Wavelet Packet Analysis of Variance of Time Series: Methods and Applications

August 3, 2002

Abstract

In this paper we extend the definition of wavelet variance to wavelet packets. We also show how some of the statistical properties of the wavelet variance nicely extend to wavelet packets. Specifically, the wavelet packet variance estimator is still unbiased and asymptotically Gaussian. We then investigate the use of the wavelet packet variance in the analysis of time series with non-constant variance and adapt to wavelet packets the ICSS algorithm for the location of variance change-points. The employment of wavelet packets makes the algorithm more flexible than when standard wavelets are used. Wavelet packets, in fact, induce a finer partition of the frequency domain of the process generating the data, therefore allowing greater decorrelation properties. We show this on simulated data and on a benchmark time series. We also describe an application to time series of crack widths on the Brunelleschi Dome of Santa Maria del Fiore Cathedral in Florence.

Running title: Wavelet Variance with Packets.

Key Words: Change-Points Detection, Crack Widths, Undecimated Transforms, Wavelet Variance, Wavelet Packets.

1 Introduction

The wavelet variance was first introduced by Percival (1995) as a tool for the decomposition of the variance of a time series into different components, each as-
sociated to a different scale (or resolution level). Serroukh et al. (2000) proved that the wavelet variance estimator is asymptotically Gaussian in the case of various nonlinear, non-Gaussian and stationary or locally stationary processes.

Given a time series \( \{Y_t, 0 \leq t \leq N - 1\} \), one of the main conditions for the wavelet variance estimator to be unbiased is that the process \( \{Y_t\} \) must be \( I(d) \), that is its \( d \)-th order backward difference must be a stationary process. However, with real time series, it is not uncommon to encounter departures from this assumption. For instance, in physical sciences, atmospheric variables often show increasing variability in certain months of the year. In such cases, one possible approach is to use testing procedure to identify variance change-points, therefore locating intervals where the variance can be considered constant. Wavelet variance values can then be computed in each of these intervals. Inclán and Tiao (1994) proposed the ICSS algorithm to identify multiple change-points in the variance of a series of independent observations. Whitcher et al. (2000, 2002), applied the ICSS algorithm to coefficients from standard discrete wavelet transforms. However, since the procedure requires independent observations, they restricted themselves to data from long-memory processes. Decorrelation properties of the wavelet transforms for such processes are, in fact, well studied and allowed them to assume wavelet coefficients as approximately uncorrelated.

In this paper we explore some new techniques, based on wavelet packet transforms. We first define a wavelet variance estimator based on wavelet packets and show how statistical properties of the wavelet variance easily extend to packets. We then investigate the use of the wavelet packet variance in the analysis of time series with non-constant variance via the ICSS algorithm. We show how the employment of wavelet packets makes this algorithm more flexible than when standard wavelets are used. Wavelet packets, in fact, induce a finer partition of the frequency do-
main of the process generating the data, allowing greater decorrelation properties. We exploit performances of the proposed methods on simulated data, with both single and multiple change-points, by computing empirical size and power of the testing procedure. We apply the method to a time series of subtidal coastal sea levels previously analysed with wavelet variance. We also describe an application to time series of crack widths on the Brunelleschi Dome of Santa Maria del Fiore Cathedral in Florence. In the examples we use the classical Ljung-Box test for autocorrelation, Box et al. (1994), to exploit correlation at the different packet levels. Examples show the great potential of wavelet packet variance techniques as exploratory tools for time series analysis.

The paper is organized as follows: We begin with a review in Section 2 of the most important concepts regarding standard wavelet transforms and wavelet variance. In Section 3 we briefly introduce discrete wavelet packet transforms and define the wavelet packet variance as a generalization of the wavelet variance. In Section 4 we outline the ICSS algorithm applied to wavelet packet variance. We report simulation studies and real examples in Section 5 and concluding remarks in Section 6. Theoretical results are in the Appendix.

2 Preliminaries

2.1 The maximal overlap discrete wavelet transform

The maximal overlap wavelet transform (MODWT) of Percival and Walden (2000) is basically an undecimated version of the discrete wavelet transform (DWT) of Mallat (1989). The level- \( j \) MODWT of a time series \( \{ Y_t, 0 \leq t \leq N - 1 \} \) is defined
using circular linear filtering as
\[
\tilde{W}_{j,t} = \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} Y_{(t-l)\text{mod}N}, \quad 0 \leq t \leq N - 1
\] (1)
with \(\tilde{h}_{j,l} = h_{j,l}/2^{j/2}\) and \(\{h_{j,l}\}\) representing a level-\(j\) wavelet filter of length \(L_j = (2^j - 1)(L - 1) + 1\). Filter coefficients vary according to the wavelet family. Here we are concerned with Daubechies (1992) wavelets, which have compact support, implying filters with a finite number of nonzero coefficients. Coefficients \(\{\tilde{W}_{j,t}\}\) represent differences between generalized averages of the time series \(\{Y_t\}\) on a scale \(\tau_j = 2^{j-1}\) (or level \(j\)). Undecimated transforms with strong analogies with the MODWT are defined in Shensa (1992), Coifman and Donoho (1995), Nason and Silverman (1995).

### 2.2 The wavelet variance

Let \(\{Y_t, t \in \mathbb{Z}\}\) be a discrete parameter real-valued stochastic process. By filtering \(\{Y_t\}\) with a level-\(j\) MODWT wavelet filter of length \(L_j\) we obtain the stochastic process
\[
\tilde{W}_{j,t} = \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} Y_{t-l}, \quad t \in \mathbb{Z}.
\] (2)

The time dependent wavelet variance of \(\{Y_t\}\) at level \(j\) is then defined as
\[
\nu_Y^2(j) = \text{var}\{\tilde{W}_{j,t}\}.
\] (3)

In practice an estimate is obtained by assuming that the variance of \(\{\tilde{W}_{j,t}\}\) is constant over time, therefore defining
\[
\nu_Y^2(j) = \text{var}\{\tilde{W}_{j,t}\},
\] (4)
and an unbiased estimate is constructed using the MODWT coefficients of equation (1) that are not affected by the modulus operation as

$$\hat{\nu}_Y^2(j) = \frac{1}{M_j(N)} \sum_{t=L_j-1}^{N-1} \hat{W}_{j,t}^2,$$

with $M_j(N) = N - L_j + 1$. It is possible to prove that the asymptotic distribution of $\hat{\nu}_Y^2(j)$ is Gaussian, a result that allows the formulation of confidence intervals for the estimate. See Percival (1995) and Serroukh et al. (2000) for methods and proofs.

3 Packet variance

3.1 The discrete wavelet packet transform

The discrete wavelet packet transform (DWPT) of a time series $\{Y_t\}$ is a generalization of the DWT. At the first level of the transform a low and a high filter are applied to $\{Y_t\}$ to obtain, after subsampling, $\{W_{1,0,t}\}$, corresponding to the frequency band $[0, 1/4]$, and $\{W_{0,1,t}\}$, corresponding to the frequency band $[1/4, 1/2]$. At level $j$ the same steps are repeated by filtering and subsampling the low passed and high passed coefficients obtained at the previous level. From a computational point of view the DWPT can be readily computed using a very simple modification of the pyramid algorithm introduced by Mallat (1989) to compute the DWT. Details are given in Wickerhauser (1994).

The maximal overlap discrete wavelet packet transform (MODWPT) is basically an undecimated version of the DWPT, introduced by Walden and Contreras Cristan (1998). As a filtering of the original time series it can be written as

$$\tilde{W}_{j,n,t} = \sum_{l=0}^{L_j-1} \tilde{f}_{j,n,l} Y_{(t-l)\bmod N}, 0 \leq t \leq N - 1$$

(6)
for \( n = 0, \ldots, 2^j - 1 \), where

\[
\tilde{f}_{j,n,l} = \sum_{k=0}^{L_j-1} \tilde{f}_{n,k} \tilde{f}_{j-1,\lfloor n/2 \rfloor - 2^{j-1} k} , 0 \leq l \leq L_j - 1
\]  

(7)

with

\[
\tilde{f}_{n,l} = \begin{cases} 
\tilde{g}_l & \text{if } n \mod 4 = 0 \text{ or } 3 \\
\tilde{h}_l & \text{if } n \mod 4 = 1 \text{ or } 2
\end{cases}
\]

(8)

and \( \tilde{g}_l = (-1)^{l+1} \tilde{h}_{L-l-1} \), and such that \( \{ \tilde{f}_{1,0,l} = \tilde{g}_l, 0 \leq l \leq L \} \) and \( \{ \tilde{f}_{1,1,l} = \tilde{h}_l, 0 \leq l \leq L \} \). For additional details see Walden and Contreras Cristan (1998).

### 3.2 The wavelet packet variance

Following the same approach described for the wavelet variance we now introduce the wavelet packet variance based on the undecimated discrete wavelet packet transform (MODWPT). Given \( \{ Y_t, t \in \mathbb{Z} \} \) a discrete parameter real-valued stochastic process, we therefore define the \([j,n]\) packet variance \( \nu_{Y,t}^2(j,n) \) as the variance of \( \tilde{W}_{j,n,t} \), if it exists and it is finite,

\[
\nu_{Y,t}^2(j,n) = \text{var}\{ \tilde{W}_{j,n,t} \}
\]

(9)

with

\[
\tilde{W}_{j,n,t} = \sum_{l=0}^{L_j-1} \tilde{f}_{j,n,l} X_{t-l} , n > 0 , \ t \in \mathbb{Z}.
\]

(10)

Assuming a process with constant variance, the packet variance is not time dependent and an unbiased estimates of \( \nu_{Y,t}^2(j,n) \) can be computed using the MODWPT coefficients in equation (6) that do not involve the use of circularity as

\[
\hat{\nu}_{Y,t}^2(j,n) = \frac{1}{M_j(N)} \sum_{t=L_j-1}^{N-1} \tilde{W}_{j,n,t}^2.
\]

(11)
It can be shown that the wavelet packet variance estimator is unbiased and asymptotically Gaussian. Proofs of these results can be readily obtained along the lines of those proved for the case of the wavelet variance and are briefly sketched in the Appendix.

4 Testing for variance changes

Both the estimator (5) of the wavelet variance and the one (11) of the packet variance are based on the hypothesis that the variance of the underlying stochastic process is constant. When this is not the case, testing procedures can be used to identify variance change-points, therefore locating intervals where the variance can be considered constant. Wavelet variance values can then be computed in each of these intervals.

The problem of identifying multiple change-points in the variance of a sequence of independent observations has been addressed by several authors. Inclán and Tiao (1994) proposed a procedure based on an iterated cumulative sum of squares (ICSS) algorithm. Whitcher et al. (2000, 2002) applied the ICSS procedure to wavelet coefficients obtained from the DWT and performed extensive simulation studies for both single and multiple change-points. Since the procedure requires independent coefficients, they assumed data from long memory processes, specifically stationary Gaussian fractionally differenced. Decorrelation properties of the DWT for such processes are, in fact, well studied, see for example Tewfik and Kim (1992), and allow to assume wavelet coefficients at a given level as approximately uncorrelated.
4.1 Detection and location of multiple variance changes via wavelet packets

We have investigated the use of the ICSS algorithm with wavelet packet coefficients. The adaptation of the procedure to packets is pretty straightforward. The redundancy of wavelet packets, however, implies greater flexibility and better decorrelation properties than the DWT, therefore allowing the algorithm to perform well for a wider class of processes than the long-memory. Intuitively, at level \( j \) a wavelet packet transform partitions the frequency interval \([0, 1/2]\) into \( 2^j \) equal intervals of the form

\[
\left[ \frac{n}{2^j+1}, \frac{n+1}{2^j+1} \right], \quad n = 0, 1, \ldots, 2^j - 1
\]

and we can therefore expect that a process with a spectral density function relatively flat in any of the intervals of the partition would result in approximately uncorrelated coefficients at the corresponding packet. The standard wavelet transform, on the other hand, produces at level \( j \) a single set of coefficients, associated with the single interval \([1/2^{j+1}, 1/2^j]\), and in this sense it is less flexible.

Before reporting on examples let us briefly summarize the test procedure of Inclán and Tiao and discuss its adaptation to wavelet packet coefficients. Let \( \{Y_t, 0 \leq t \leq N - 1\} \) be a finite realization of a sequence of uncorrelated random variables with zero means and variances \( \sigma^2_0, \sigma^2_1, \ldots, \sigma^2_{N-1} \). We want to test the hypothesis \( H_0 : \sigma^2_0 = \ldots = \sigma^2_{N-1} \) against the alternative \( H_1 : \sigma^2_0 = \ldots = \sigma^2_k \neq \sigma^2_{k+1} = \ldots = \sigma^2_{N-1} \). The test uses a normalized cumulative sum of squared test statistic \( D = \max(D^+, D^-) \) where

\[
D^+ = \max_{0 \leq k \leq N-2} \left( \frac{k+1}{N-1} - P_k \right), \quad D^- = \max_{0 \leq k \leq N-2} \left( P_k - \frac{k}{N-1} \right),
\]

and \( P_k = \frac{1}{k} \sum_{j=0}^{N-1} Y^2_j / \left( \sum_{j=0}^{N-1} Y^2_j \right) \). When testing for multiple change-points the ICSS
algorithm iteratively computes $D$ on subseries obtained from $\{Y_t\}$. Critical levels for $D$ under the null hypothesis can be obtained via Monte Carlo simulation. Inclàn and Tiao also proved that the asymptotic distribution of the statistic is that one of a Brownian bridge. This allows to perform the test using an asymptotic approximation for the critical values of $D$ when the sample size is at least 128.

We compute the test statistic $D$ using the DWPT wavelet packet coefficients, after correcting for phase effects and discarding the coefficients that involve the use of circularity. Phase effects are fully investigated by Wickerhauser (1994) and Walden and Contreras Cristan (1998). The location of the variance change-points is estimated using the undecimated packet transform as

$$\hat{k} = \text{argmax}(D)$$

with $D$ computed on the MODWPT coefficients.

5 Applications

In this section we show possible situations where estimation of variance change-points requires the use of the DWPT based procedure. We then discuss an application to a time series of subtidal coastal sea levels previously investigated by Percival and Mofjeld (1997) with the DWT. We conclude with the analysis of a time series of crack widths on the Brunelleschi Dome in Florence.

In the sequel we indicate with $D(L)$ and $\text{LA}(L)$ extremal phase and least asymmetric wavelet filters of length $L$, respectively, see Daubechies (1992). Whitcher et al. (2000, 2002) performed extensive simulation studies, using fractionally differenced processes only, with three different wavelet families, Haar, $D(4)$ and $\text{LA}(8)$. They found that power and size of the ICSS algorithm decrease with the wavelet level and recommended not to use levels greater than 3. They also found very little
differences in power and size among the different wavelet filters they used. In our study we investigated performances of the ICSS procedure on wavelet coefficients from different processes than the long-memory and made use of the Ljung-Box test for the presence of autocorrelation, see for example Box et al. (1994), to test whether the packet transform produces uncorrelated coefficients at a given packet. We report on results with the filter that best decorrelates the data.

5.1 Simulation study

Let \( \{X_t\} \) be an ARMA(2,2) stochastic process defined as

\[
(1 - 0.7B + 0.2B^2)X_t = (1 + 0.9B + 0.6B^2)\epsilon_t
\]

where \( B \) is the backward shift operator, \( BX_t = X_{t-1} \). An estimate of the spectral density function (SDF) of this process is shown in Figure 1. The SDF is quite variable and we cannot expect to obtain time series of uncorrelated DWT wavelet coefficients. This is confirmed by the autocorrelation functions of the coefficients at the first two DWT levels shown in Figure 2 and by the Ljung-Box test, in Table 1, performed with a 5% significance level at lags 10, 20 and 30, using least asymmetric wavelets with filter length 8, Daubechies (1992). Better results are instead obtained with wavelet packets, see again Figure 2 and Table 1, and indeed coefficients at packet \([2,2]\) can be considered uncorrelated.

In order to better appreciate the consequences of having uncorrelated wavelet coefficients, let us now examine empirical size and power of the ICSS algorithm for both DWT and DWPT. Figure 3 shows rejection rates obtained by simulating, for each \( N = 2^j, 5 \leq j \leq 11 \), 10,000 realizations of process (15), with constant variance, and performing the test using Monte Carlo critical values with significance levels \( \alpha = 0.01 \) and \( \alpha = 0.05 \). When using DWT coefficients the presence of
correlation determines an over rejection of the (true) null hypothesis of constant variance. At DWT levels 1 and 2 rejection rates are in fact well above 10% with \( \alpha = 0.05 \) (see the lines marked with squares and diamonds in the plot on the left) and above 5% with \( \alpha = 0.01 \) when \( N \geq 128 \) (plot on the right). When, instead, the \( D \) statistic is computed using packet \([2,2]\) DWPT coefficients rejection rates are very close to the values of 5% and 1%.

Results from the study of the empirical power of the ICSS procedure are shown in Table 2. They were obtained with 10,000 replicates of the ICSS algorithm on time series with fixed length \( N = 2048 \) and one variance change-point at \( k = 1024 \). The parameter \( \Delta \) indicates the ratio of the two variance values used to simulate the change-point. When using DWT coefficients, due to the presence of correlation among the coefficients, the procedure tends to overestimate the number of variance change-points. For example, if the ratio of the two variances is \( \Delta = 3 \), two or more change-points are identified in 21.6% of the replicates when using level 1 DWT coefficients. The rejection rates obtained with packet \([2,2]\) DWPT coefficients are, instead, always around 90%, and only slightly worse when \( \Delta = 1.5 \).

We also exploited performances of the procedure in the detection of multiple change-points by using a more complicated process given by the summation of a sinusoidal component with frequency 1/1460, a sinusoidal component with frequency 1/4 and an ARMA(4,4) model

\[
(1 - 0.5B + 0.4B^2 - 0.1B^3 - 0.2B^4)X_t = (1 - 0.2B + 0.1B^4)\varepsilon_t.
\]  

This model mimics the behavior of measurements of crack widths that we will later analyse. We used \( N = 5840 \) and 7 change-points at \( k_1 = 900, k_2 = 1460, k_3 = 2360, k_4 = 2920, k_5 = 3820, k_6 = 4380, k_7 = 5280 \). The ratio between variances was set to \( \Delta = 10 \).

Rejection rates with Haar wavelets for the null hypothesis of no change-points
are shown in Table 3. We also show the estimated locations of the variance change-points in Figures 4 and 5. Level 1 DWT coefficients are heavily autocorrelated and this causes change-points to be overestimated. Eight or more change-points are identified in 41% of the replicates. At packet [3,5], instead, the algorithm correctly identifies seven change-points in 73% of the replicates and there is only a very small tendency to underestimate the number of change-points. Also, the estimates obtained from the level 1 DWT are more spread out around the true change-point locations, as it can be seen from Figures 4 and 5.

5.2 Subtidal sea levels

Percival and Mofjeld (1997) studied a time series of subtidal coastal sea levels, measured in centimeters, at Crescent City (CA). The measurements are transmitted by a permanent tide gauge every 6 minutes. The data, shown in the top plot of Figure 7, are low-passed and subsampled every 1/2 day. They range from the beginning of 1980 to the end of 1991, for a total of $N = 8746$ observations. Percival and Mofjeld computed a time-dependent wavelet variance assuming that, in time segments including a given number of consecutive observations, the wavelet coefficients are a portion of a realization of a process with constant variance. They concluded that the variance of the series is high during the winter months and low during the summer months.

Our analysis revealed that packet [2,3] Haar coefficients are uncorrelated, as it appears from the Ljung-Box test statistic shown in Table 4 and from the autocorrelation functions shown in Figure 6. DWT coefficients at levels 1 and 2 resulted, instead, autocorrelated. The ICSS algorithm applied to packet [2,3] coefficients detected 36 variance change-points, therefore partitioning the time series in 37 intervals where the variance can be considered constant. The vertical lines in Figure
7 indicate the estimated locations of the variance change-points. The middle plot shows the level [2,3] MODWPT coefficients; the bottom plot shows estimates of the packet [2,3] wavelet variance in the intervals defined by the variance change-points, together with 95% confidence intervals. The ICSS procedure nicely isolates intervals with similar variance, it clearly shows that the variance is periodic, with a fairly regular evolution, and that it is highest during the winter months, generally from November until April.

5.3 Crack widths in the Brunelleschi Dome

The dome of Santa Maria del Fiore Cathedral was designed by Filippo Brunelleschi, who also directed its construction (1434-1472). The structure includes an internal thick dome, with a structural support function, and an external thin one, whose function is to protect from atmospheric agents, linked by several joining elements located at the edges between adjacent webs and inside the webs themselves. The first cracks appeared soon after the construction was completed, the main cause being the dome itself, its weight and the insufficient resistance of the tambour, see Chiarugi et al. (1983). If we assign numbers to the eight webs of the dome, from 1 to 8, with 1 being the web in front of the nave, at present time the main cracks are in the even webs. They start from the tambour and stretch to the higher part of the dome, passing through both the internal and external dome. Several control devices have been installed to monitor the evolution of the cracks. The most recent is a large digital monitoring system installed on January 8, 1988, that includes 166 instruments recording four measurements per day. Measured variables are temperatures (in the air and in the masonry) and variations of crack widths relative to the day in which the monitoring system began working.

In this section we apply wavelet methods for the analysis of variance to time
series of variations of crack widths measured by instruments located in the inner
dome of web number 4. We report here only the analysis of the instrument (df406)
located on the main crack of the web, in the inner part of the inner dome. Data are
shown in the top plot of Figure 8. Measurements are in millimeters. A negative
variation means that the crack is getting wider.

With Haar wavelets, the Ljung-Box test identified wavelet packet coefficients
at packet [3, 5] as a time series of uncorrelated DWPT coefficients. The ICSS algo-
rithm then located 13 change-points, defining 14 intervals with constant variance
in which an estimate of the wavelet packet variance can be computed via (11).
This is shown in Figure 8. The intraday variance appears seasonally dependent. It
seems to increase during spring and to achieve its maxima in the summer months,
while it is lowest in the winter months. Also, the variance tends to achieve its max-
ima when the crack gets close to the position of minimum width, so that greater
variance corresponds to the phase when the crack is closing. The dates correspond-
ing to the variance change-points are: August 29 1989, January 30 1990, August

We performed similar analyses using data from the instruments installed in
the webs affected by the main cracks. As a common feature, the instruments
positioned in the inner dome always show seasonal behavior and high values of
the wavelet variance. For those in the outer dome, instead, the wavelet variance
exhibits only very small variations and no seasonal dependence. This may be
partially explained by the fact that the thick inner dome has a structural function,
supporting the weight of the entire monument, while the outer dome was designed
by Filippo Brunelleschi as a simple shield against atmospheric agents.
Results here reported are part of a preliminary exploratory analysis of the monitoring system of the Brunelleschi dome. This analysis has motivated a modeling approach to the data via ARMA-GARCH models with seasonal components to describe the dynamics of the cracks. The complete study with the results of the modeling analysis will be reported elsewhere.

6 Concluding remarks

In this paper we have extended to wavelet packets recently wavelet approaches to variance estimation for time series. We have defined the wavelet packet variance as a generalization of the wavelet variance of Percival (1995). We have shown that the estimator is still unbiased and asymptotically Gaussian when the underlying process is Gaussian. Serroukh et al. (2000) have generalized some results about the distribution of the wavelet variance estimator to processes that are not necessarily Gaussian and we anticipate that similar extensions may be possible for the wavelet packet variance as well. This is left to future research.

For the case of processes with non constant variance we have adapted to wavelet packets the procedure of Inclán and Tiao (1994) to identify variance change-points. There the test statistic is computed using wavelet packet coefficients, while the location of the variance change-points is estimated using undecimated wavelet packet coefficients. We have shown via simulations how this procedure can be applied to time series coming from a wide class of stochastic processes. Wavelet packets have in fact greater decorrelation abilities than standard transforms, due to their finer partition of the frequency domain. We have studied empirical size and power of the testing procedures. We have also applied wavelet packet techniques to time series collected in a variety of fields, first to a time series of subtidal coastal sea
levels, previously studied by Percival and Mofjeld (1997) using wavelet variance, and then to a time series of crack width variations recorded in the Brunelleschi Dome in Florence. In the latter example, our analysis has revealed some interesting aspects regarding the dynamic of crack evolutions. Examples have highlighted the great potential of wavelet packet variance techniques as exploratory tools for time series analysis.

Appendix

In this appendix we give some theoretical results to support our statement that the estimator of the packet variance (11) is unbiased and asymptotically Gaussian. These results are largely based on those presented by Percival (1995) and Percival and Walden (2000) for the wavelet variance (5). We give here brief sketches of their proofs. These are mainly based on the fact that wavelet packet filters \( \tilde{f}_{j,n,l} \) can be computed from wavelet and scaling filters through equations (8) and (7) and that the squared gain function of wavelet packet filters can be obtained as a product of those of scaling and wavelet filters. Complete proofs would require introducing tedious elements of notation. We refer interested readers to Gabbanini (2002). Proposition 5 below gives an unbiased estimator of the packet variance \( \nu^2_X(j,n) \). Proposition 6 establishes that this estimator is asymptotically Gaussian with Gaussian data. We remind the reader that the transfer function of a filter \( \tilde{f}_{j,n,l} \) is defined as

\[
\tilde{F}_{j,n}(f) = \sum_{l=0}^{L_{j,1}} \tilde{f}_{j,n,l} e^{-i2\pi f l}.
\]

**Proposition 1** Let \( \{X_t\} \) be a stationary process with constant mean \( E\{X_t\} = \mu_X \). Then \( E\{\tilde{W}_{j,n,t}\} = 0 \ \forall n > 0 \).
Proof: From the definition of $\{\tilde{W}_{j,n,t}\}$ we have

\[
E\{\tilde{W}_{j,n,t}\} = \sum_{l=0}^{L_j-1} \tilde{f}_{j,n,t} \cdot E\{X_{t-l}\} = \mu_X \cdot \sum_{l=0}^{L_j-1} \tilde{f}_{j,n,t}.
\]

The result is obvious if $\mu_X = 0$. Otherwise it can be proven by using

\[
\tilde{F}_{j,n}(0) = \sum_{l=0}^{L_j-1} \tilde{f}_{j,n,t} = 0. \quad \square
\]

Proposition 2 Let $\{X_t\}$ be a stationary process and $J_0 \geq 1$,

\[
\text{var}\{X_t\} = \text{var}\{\tilde{W}_{J_0,0,t}\} + \sum_{n=1}^{2^J_0-1} \nu_X^2 (j,n).
\]

Proof: The result is obvious for $J_0 = 1$, as it follows from the properties of the wavelet variance. Let $\tilde{F}_{j,n}(f) = |\tilde{F}_{j,n}(f)|^2$ denote the squared gain function of the filter $\{\tilde{f}_{j,n,l}\}$. With $\{\tilde{W}_{0,0,t}\} \equiv \{X_t\}$, we have for all $j > 1$

\[
\text{var}\{\tilde{W}_{j-1,n,t}\} = \int \tilde{F}_{j-1,n}(f) S_X(f) df.
\]

Given the fundamental property of wavelet filters $\tilde{G}(2^j f) + \tilde{H}(2^j f) = 1$ we also have

\[
\text{var}\{\tilde{W}_{j-1,n,t}\} = \int \left[ \tilde{G}(2^{j-1} f) + \tilde{H}(2^{j-1} f) \right] \tilde{F}_{j-1,n}(f) S_X(f) df.
\]

From this one it is straightforward to derive

\[
\text{var}\{\tilde{W}_{j-1,n,t}\} = \text{var}\{\tilde{W}_{j,2n,t}\} + \text{var}\{\tilde{W}_{j,2n+1,t}\} \quad n = 0, 1, \ldots, 2^{j-1} - 1.
\]

The result follows from successive substitutions. \quad \square

Proposition 3 Let $\{X_t\}$ be an $I(d)$ stochastic process and suppose that $\{\tilde{f}_{j,n,t}\}$ is obtained from a $D(L)$ or a $LA(L)$ wavelet filter of length $L \geq 2d$. Then, if $n > 0$, $\{\tilde{W}_{j,n,t}\}$ is a stationary process.
Proof: If $j = 1$ notice that $\{\tilde{W}_{1,1,t}\} \equiv \{\tilde{W}_{1,t}\}$. If $j > 1$ let $\{\tilde{h}_t\}$ be a Daubechies D(L) or LA(L) wavelet filter and let $\tilde{H}(f) = |\tilde{H}(f)|^2$ be its squared gain function. As in Daubechies (1992) the squared gain function can be factorized as

$$\tilde{H}(f) = {\mathcal{D}}^{L/2}(f) \cdot \tilde{A}_L(f),$$

with

$$\tilde{A}_L(f) = \frac{1}{2L} \sum_{l=0}^{\frac{L}{2}-1} \left( \frac{L}{2} - 1 + l \right) \cos^{2l}(\pi f).$$

${\mathcal{D}}^{L/2}(f)$ corresponds to a backward difference operator with order $L/2$, whereas $\tilde{A}_L(f)$ corresponds to a finite length filter. Using this result, if $n > 0$ we can write the squared gain function of $\{\tilde{f}_{j,n,t}\}$ as the product

$$\tilde{F}_{j,n}(f) = {\mathcal{D}}^d(f) \cdot {\mathcal{B}}_{j,n,m}(f),$$

where $\mathcal{D}^d(f)$ is the squared gain function of a backward difference operator of order $d$, and $\mathcal{B}_{j,n,m}(f)$ is a squared gain function that corresponds to a finite length filter (see Gabbanini (2002) for details on how to derive its expression). This allows to state that $\tilde{W}_{j,n,t} = \sum_{l=0}^{L-1} \tilde{f}_{j,n,l} X_{t-l}$ is a stationary process, given that $\{X_t\}$ is an $I(d)$ process. \(\square\)

Proposition 4 Let $\{X_t\}$ be an $I(d)$ stochastic process and suppose that $\{\tilde{f}_{j,n,t}\}$ is obtained from a $D(L)$ or $LA(L)$ wavelet filter of length $L$.

- If $L > 2d$, $E\{\tilde{W}_{j,n,t}\} = 0$, $\forall n > 0$;
- if $L = 2d$ and $E\{X_t\} = 0$, $E\{\tilde{W}_{j,n,t}\} = 0$, $\forall n > 0$;
- if $L = 2d$ and $E\{X_t\} \neq 0$, $\exists n : E\{\tilde{W}_{j,n,t}\} \neq 0$.  

18
Proof: Let \( \{Y_t\} \) be a stochastic process obtained by filtering \( \{X_t\} \) with a backward difference filter corresponding to the squared gain function \( D^d(f) \) as in proposition 3. Notice that

\[
E\{W_{j,n,t}\} = E\{Y_t\} \cdot B_{j,n,m}(0).
\]

If \( E\{X_t\} = 0 \), then clearly \( E\{Y_t\} = 0 \). Otherwise one can use the factorization of \( \tilde{F}_{j,n}(f) \) given in proposition 3 to prove the result.  

Proposition 5 Let \( \{X_t, 0 \leq t \leq N - 1\} \) be a time series from an \( I(d) \) stochastic process and suppose that \( \{\tilde{f}_{j,n,1}\} \) is obtained from a \( D(L) \) or \( LA(L) \) wavelet filter of length \( L \) such that \( E(\tilde{W}_{j,n,t}) = 0 \). Then

\[
\hat{\nu}^2_X(j, n) = \frac{1}{M_j} \sum_{t=L_j-1}^{N-1} \tilde{W}_{j,n,t}^2
\]

is an unbiased estimator for the packet variance \( \nu^2_X(j, n) \).

Proof: Straightforward given the definition of the packet variance (11) and the results of the previous propositions.  

Some results are needed in order to prove Proposition 6.

Lemma 1 Let \( \{\zeta^2_1, \zeta^2_2, \ldots, \zeta^2_n\} \) be a realization of a strictly stationary stochastic process. Suppose that \( \delta > 0 \) exists such that \( E(|\zeta|^4+\delta) < \infty \). Suppose that the spectral density function \( S_{\xi^2}(f) \) of the process \( \{\zeta^2_t\} \) is continuous on \( [-\frac{1}{2}, \frac{1}{2}] \) and strictly positive. Then \( \hat{\nu}^2_\zeta = \frac{1}{n} \sum_{t=1}^{n} \zeta^2_t \) is asymptotically Gaussian with mean \( \nu^2_\zeta = E(\zeta^2_t) \) and variance \( \frac{S_{\xi^2}(0)}{n} \).

A proof can be found in Percival (1995).  

Lemma 2 If \( \{G_t\} \) is a stationary Gaussian process whose spectral density function is \( S_G(f) \), then \( \{G_t^2\} \) is also stationary with spectral density function

\[
S_{G^2}(f) = 2 \int_{-1/2}^{1/2} S_G(f')S_G(f - f')df'.
\]

A proof is given in Hannan (1970), page 83. \( \square \)

Lemma 3 If \( \{G_t\} \) is a stationary Gaussian process whose spectral density function is \( S_G(f) \). If \( S_G \) is square integrable and strictly positive a.e. in \([-\frac{1}{2}, \frac{1}{2}]\), then \( S_{G^2} \) is continuous in \([-\frac{1}{2}, \frac{1}{2}]\) and strictly positive.

A proof can be found in Percival (1995). \( \square \)

Proposition 6 Let \( \{X_t, 0 \leq t \leq N - 1\} \) be a time series from a Gaussian I(d) stochastic process. Suppose that \( \{Y_{j,n,t}\} \) is obtained from a D(L) or LA(L) wavelet filter of length \( L \) such that \( E(W_{j,n,t}) = 0 \). If \( S_{W_{j,n}} > 0 \) a.e. in \([-\frac{1}{2}, \frac{1}{2}]\) and if

\[
\int_{-1/2}^{1/2} S_{W_{j,n}}^2(f) df < \infty,
\]

then \( \nu_X^2(j,n) \) is asymptotically Gaussian with mean \( \nu_X^2(j,n) \) and variance

\[
\frac{2}{M_j} \int_{-1/2}^{1/2} S_{W_{j,n}}^2(f) df.
\]

Proof: Proposition 3 ensures that \( \{W_{j,n,t}\} \) is stationary with zero mean. Since it is also Gaussian, \( \{W_{j,n,t}\} \) is strictly stationary. Moreover, \( \{W_{j,n,t}^2\} \) is also strictly stationary and, for lemma 3, its spectral density \( S_{W_{j,n}}^2 \) is continuous and strictly positive. Then, once we have defined \( \hat{\nu}_X(j,n) \) as in equation (11), we can use lemma 1 and lemma 2 to state that \( \hat{\nu}_X(j,n) \) is asymptotically Gaussian with mean \( \nu_X(j,n) \) and variance given by \( \frac{S_{W_{j,n}}^2(0)}{M_j} = \frac{2}{M_j} \int_{-1/2}^{1/2} S_{W_{j,n}}^2(f) df. \) \( \square \)
References


Table 1: ARMA(2,2): Ljung-Box $Q$ test statistic values for level 1 DWT coefficients, level 2 DWT coefficients and packet $[2,2]$ DWPT coefficients, from a realization of process (15) with $N = 2048$. 5% critical values are in parentheses.

<table>
<thead>
<tr>
<th>Level</th>
<th>$Q_{10}$</th>
<th>$Q_{20}$</th>
<th>$Q_{30}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>337.06 (18.31)</td>
<td>349.41 (31.41)</td>
<td>358.81 (43.77)</td>
</tr>
<tr>
<td>2</td>
<td>169.39</td>
<td>174.48</td>
<td>185.57</td>
</tr>
<tr>
<td>2,3</td>
<td><strong>9.29</strong></td>
<td><strong>12.92</strong></td>
<td><strong>16.75</strong></td>
</tr>
</tbody>
</table>


![Figure 1: ARMA(2,2): Spectral density function estimate of process (15).](image)
Figure 2: ARMA(2,2): Autocorrelations of level 1 DWT coefficients, level 2 DWT coefficients and packet [2,2] DWPT coefficients, from a realization of process (15) with \( N = 2048 \).

Figure 3: ARMA(2,2): Empirical size of the ICSS algorithm with level 1 DWT, level 2 DWT and packet [2,2] DWPT coefficients.
Figure 4: Model (16): Locations of variance change-points estimated at level 1 DWT, for $N = 5840$ and seven change-points.

Figure 5: Model (16): Locations of variance change-points estimated at packet $[3,5]$ DWPT, for $N = 5840$ and seven change-points.
Figure 6: Subtidal sea levels: Autocorrelation functions at level 2 DWPT.

Figure 7: Subtidal sea levels: Bars indicate the estimated locations of variance changes. Top plot: data. Middle plot: packet [2,3] MODWPT coefficients. Bottom plot: estimates of packet [2,3] wavelet variance in the intervals defined by the variance change-points, with a 95% c.i.
Table 2: ARMA(2,2): Empirical power of the ICSS algorithm for $N = 2048$ and one variance change-point at $k = 1024$.

<table>
<thead>
<tr>
<th>Level</th>
<th>$\Delta$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>&gt;2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.5</td>
<td>4.2</td>
<td><strong>78.7</strong></td>
<td>13.5</td>
<td>3.6</td>
</tr>
<tr>
<td>2</td>
<td>1.5</td>
<td>18.1</td>
<td><strong>70.8</strong></td>
<td>9.7</td>
<td>1.5</td>
</tr>
<tr>
<td>2.2</td>
<td>1.5</td>
<td>19.4</td>
<td><strong>75.5</strong></td>
<td>4.8</td>
<td>0.3</td>
</tr>
<tr>
<td>1</td>
<td>2.0</td>
<td>0.7</td>
<td><strong>79.5</strong></td>
<td>15.2</td>
<td>4.6</td>
</tr>
<tr>
<td>2</td>
<td>2.0</td>
<td>1.8</td>
<td><strong>80.3</strong></td>
<td>14.7</td>
<td>3.3</td>
</tr>
<tr>
<td>2.2</td>
<td>2.0</td>
<td>0.9</td>
<td><strong>93.2</strong></td>
<td>5.7</td>
<td>0.2</td>
</tr>
<tr>
<td>1</td>
<td>3.0</td>
<td>0.5</td>
<td><strong>77.9</strong></td>
<td>16.6</td>
<td>5.0</td>
</tr>
<tr>
<td>2</td>
<td>3.0</td>
<td>0.5</td>
<td><strong>81.3</strong></td>
<td>15.3</td>
<td>2.9</td>
</tr>
<tr>
<td>2.2</td>
<td>3.0</td>
<td>0.6</td>
<td><strong>92.4</strong></td>
<td>6.4</td>
<td>0.7</td>
</tr>
<tr>
<td>1</td>
<td>10.0</td>
<td>0.7</td>
<td><strong>77.7</strong></td>
<td>16.3</td>
<td>5.4</td>
</tr>
<tr>
<td>2</td>
<td>10.0</td>
<td>0.9</td>
<td><strong>79.9</strong></td>
<td>15.9</td>
<td>3.4</td>
</tr>
<tr>
<td>2.2</td>
<td>10.0</td>
<td>0.8</td>
<td><strong>89.7</strong></td>
<td>8.6</td>
<td>0.9</td>
</tr>
</tbody>
</table>

Table 3: Model (16): Empirical power of the ICSS algorithm for $N = 5840$ and seven variance change-points.

<table>
<thead>
<tr>
<th>Livello</th>
<th>$\Delta$</th>
<th>&lt;6</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>&gt;8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>2</td>
<td>1.5</td>
<td><strong>28.8</strong></td>
<td>26</td>
<td>41.7</td>
</tr>
<tr>
<td>3.5</td>
<td>10</td>
<td>8.3</td>
<td>9.7</td>
<td><strong>73</strong></td>
<td>8.6</td>
<td>0.4</td>
</tr>
</tbody>
</table>

Table 4: Subtidal sea levels: Ljung-Box $Q$ statistic for DWPT coefficients at level 2. Critical values are in parentheses.

<table>
<thead>
<tr>
<th>Level</th>
<th>$Q_{10}$</th>
<th>$Q_{20}$</th>
<th>$Q_{30}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.0</td>
<td>6311.20 (18.31)</td>
<td>9655.61 (31.41)</td>
<td>11309.60 (43.77)</td>
</tr>
<tr>
<td>2.1</td>
<td>128.35</td>
<td>153.85</td>
<td>163.54</td>
</tr>
<tr>
<td>2.2</td>
<td>172.36</td>
<td>202.00</td>
<td>215.13</td>
</tr>
<tr>
<td>2.3</td>
<td>13.76</td>
<td>25.80</td>
<td>39.19</td>
</tr>
</tbody>
</table>
Figure 8: Crack widths in the Brunelleschi Dome: Bars indicate the estimated locations of variance changes. Top plot: data. Middle plot: estimates of packet [3,5] wavelet variance in the intervals defined by the variance change-points, with a 95% C.I. Bottom plot: packet [3,5] MODWPT coefficients.