/2+\pi).$$

Since $P_{ii'}(\pi) = P_{ii'}(-\pi)$ for $i + i' \leq 2N$, also $(f_1^d)^{(l)}(-\pi) = (f_1^d)^{(l)}(\pi)$.

As $V^{(i')}(x) = \mathcal{O}(|x|^{2N-i'})$ by Example 2.4 in Section 2, the first component of $P_{ii'}$ is then $\mathcal{O}(f^{(i)}(x/2)|x|^{2N-i'})$ and thus is integrable by assumption, while the second is bounded, as $f^{(i)}(x/2+\pi)$ is bounded in $[-\pi,\pi)$ and $V^{(i')}$ exists and is bounded everywhere.

In order to prove the result for an arbitrary $j$, it is enough to show that $f_1^a$ shares the properties of $f$. Then, by induction, $f_j^a$ satisfies the assumptions of the proposition and the above argument applied to $f_{j+1}^d$ yields $r_{j+1}^d(k) = o(|k|^{-2N})$. Again, by Proposition 3.10

$$(f_1^a)^{(l)}(x) = \frac{1}{2^{l+1}} \sum_{i+i'=l} f^{(i)}(x/2)U^{(i')}(x/2) + f^{(i)}(x/2+\pi)U^{(i')}(x/2+\pi).$$

Since $U^{(i')}$ exist and are bounded everywhere, derivatives of $f_1^a$ exist almost everywhere for $l = 0, 1, \ldots, 2N$ and are bounded on compact subsets of $[-\pi,\pi)\backslash\{0\}$. Integrability of $x^l(f_1^a)^{(l)}(x)$ follows from integrability of $x^i f^{(i)}(x/2)$ and
$x^{i'}U^{(i')}(x/2)$. ■

By a simple induction we get from Proposition 3.10 that for $j = 1, 2, \ldots$

$$f_j^d(\lambda) = \frac{1}{2^j} \sum_{k=0}^{2^j-1} V_j\Big(\frac{\lambda+2k\pi}{2^j}\Big) f\Big(\frac{\lambda+2k\pi}{2^j}\Big), \tag{17}$$

where $V_j(\lambda) := V(2^{j-1}\lambda)\prod_{l=0}^{j-2} U(2^l\lambda)$ and $f(\cdot)$ is periodically extended to $\mathbb{R}$ (see Percival and Walden, 2000, exercise 348b).

We will prove now that the rate of decorrelation established in Theorem 3.1 for the FGN process also holds for the FARIMA process, i.e. in the case of the FARIMA process we can obtain a more stringent bound on the rate of decorrelation.

THEOREM 4.1 *Let $X(n)$ be the FARIMA$(0,d,0)$ process with $0 < d < 1/2$ and $d_{j,k}$ are defined as in Proposition 4.3. Then*

$$|r_j^d(k)| = \mathcal{O}(|k|^{-2(N-H)-2}) \qquad \text{when} \quad k \to \infty.$$

*Proof.* We prove the result for $j = 1$, the general case is proved similarly using (17). In view of the proof of Proposition 4.3 and the fact that the spectral density of the FARIMA$(0,d,0)$ process equals $f(\lambda)=(\sigma^2/2\pi)2^{-2d}\sin^{-2d}(|\lambda|/2)$ for $\lambda\in(-\pi,\pi]$, it is enough to prove that

$$\int_0^\pi (\sin^{-2d}(\lambda/2)V(\lambda/2))^{(2N)}\cos k\lambda\,d\lambda=\mathcal{O}(k^{2d-1}),$$

taking into account symmetry of $f$ and $H=d+1/2$. The above equality will follow from

$$\int_0^\pi (\sin^{-2d}(\lambda/2))^{(s)}(V(\lambda/2))^{(2N-s)}\cos k\lambda\,d\lambda=\mathcal{O}(k^{2d-1}) \tag{18}$$

for $s=0,1,\ldots,2N$. Observe that in view of (6) $V^{(i)}(\lambda)=C(\sin^{2N-1}\lambda)^{(i-1)}$ for $i\geq 1$ and thus for $s<2N$ the integrand in (18) equals $C(\sin^{-2d}(\lambda/2))^{(s)}(\sin^{2N-1}(\lambda/2))^{(2N-s-1)}\cos k\lambda$. We expand the derivatives and note that that the only unbounded term in the integrand is of the form $C(\sin^{-2d}(\lambda/2))\cos^{2N-1}\cos k\lambda$. We consider this term, the other terms are treated using analogous but simpler reasoning. Thus, it is sufficient to prove that

$$\int_0^\pi \sin^{-2d}(\lambda/2)\cos^{2N-1}(\lambda/2)\cos k\lambda\,d\lambda=\mathcal{O}(k^{2d-1}).$$

Integrating by parts we see that the above integral equals

$$\frac{d}{k}\int_0^\pi \sin^{-2d-1}(\lambda/2)\cos^{2N}(\lambda/2)\sin k\lambda\,d\lambda$$
$$+\frac{2N-1}{2k}\int_0^\pi \sin^{-2d+1}(\lambda/2)\cos^{2N-2}(\lambda/2)\sin k\lambda\,d\lambda.$$

The integrand of the second integral is bounded and thus the second term is $\mathcal{O}(k^{-1})$. In order to treat the first integral we approximate it by the analogous integral with the term $\sin^{-2d-1}(\lambda/2)$ replaced by $(\lambda/2)^{-2d-1}$. Substituting $\lambda:=k\lambda/2$ we see that the difference between the two integrals is not larger than

$$\frac{C}{k^2}\int_0^{k\pi/2}|(\sin\lambda/k)^{-2d-1}-(\lambda/k)^{-2d-1}|(\cos\lambda/k)^{2N}|\sin 2\lambda|\,d\lambda. \tag{19}$$

We have that

$$\begin{aligned}&|\sin^{-2d-1}(\lambda/k)-(\lambda/k)^{-2d-1}|\leq\\ &\qquad\leq C\max(\sin(\lambda/k)^{-2d-2},(\lambda/k)^{-2d-2})|\sin(\lambda/k)-(\lambda/k)|\\ &\qquad=\mathcal{O}((\lambda/k)^{(-2d+1)})\end{aligned}$$

in view of $|\sin(\lambda/k)-(\lambda/k)|=\mathcal{O}((\lambda/k)^3)$ and $\sin(\lambda/k)\geq(2/\pi)(\lambda/k)$ for $\lambda\in[0,k\pi/2]$. Thus, (19) is $\mathcal{O}(k^{-1})$. Moreover, observe that the approximating integral

$$\begin{aligned}
\frac{C}{k^2}\int_0^{k\pi/2} &(\lambda/k)^{-2d-1}(\cos\lambda/k)^{2N}\sin 2\lambda\, d\lambda \\
&= Ck^{2d-1}\int_0^{k\pi/2}\lambda^{-2d-1}\cos^{2N}(\lambda/k)\sin 2\lambda\, d\lambda \\
&= \mathcal{O}(k^{2d-1})
\end{aligned}$$

since the above integral is bounded for $k \in \mathbb{N}$ as the integrand is $\mathcal{O}(\lambda^{-2d-1})$ for large $\lambda$ and $\mathcal{O}(\lambda^{-2d})$ for $\lambda$ close to 0. ■

REMARK 4.1 Observe that in both cases of the FGN and the FARIMA process considered in Theorems 3.1 and 4.1 it was essential to use the exact form of either the covariance function or the spectral density to prove the rate $\mathcal{O}(n^{-2(N-H)-2})$ of within-scale decorrelation.

Intuitively, in view of (17), $f_j^d(\cdot)$ corresponds to the spectral density of the sequence obtained by taking every $2^j$th element of the sequence pertaining to the the sequence $X(n)$ filtered by a filter $\mathbf{v}_j = (v_{j,k})$ with the power function $V_j$. Indeed, we have

PROPOSITION 4.4 *For $l \in \mathbb{Z}$*

$$\int_{-\pi}^{\pi} f_j^d(\lambda)e^{i\lambda l}\, d\lambda = \int_{-\pi}^{\pi} f(\lambda)V_j(\lambda)e^{i2^j\lambda l}\, d\lambda. \tag{20}$$

*Proof.* Observe that in view of (17) the left-hand side of (20) can be written as

$$\frac{1}{2^j}\sum_{k=0}^{2^j-1}\int_{-\pi}^{\pi} V_j\Big(\frac{\lambda+2k\pi}{2^j}\Big) f\Big(\frac{\lambda+2k\pi}{2^j}\Big)e^{i\lambda l}\, d\lambda.$$

Changing variable to $\lambda := (\lambda + 2k\pi)/2^j$ in each integral above separately and using periodicity of $e^{il\cdot}$ and $V_j f(\cdot)$ we get that the above expression equals to

$$\begin{aligned}
\sum_{k=0}^{2^j-1}\int_{(2k-1)\pi/2^j}^{(2k+1)\pi/2^j} V_j(\lambda)f(\lambda)e^{i2^j\lambda l}\, d\lambda &= \int_{-\pi/2^j}^{2\pi-\pi/2^j} V_j(\lambda)f(\lambda)e^{i2^j\lambda l}\, d\lambda \\
&= \int_{-\pi}^{\pi} V_j(\lambda)f(\lambda)e^{i2^j\lambda l}\, d\lambda.
\end{aligned}$$

■

The proposition is equivalent to equation (348b) in Percival and Walden (2000). Here, our line of argument is reversed as Proposition 3.10 and (17) are proved first. Moreover, observe that equality (17) leads to the following analogue of Proposition 4.3 stated for a general finite filter $v$.

THEOREM 4.2 *Let $(v_i)_{i=0}^{L-1}$ be a finite wavelet filter such that its transfer function $v(\lambda) = \sum_{k=0}^{L-1} v_k e^{ik\lambda}$ satisfies $v^{(i)}(0) = 0$ for $0 \le i \le N-1$. Then Proposition 4.3 holds under conditions on $f$ assumed there.*

*Proof.* Observe that the condition on $v(\lambda)$ translates to $\sum_{k=0}^{L-1} k^j v_k = 0$ for $0 \le j \le N-1$. Thus

$$\sum_{0\le k,l\le L-1} (k-l)^j v_k v_l = \sum_{s=0}^{j} \binom{j}{s} \sum_{k=0}^{L-1} k^s v_k \sum_{l=0}^{L-1} l^{j-s} v_l = 0$$

for $0 \le j \le 2N-1$, which is equivalent to $V^{(j)}(0) = 0$ for $0 \le j \le 2N-1$. Now the proof follows analogously to the proof of Proposition 4.3 by using (20) and noting that since $V^{(2N)}(0) = \mathcal{O}(1)$ we have that $V^{(k)}(\lambda) = \mathcal{O}(|\lambda|^{2N-k})$ in $[-\pi, \pi)$ and this order is inherited by $V_j(\lambda)$ as derivatives of $U(\cdot)$ exist and are bounded in $[-\pi, \pi)$. ■

REMARK 4.2 A simple inductive argument (see Vidakovic, 1999, p. 83) shows that the condition $v^{(i)}(0) = 0$ for $0 \le i \le N-1$ is equivalent to vanishing of first $N$ moments of $\psi$. Observe also that the above proof indicates that, heuristically, $\int (f_j^d)^{(2N)}(x) e^{ikx}\, dx$ should behave as $C_1 \int_{-\pi}^{\pi} x^{-\gamma} e^{ikx}\, dx \sim C_2 k^{\gamma-1}$ (Zygmund, 1959, p. 186). Thus it is plausible that $r_j^d(k) = \mathcal{O}(k^{-2N-1-\gamma})$ under assumptions of Theorem 4.2. We were unable, however, to make this argument formal apart from a special case of the FARIMA process studied in Theorem 4.1.

Let us note that Craigmile and Percival (2005) studied the rate of between-scales decorrelation for certain generalized fractionally differenced processes when the length of the filter $L$ tends to infinity. For other results concerning between-scales decorrelation see Dijkerman and Mazumdar (1994) and McCoy and Walden (1996).

Observe that the filter $\mathbf{v}_j = (v_{j,k})$ with the power function $V_j$ has a unit energy, i.e. the following proposition holds which will be used in Section 5.2.

PROPOSITION 4.5 *For $j \in \mathbb{N}$ $\sum_{k\in\mathbb{Z}} v_{j,k}^2 = 1$.*

*Proof.* Proposition holds for $j = 1$ as $v_{1,k} = v_k$ defined in Section 1 and $\sum_{k\in\mathbb{Z}} v_k^2 = \| \psi_{1,0} \|_2^2 = 1$. Observe that

$$2\pi \sum_{k\in\mathbb{Z}} v_{j,k}^2 = \int_{-\pi}^{\pi} \sum_{m,n\in\mathbb{Z}} v_{j,m} v_{j,n} e^{i\lambda(m-n)}\, d\lambda = \int_{-\pi}^{\pi} V_j(\lambda)\, d\lambda.$$

Thus, it is enough to prove that $\int_{-\pi}^{\pi} V_j(\lambda)\, d\lambda = 2\pi$. Noting that $V_j(\lambda) =$

$V_{j-1}(2\lambda)U(\lambda)$ and using periodicity of $V_{j-1}(\cdot)$ and $U(\cdot)$ we have

$$\begin{aligned}\int_{-\pi}^{\pi} V_j(\lambda)\,d\lambda &= \int_{-\pi}^{\pi} V_{j-1}(2\lambda)U(\lambda)\,d\lambda = \frac{1}{2}\int_{-2\pi}^{2\pi} V_{j-1}(\lambda)U(\lambda/2)\,d\lambda \\ &= \frac{1}{2}\int_{-\pi}^{\pi} V_{j-1}(\lambda)\{U(\lambda/2)+U(\lambda/2+\pi)\}\,d\lambda \\ &= \int_{-\pi}^{\pi} V_{j-1}(\lambda)\,d\lambda\end{aligned}$$

as $U(\lambda)+U(\lambda+\pi)=2$ (see, e.g. Wojtaszczyk, 1997, Lemma 3.12). Thus, the proof follows by induction argument. ■

REMARK 4.3 Observe that it follows from the last two propositions that if $X(n)$ is the noise process with the variance $\sigma^2$ and with the corresponding spectral density $f(\lambda)=(\sigma^2/2\pi)I_{[-\pi,\pi)}(\lambda)$, then

$$\mathbb{E}d_{j,k}^2 = \int_{-\pi}^{\pi} f_j^d(\lambda)\,d\lambda = \frac{\sigma^2}{2\pi}\int_{-\pi}^{\pi} V_j(\lambda)\,d\lambda = \sigma^2.$$

Thus, the variance of wavelet coefficients at each level equals to the variance of the underlying process. Moreover, in this case equality holds in Proposition 3.4.

In order to understand better the decorrelation property of the wavelet coefficients in this case consider for a moment the situation when the sequences $(u_i)$ and $(v_i)$ correspond to ideal filters. Namely, let $(u_i)$ be the ideal (Shannon) low-pass filter such that the corresponding power function $U(\lambda)=2I_{[-\pi/2,\pi/2]}$ and $(v_i)$ is the ideal high-pass filter such that $V(\lambda)=2(I_{[-\pi,\pi)}-I_{[-\pi/2,\pi/2]}]$ for $\lambda\in[-\pi,\pi)$. Fig. 1 indicates how much power functions $U(\cdot)$ and $V(\cdot)$ for the Daubechies wavelet of order 5 differ from the ideal ones.

Equation (14) implies that in order to obtain $f_1^a$, the spectral density $f$ is dilated from the interval $[-\pi/2,\pi/2)$ to $[-\pi,\pi)$ and multiplied by 1/2. On the other hand, $f_1^d$ is obtained from $f$ in the following way: the spectral density is dilated from the interval $[\pi/2,\pi)$ to $[\pi,2\pi)$, multiplied by 1/2, moved by $2\pi$ to the left and symmetrically extended to $[0,\pi)$. In this way $f_1^d$ depends only on the restriction of $f$ to $[\pi/2,\pi)$. Analogously, $f_i^d$ is based on values of $f$ only for the frequencies $\lambda\in[\pi/2^i,\pi/2^{i-1})$ and thus it is not influenced by a possible pole at 0 of the spectral density $f$ in the long-range dependent case. Moreover, $(i-1)$fold decimation results in considerable flattening of the original part of the spectral density $f$ for large $i$. As a result, after few iterations $f_i^d$ resembles a constant function which is a spectral density of a white noise. Fig. 2 shows the spectral densities of autoregressive processes $AR(1,-0.6)$ and $AR(1,0.6)$ together with $f_i^d$ for $i=1,2,3$. We see that the shape of $f_1^d$ in these cases is considerably different as the region $[\pi/2,\pi)$ corresponds to the lowest

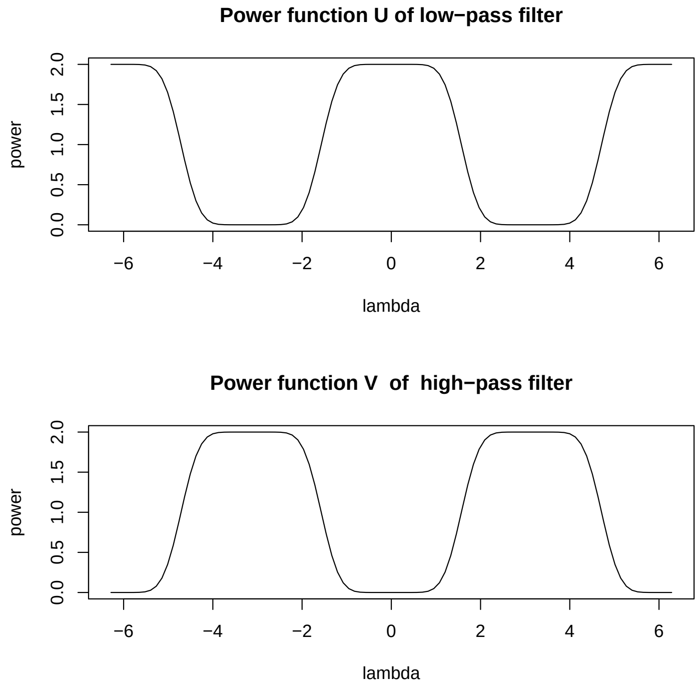


Figure 1. Power functions $U(\cdot)$ and $V(\cdot)$ for Daubechies wavelet of order 5

(respectively, the highest) part of the spectral density. However, most of the difference disappears for spectral densities of the wavelet coefficients of the third octave. Fig. 3 shows slow decay of correlations of $(a_{1,\cdot})$ and contrasts it with the behaviour of correlations of $(d_{1,\cdot})$ for a sample path of the FGN with $H = 0.9$ consisting of 2048 observations.

## 5. Applications

We focus on two applications related to the properties studied in the previous sections, namely on estimation of an exponent $\gamma$ of a spectral density at 0 in (9) for long-range dependent processes or equivalently their Hurst exponent $H$. First we consider estimation of $\gamma$ based on wavelet and scaling coefficients making use of (10). We focus here on regression based methods, MLE approach is described in Wornell (1995).

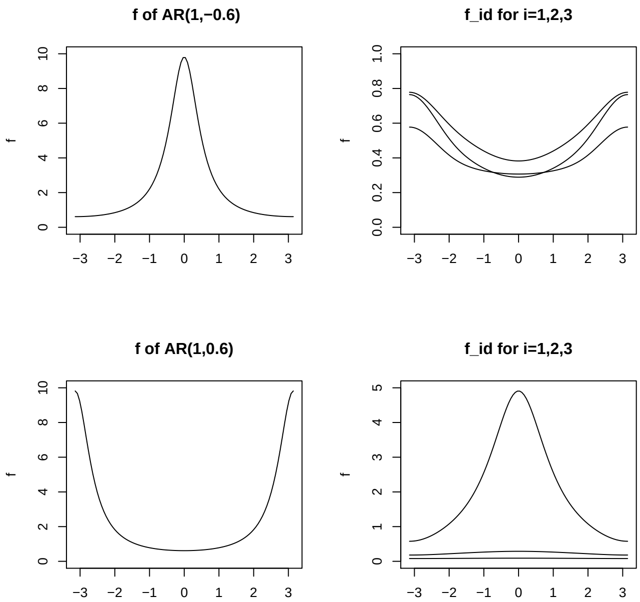


Figure 2. Spectral densities $f_i^d, i = 1, 2, 3$ for AR(1,-0.6) and AR(1,0.6) for Daubechies wavelet of order 5

### 5.1. Estimation of exponent $\gamma$ of spectral density at 0

Let $X(1), \ldots, X(n)$ be an observable part of a sample path of a discrete Gaussian LRD process, $\mu_j = \mathbb{E}\tilde{d}_{j,k}^2 = \mathbb{E}d_{j,k}^2$ and

$$\hat{\mu}_j = \frac{1}{n_j} \sum_{k=1}^{n_j} \tilde{d}_{j,k}^2,$$

where $n_j \approx n/2^j$ is a number of wavelet coefficients at the scale $2^j$ which can be calculated based on $X(1), \ldots, X(n)$ without extrapolating for the past or the future values. The coefficients $\tilde{d}_{j,k}^2$ are defined as in Section 4 and pertain to a continuous time process $\tilde{X}(t)$ in which $X(n)$ is embedded. In further considerations we assume that $\tilde{d}_{j,k}$ are independent within scales and among scales i.e. their weak dependence proved in Propositions 3.6, 3.8 and Theorem 3.1 is idealized to independence.

From (10) it follows that

$$\log_2 \mu_j \sim j\gamma + \log_2 C_1,$$

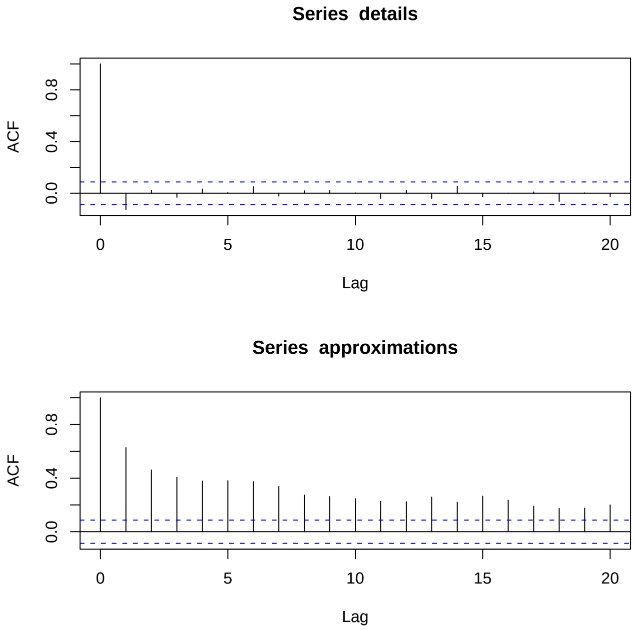


Figure 3. Autocorrelations for details $d_{1,k}$ and approximations $a_{1,k}$ for the FGN with $H = 0.9$ and $n = 2048$ and Daubechies wavelet of order 5

where $C_1 = c_f \int |x|^{-\gamma} |\hat{\psi}(x)|^2 \, dx$. Observe that as $\hat{\mu}_j$ is an unbiased estimate of $\mu_j$ and $X(n)$ are Gaussian, the assumed idealization implies that

$$\log_2 \hat{\mu}_j \quad \text{is approximately distributed as} \quad j\gamma + \log_2 C_1 + \log_2 \Big(\frac{Z_{n_j}}{n_j}\Big),$$

where $Z_{n_j}$ has $\chi^2$ distribution with $n_j$ degrees of freedom due to independence assumption. Using two-term Taylor expansion

$$\log_2 \Big(\frac{Z_{n_j}}{n_j}\Big) \approx \frac{1}{\ln 2}\Big(\frac{Z_{n_j}}{n_j} - 1 - \frac{1}{2}\Big(\frac{Z_{n_j}}{n_j} - 1\Big)^2\Big) \tag{21}$$

we get

$$g_j = \mathbb{E} \log_2 \Big(\frac{Z_{n_j}}{n_j}\Big) \approx \frac{-1}{n_j \ln 2}$$

and similarly

$$\mathrm{Var}\Big(\log_2 \Big(\frac{Z_{n_j}}{n_j}\Big)\Big) \approx \frac{2}{n_j \ln^2 2}.$$

Thus, the following equation approximately holds

$$y_j = \log_2 \hat{\mu}_j - g_j = j\gamma + \log_2 C_1 + \varepsilon_j, \tag{22}$$

for $j = 1, 2, \ldots, j_2$ where $\varepsilon_j := \log_2(\frac{Z_{n_j}}{n_j}) - g_j$ are independent and $\mathbb{E}\varepsilon_j \approx 0$, $\mathrm{Var}(\varepsilon_j) \approx 2(n_j \ln^2 2)^{-1}$. Integer $j_2$ is the maximal index for which $\hat{\mu}_j$ can be calculated based on the sample of size $n$. In order to estimate $\gamma$ Abry et al. (1995) and Abry and Veitch (1998) proposed to fit a weighted least squares (WLS) line to points $(j, \log_2 \hat{\mu}_j - g_j)$ and consider its slope $\hat{\gamma}_d$ as an estimate of $\gamma$. The corresponding estimate of $H$ is $\hat{H}_d = (1 + \hat{\gamma}_d)/2$. WLS coefficients $\beta_0$ and $\beta_1$ are obtained as minimizers of $WSST = \sum_j w_j(y_j - \beta_0 - \beta_1 j)^2$, where $w_j$ are proportional to $\mathrm{Var}(\varepsilon_j)^{-1}$. Thus, in the considered case one takes $w_j = n_j$. Veitch and Abry (1999) also proposed a method of estimating $c_f$ in (9). Bardet et al. (2000) studied the properties of $\hat{H}_d$.

Observe that (22) holds only approximately due to the fact that the relation (9) is asymptotic for $j \to \infty$. This suggests rejecting a certain number of lowest octaves when fitting WLS line, i.e. only points corresponding to $j \in \{j_1, \ldots, j_2\}$ are considered in (22) where $j_1 \geq 1$ is a chosen integer. The data-dependent choice of the lowest octave is proposed in Taqqu et al. (2003).

It turns out that assuming independence of the wavelet coefficients at each scale, $\hat{H}_d$ is an unbiased estimator of $H$ and its asymptotic variance is close to the minimal asymptotic variance given by the Cramér-Rao bound. Moreover, $\hat{H}_d$ outperforms, in terms of the Mean Squared Error, most of nonparametric estimators of the Hurst parameter based on the scaling property, such as the variogram or the $R/S$ estimator and performs a on par with the Whittle maximum likelihood estimator when a parametric form of the process is assumed. We stress that the construction the wavelet estimator, in particular the form of the bias terms $g_j$ and weights $w_j$ rests crucially on the decorrelation property of the wavelets coefficients.

Note that construction of a competing estimator of $H$ based on the analogous asymptotic equivalence for the approximation coefficients (11) is possible. However, as $a_{jk}$ are long-range dependent, no matter how many vanishing moments the wavelet $\psi$ has, all advantages of the above approach to develop $\hat{H}_d$ would be lost here. Also, let us note that other estimators of $H$ for parametric models based on approximate decorrelation of the wavelet coefficients have been proposed (see e.g. Zhang et al., 2004, and Craigmile et al., 2005).

## 5.2. Simulation of a stationary process with a given spectrum

We briefly describe the idea of generating sample paths of a stationary Gaussian process with a spectral density specified in analytic form. There are two main reasons for applying the wavelet approach. The first is that a classical Durbin-Levinson algorithm using covariances $\gamma(0), \gamma(1), \ldots, \gamma(n-1)$ to generate $X(1)$,

$X(2), \ldots, X(n)$ is quite expensive computationally as it requires $\mathcal{O}(N^2)$ operations. Secondly, in many cases it is more convenient to use frequency domain information in the form of a spectral density instead of a covariance function in time domain. Davis and Harte (1987) provide an example of such approach.

The main idea of approximate generation of a discrete time stochastic process with a given spectral density using wavelets relies on (5) and the fact that variances of detail and approximation coefficients can be estimated from the spectrum. We have in view of Proposition 4.4 $\mathrm{Var}(d_{j,k}) = \int_{-\pi}^{\pi} V_j(\lambda) f(\lambda)\, d\lambda$, where $V_j(\cdot)$ is defined below (17). Consider now the power function $V_j$ corresponding to the ideal high-pass filter. According to the disscusion below (17) $V_j(\cdot)$ is equal to a constant $C$ on intervals $[-\pi/2^{j-1}, -\pi/2^j]$ and $[\pi/2^j, \pi/2^{j-1}]$ and 0 elsewhere. It follows from the normalization condition $\int_{-\pi}^{\pi} V_j(\lambda)\, d\lambda = 2\pi$ proved in Proposition 4.5 that $C = 2^j$. Thus, using the mean value theorem for integrals we have

$$\mathrm{Var}(d_{j,k}) = 2^{j+1} \int_{\pi/2^j}^{\pi/2^{j-1}} f(\lambda)\, d\lambda \approx 2\pi f(\lambda_j),$$

where $\lambda_j$ is the midpoint of the interval $[\pi/2^j, \pi/2^{j-1}]$.

Suppose now that we would like to generate $n = 2^N$ observations pertaining to discrete time zero-mean Gaussian processes having a spectral density $f$ using resolution levels $j = 1, 2, \ldots, J$. Since $d_{j,k}$ are normal $N(0, \sigma_j^2)$ with $\sigma_j^2 \approx 2\pi f(\lambda_j)$ we generate $d_{j,k}, j = 1, \ldots, J, k = 1, \ldots, N_j = 2^{N-j}$ as *independent* normal variables with variances specified accordingly. We stress that the decorrelation property of wavelet coefficients is used here. A sample path is generated using approximation to (5) for discrete time

$$X_{sim}(n) = a_{J,1}\phi_{J,1}(n) + \sum_{j=1}^{J} \sum_{k=1}^{N_j} d_{j,k}\psi_{j,k}(n).$$

Here, $a_{J,1}$ is a normal variate with mean 0 and the variance $v_J = \int_{-\pi/2^J}^{\pi/2^J} f(\lambda)\, d\lambda$. In the case when $f$ has a pole at 0 instead of approximating the last integral it is preferable to use $v_J = \mathrm{Var}(X(n)) - 2\int_{\pi/2^J}^{\pi} f(\lambda)\, d\lambda$. Simulations in McCoy and Walden (1996) indicate that the above method performs comparably with exact but much more time consuming time domain methods.

As an aside note that in view of the previous discussion regressing $\mathbb{E}d_{jk}^2$ on $\log \lambda_j$ is tantamount to regressing $\log f(\lambda_j)$ on $\log \lambda_j$. When variants of the periodogram are used to estimate the spectral density and $\lambda_j$s are replaced by the Fourier frequencies $\lambda_j = 2\pi j/n$ one arrives at usual frequency domain estimates of $\gamma$ (see Beran, 1994).

## 6. Conclusion

In the paper we have studied stochastic properties of the wavelet and the approximation coefficients for a time series indexed by either continuous or discrete time. Starting with an analysis of the continuous time series we have discussed two approaches to extend the wavelet analysis to the discrete time series. The first one relies on embedding in the continuous time series whereas the second is a direct application of the pyramidal algorithm to the time series sequence. In particular, the second approach yields the similar results to results for continuous time due to the fact that recursive relations for the spectral densities stated in Proposition 3.10 hold in both cases. Special attention has been paid to the long-memory case and it is shown that if the considered wavelet has a sufficient number of vanishing moments the covariance of the wavelet coefficients decays arbitrarily quickly. Such property is essential for one of the methods of time series generation with the given spectral density. Moreover, it greatly facilitates studying properties of estimators of the Hurst coefficient based on the wavelet coefficients as it allows for treating them as approximately independent random variables.

## Acknowledgment

Constructive suggestions of two anonymous referees which helped to improve the previous version of the manuscript are gratefully acknowledged.

## References

Abry, P., Gonçalves, P. and Flandrin, P. (1995) Wavelets, spectrum estimation, 1/f processes. *Lecture Notes in Statistics* **103**, Springer, Berlin, 15-30.

Abry, P. and Veitch, D. (1998) Wavelet analysis of long-range dependent traffic. *IEEE Trans. Inform. Theory* **44**, 2-15.

Abry, P., Flandrin, P., Taqqu, M.S. and Veitch, D. (2002) Self-similarity and long-range dependence through the wavelet lens. In: P. Doukhan, G. Oppenheim, M. Taqqu, eds., *Theory and Applications of Long-range Dependence*. Birkhäuser, Boston.

Bardet, J.M.., Lang,G., Moulines, M. and Soulier, P. (2000) Wavelet estimator for long-range dependent processes. *Statistical Inference for Stochastic Processes* **3**, 85-99.

Beran, J. (1994) *Statistics for Long Memory Processes*. Chapman and Hall, New York.

Białasiewicz, J. (2000) *Wavelets and Approximations* (in Polish). WNT, Warsaw.

Brockwell, P.J. and Davis, R.A. (1987) *Time Series: Theory and Methods*. Springer, New York.

Burt, P.J. and Adelson, E.H. (1983) The Laplacian pyramid as a compact image code. *IEEE Trans. Commun.* **COM-31**, 532-540.

Craigmile, P.F. , Guttorp, P. and Percival, D.B. (2005) Wavelet-based estimation for polynomial contaminated fractionally differenced process *IEEE Trans. Signal Processing* **53**, 3151-3161.

Craigmile, P.F. and Percival, D.B. (2005) Asymptotic decorrelation of between-scale wavelet coefficients. *IEEE Trans. Inform. Theory* **51**, 1039-1048.

Davis, R. B. and Harte, D.S. (1987) Tests for Hurst effect. *Biometrika* **74**, 95-101.

Daubechies, I. (1992) *Ten Lectures on Wavelets* **61**, CBMS-NSF Regional Conference Series in Applied Mathematics. SIAM, Philadelphia.

Dijkerman, R.W. and Mazumdar, R.R. (1994) Wavelet representation of stochastic processes. *IEEE Trans. Signal Processing* **42**, 1640-1652.

Flandrin, P. (1992) Wavelet analysis and synthesis of fractional Brownian motion. *IEEE Trans. Inform. Theory* **38**, 910-917.

Gajek, L. and Mielniczuk, J. (1999) Long- and short-range dependent sequences under exponential subordination. *Statistics and Probability Letters* **43**, 113-121.

Gençay, R., Selçuk, F. and Whitcher, B. (2002) *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics.* Academic Press, San Diego.

Gray, H.L., Zhang, N.-F. and Woodward, W.A. (1989) On generalized fractional processes. *J. Time Ser. Anal.* **10**, 233-257.

McCoy, E.T. and Walden, A.T. (1996) Wavelet analysis and synthesis of stationary long-memory processes. *J. Comp. and Graph. Statist.* **5**, 26-56.

Meyer, I. (1993) *Wavelets. Algorithms and Applications.* SIAM, Philadelphia.

Nason, G. P. and von Sachs, R. (1999) Wavelets in time series analysis. *Phil. Trans. R. Soc. London Series A*, **357**, 2511-2526.

Percival , D.B. and Walden, A.T. (2000) *Wavelet Methods for Time Series Analysis.* Cambridge University Press, Cambridge.

Samorodnitsky, G. and Taqqu, M. (1994) *Stable Non-Gaussian Random Processes.* Chapman and Hall, New York.

Taqqu, M., Teverovsky, V. and Willinger, W. (1995) Estimators for long-range dependence: an empirical study. *Fractals* **3**, 785-798.

Tewfik, A.H. and Kim, M. (1992) Correlation structure of discrete wavelet coefficients of fractional Brownian motion. *IEEE Trans. Inform. Theory* **38**, 904-909.

Vanucci, M. and Corradi, F. (1999) Covariance structure of wavelet coefficients. Theory and methods in Bayesian perspective. *J. R. Stat. Soc. B* **61**, 971-986.

Veitch, D. and Abry, P. (1999) A wavelet-based joint estimator of the parameters of long-range dependence. *IEEE Trans. Inform. Theory*, **45**, 878-897.

Veitch, D., Abry, P. and Taqqu, M.S. (2003) On automatic selection of the onset of scaling. *Fractals* **11**, 377-390.

Veitch, D., Taqqu, M.S. and Abry, P. (2000) Meaningful MRA initialisation for discrete time series. *Signal Processing* **80**, 1971-1983.

Vidakovic, B. (1999) *Statistical Modeling by Wavelets.* Wiley, New York.

Walter, G.G. (1994) *Wavelets and Other Orthonormal Systems with Applications.* CRC Press, Boca Raton.

Wojtaszczyk, P. (1997) *A Mathematical Introduction to Wavelets.* Cambridge University Press, Cambridge.

Wornell, G.W. (1995) *Signal Processing and Fractals - a Wavelet-based Approach.* Prentice-Hall, Upper Saddle River, New Jersey.

Yong, C.H. (1974) *Asymptotic Behaviour of Trygonometric Series.* Chinese Univ. Hong Kong.

Zhang, L., Bao, P. and Wu, X. (2004) Wavelet estimation of fractional Brownian motion in a noisy environment. *IEEE Trans. Inform. Theory* **50**, 2194-2200.

Zygmund, A. (1959) *Trigonometric Series*, 2nd ed., Cambridge University Press, Cambridge.