For SIP Topical Issue: Wavelet spectrum analysis of EIT/SoHO images

V. Delouille (veronique.delouille@oma.be), J. de Patoul (judith.depatoul@oma.be) and J.-F. Hochedez (hochedez@oma.be)
Royal Observatory of Belgium

L. Jacques and J.-P. Antoine
Inst. de Physique Théorique, Université catholique de Louvain

April 13, 2005

Abstract.

The Extreme ultraviolet Imaging Telescope (EIT) of SoHO offers a unique record of the solar atmosphere for its sampling in temperature, field-of-view, resolution, duration, and cadence. To investigate globally and locally its topology and evolution during the solar cycle, we consider a multiscale approach, and more precisely we use the wavelet spectrum.

We present three results among the applications of such a procedure. First, we estimate the typical dimension of the supergranules as seen in the 30.4 nm passband, and we show that the evolution of the characteristic network scale is almost in phase with the solar cycle. Second, we build pertinent time series that give the evolution of the signal energy present in the corona at different scales. We propose a method that detects eruptions and post-flaring activity in EUV image sequences. Third, we introduce a new way to extract Active Regions in EIT images, with perspectives in e.g. long-term irradiance analysis.

Keywords: solar corona, chromospheric network, supergranule, flare, post-flare eruption, wavelet, wavelet spectrum, scale, scale measure

1. Introduction

Since 1990, wavelets and related multiscale representations have been widely applied in signal processing, and in particular in astrophysical data analysis. In this paper, we use the Continuous Wavelet Transform, and more precisely its *wavelet spectrum*, also called *scale measure*. We analyze individual EUV solar images, as well as large time series thereof, such as the ones offered by the EIT/SoHO archive. The application of the scale measure to solar images analysis was first proposed with early results by Hochedez *et al.* (2002a). Other type of wavelet transforms have been previously used in solar physics. For example, Portier-Fozzani *et al.* (2001) use a Discrete Wavelet Transforms and the 'à trous' algorithm. Their images analysis revealed hidden structure such as Coronal Mass Ejection formation and the evolution of complex Active Regions.

Below we present three applications that exploit differently the ability of the scale measure to abridge the information present in solar images. First, we consider observations at 30.4 nm (dominated by He II emission) and show

that they exhibit a characteristic scale. Indeed, the supergranules of the chromospheric network are visible in the He II images (Reeves, 1976; Worden et al., 1999). They have a size distribution which has been well studied (Simon and Leighton, 1964; Hagenaar et al., 1997; Meunier, 2003; Del Moro et al., 2004; DeRosa and Toomre, 2004). The wavelet spectrum detects this structure, in the sense that for the scales close to the range of supergranules sizes, the scale measure has a maximum. It is of interest to see how the characteristic scale, and hence the network cell size, evolves over the whole solar cycle (Meunier, 2003).

Second, we analyze a sequence at 19.5 nm (dominated by Fe XII emission) when the Sun was active in early May 1998. The 1024×1024 pixel size images are recorded every 12 minutes. In quiet periods, the wavelet spectrum can be well fitted with a linear model; there is no characteristic scale (Aletti et al., 2000). However, during flares or post-flare eruptions, the wavelet spectrum changes, and some characteristic scales may appear. Parametric fits are successfully applied, providing scalar time series that recapitulate the activity as seen by EIT in the given passband.

Third, we use a local version of the scale measure to segment the image, and in particular to extract automatically the active regions.

The images analyzed here have been preprocessed using the standard *eit-prep* procedure of the *solar software (ssw)* library. Unless stated otherwise, we use the linear *level 1* images (which are photometrically normalized) as input to our procedures.

This paper is organized as follows. Section 2 gives some basic background on wavelet analysis, and presents the Mexican Hat wavelet used in this work. Section 3 exploit the wavelet spectrum to extract a characteristic scale in the 30.4 nm passband and to monitor its evolution over the solar cycle. In Section 4, we analyze a six day time series at 19.5 nm passband in order to detect flaring events. Finally, we explain in Section 5 how to segment an image using a local version of the scale measure. Section 6 is dedicated to a discussion and prospects for future research.

2. Background

2.1. C (CWT)

Wavelet transforms have proved to be a useful tool for analyzing structures of different sizes in a bidimensional signal. An image is a function $f \in \mathbf{L}_2(\mathbb{R}^2)$ which associates a value at a position. In practice, this function is discretized in pixel and DN. A $\mathbf{L}_1$-normalized wavelet is defined as a function $\psi \in \mathbf{L}_1(\mathbb{R}^2)$

that satisfies the *admissibility condition*:

$$C_\psi = (2\pi)^2 \int_{\mathbb{R}^2} d^2\vec{k} \frac{\left|\widehat{\psi}(\vec{k})\right|^2}{\left\|\vec{k}\right\|^2} \text{ is such that } 0 < C_\psi < \infty \,,$$

where $\widehat{\psi}(.)$ denotes the Fourier transform of ψ, and $\left\|\vec{x}\right\| = (x_1^2 + \ldots + x_n^2)^{1/2}$ is the norm of a vector $\vec{x} = (x_1, \ldots, x_n)$.

If $\psi \in \mathbf{L}_1(\mathbb{R}^2) \cap \mathbf{L}_2(\mathbb{R}^2)$, this condition is equivalent to the fact that ψ is localized around the origin and has zero-average:

$$\int_{\mathbb{R}^2} \psi(\vec{x}) d^2\vec{x} = 0. \tag{1}$$

A family of continuous wavelets is obtained by scaling and translation:

$$\psi_{\vec{b},a}(\vec{x}) = \frac{1}{a^2}\psi\left(\frac{\vec{x} - \vec{b}}{a}\right), \qquad \vec{b} \in \mathbb{R}^2, \quad a \in \mathbb{R} \,,$$

where $\psi_{\vec{b},a}$ is now centered at location $\vec{b}$, and the scaling factor $1/a^2$ ensures the $\mathbf{L}_1-$normalization of $\psi_{\vec{b},a}$.

By convolving each of these functions $\psi_{\vec{b},a}$ with the function f, it is possible to explore f at different scales and locations:

$$\mathcal{W}_f(\vec{b}, a) = \int f(\vec{x}) \psi^*_{\vec{b},a}(\vec{x}) d^2\vec{x} \,, \tag{2}$$

where $*$ denotes the complex conjugation.

Equation (2) defines the Continuous Wavelet Transform (CWT), which acts as a mathematical microscope: the *wavelet coefficients* $\mathcal{W}_f(\vec{b}_0, a_0)$ measures the component of f that is proportional to a_0 and is located in a neighborhood of $\vec{b}_0$. Equation (2) entails that if an image f contains some structures having a size close to the essential support[1] of $\psi_{\vec{b},a_0}$, then $\left|\mathcal{W}_f(\vec{b}, a_0)\right|$ will be large.

The wavelet functions possess $N \geq 1$ vanishing moments if (Antoine et al., 2004)

$$\int_{\mathbb{R}^2} \psi(\vec{x}) x^\alpha y^\beta d^2\vec{x}, \quad \vec{x} = (x, y), 0 \leq \alpha + \beta \leq N.$$

When a wavelet has N vanishing moments, its wavelet coefficients are zero whenever the signal is a polynomial of order $N - 1$. This means that the wavelet transform only detects the sharp features and singularities such as edges or ridges. It remains blind to the smoothest parts of the signal.

[1] The essential support of a function is defined as the support where the function is larger than a given small value.

In practice, the convolution in (2) is computed using the fast Fourier Transform. The discretisation of a function then imposes a minimal scale in order to avoid numerical instabilities. Our first scale is higher than this minimal scale; it is taken equal to one pixel.

2.2. W

Plancherel's formula can be applied to measure the energy of a signal from its CWT:

$$\|f\|_2^2 = \int_{\mathbb{R}^2} \left|f(\vec{x})\right|^2 d^2\vec{x} = \frac{1}{C_\psi} \int_{\mathbb{R}^2 \times R_+^0} \left|\mathcal{W}_f(\vec{b},a)\right|^2 \frac{da}{a^2} d^2\vec{b}\,.$$

The term $\left|\mathcal{W}_f(\vec{b},a)\right|^2$ must be seen as an energy density in the wavelet space. Let Ω be the domain of interest (in our application, it will be either a part or the whole solar disk). Integrating $\left|\mathcal{W}_f(\vec{b},a)\right|^2$ over Ω gives the *scale measure* (Antoine et al., 1998) or *wavelet spectrum*:

$$\mu(a) = \int_\Omega \left|\mathcal{W}_f(\vec{b},a)\right|^2 d^2\vec{b}\,. \tag{3}$$

In Section 4 and 5, we demonstrate that the scale measure allows one to separate different types of activity in the EIT/SoHO images.

It is possible to compute analytically the scale measure for a white noise random process $\{X_{\vec{b}}\}$ having a standard deviation σ. Indeed, by the Wiener-Khintchine theorem (Kittel, 1958; Torrésani, 1997), we know that the wavelet coefficients of such a noisy process have a standard deviation equal to $\left\|\psi_{\vec{b},a}\right\|_{\mathbf{L}_2} \sigma = \|\psi\|_{\mathbf{L}_2} \sigma/a$. Since the integration of the wavelet coefficients over Ω is equal to zero, the wavelet spectrum $\mu(a)$ in (3) can be seen as the variance of the wavelet coefficients. It follows that $\mu(a)$ is proportional to:

$$\mu(a) \simeq \left\|\psi(\vec{x})\right\|_{\mathbf{L}_2}^2 \frac{\sigma^2}{a^2},$$

Hence if an image is essentially white noise, then for any CWT with L_1-normalization, the scale measure $\mu(a)$ follows a power law with index equal to -2.

2.3. M H

Features present in EIT images, such as the chromospheric network in 30.4 nm, the quiet Sun or Active Region (AR) in the solar corona, can be approximated as a superposition of Gaussian peaks of different sizes. For this reason, we use in our subsequent analysis the Mexican Hat (MH) wavelet transform, since it

provides an optimal filter for detection of Gaussian peaks on $1/f$ noise (Sanz et al., 2001). The Mexican Hat wavelet and its Fourier transform are both real and have an analytical expression:

$$\psi(\vec{x}) = \left(2 - \left\|\vec{x}\right\|^2\right) \exp\left(-\left\|\vec{x}\right\|^2/2\right) \quad ; \quad \widehat{\psi}(\vec{k}) = 2\pi \left\|\vec{k}\right\|^2 \exp\left(-\left\|\vec{k}\right\|^2/2\right) \quad (4)$$

The MH wavelet is the second derivative of a Gaussian, and hence possess two vanishing moments (Antoine et al., 1998). It is possible to compute analytically $\mu(a)$ if the signal $I(\vec{x})$ is a Gaussian peak with standard deviation τ centered at $\vec{b} = \vec{0}$:

$$I(\vec{x}) = \frac{A}{2\pi\tau^2} \exp\left(\frac{-\left\|\vec{x}\right\|^2}{2\tau^2}\right) \quad ; \quad \widehat{I}(\vec{\omega}) = \frac{A}{\tau} \exp\left(-\frac{\tau^2 \left\|\vec{\omega}\right\|^2}{2}\right), \qquad (5)$$

where $\widehat{I}(\vec{\omega})$ is the Fourier transform of $I(\vec{x})$. By the convolution theorem we have

$$\mu(a) = \int |\hat{I}(\vec{\omega})\hat{\psi}(a\vec{\omega})|^2 d^2\vec{\omega} = \frac{2\pi a^4 A^2}{\tau^2(a^2+\tau^2)^3}\,. \qquad (6)$$

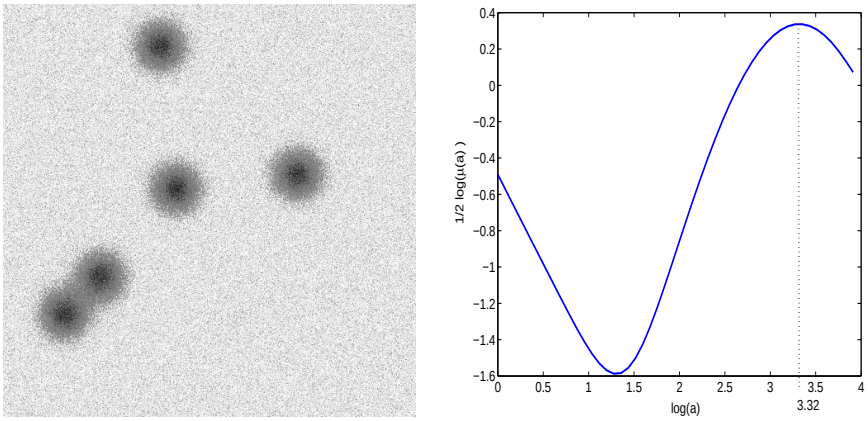
The maximum of $\mu(a)$ is attained for $a_{\max} = \sqrt{2}\tau$: we say that there is a *characteristic scale* at $\sqrt{2}\tau$. Intuitively, this means that the shape of $\psi_{\vec{0},a_{\max}}$ follows as closely as possible the function $I(\vec{x})$.

Since $\mu(a)$ is a squared quantity that often follows a power law as will be shown in our applications, we took the convention of representing $\log(\sqrt{\mu(a)}) = \frac{1}{2}\log\mu(a)$.

Figure 1 illustrates the behavior of $\mu(a)$ on a signal composed of Gaussian peaks having a standard deviation $\tau = 20$ and a unit amplitude A. The signal is also perturbed by Gaussian white noise having a standard deviation $\sigma = 0.01$. It is clear that the scale measure separates the features present at different scales: at small scales, the decreasing slope of $\frac{1}{2}\log(\mu(a))$ represents the white noise background, whereas the maximum at $a = \sqrt{2}\tau$ indicates the presence in the signal of the Gaussian peaks.

3. Time evolution of the chromospheric network size

The 30.4 nm passband of EIT is dominated by He II emission and shows the chromospheric network. In this section, we compute a *characteristic scale* $a_{\max}$ associated to the network, that is, the scale for which the wavelet spectrum $\mu(a)$ is maximum. We indicate how this particular scale can be related to the size of the network cell, and show its evolution over the solar cycle.



(a) Signal (b) Wavelet spectrum

Figure 1. (a) Signal composed of 5 Gaussian peaks ($\tau = 20, A = 1$) on Gaussian white noise with $\sigma = 0.01$ (b) Graph of $\log(\mu(a)^2)$ versus $\log(a)$, where a is the scale and $\mu(a)$ is the wavelet spectrum. At small scales the white noise gives a decreasing wavelet spectrum (with slope close to -1), whereas at larger scale the wavelet spectrum presents a maxima at $\log(\tau\sqrt{2}) = \log(20\sqrt{2}) = 3.32$.

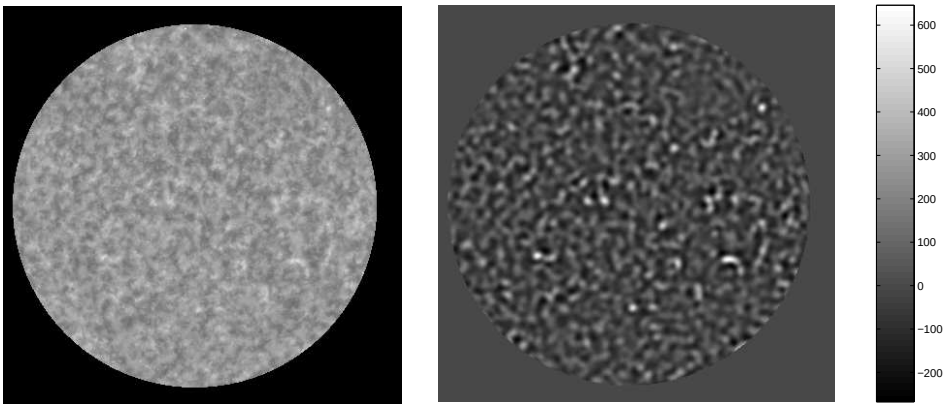
3.1. C EIT 30.4

In this study, the scale measure $\mu(a)$ is computed by integration over a disk Ω centered at the center of the Sun, and having a radius equal to $0.5R_\odot$, $R_\odot$ being the solar radius. This choice of Ω limits the distortion (or shortening) of the network's size due to the projection of the solar sphere on the two-dimensional plane of the sky. We choose the value $0.5R_\odot$ because the projection then induces a bias always smaller than 15%, which is comparable to other approximation errors. Indeed, the observed characteristic scale variation exhibits instantaneous fluctuations of the order of 10%, see Figure 4. These fluctuations result from the method (finite grid of scales a, scale plate variation), and from possible intrinsic solar variability.

The image I_{304} in Figure 2(a) represents the $0.5R_\odot$ solar disk taken on January 2, 1997 in the 30.4 nm channel. The corresponding scale measure exhibits a maximum at $\log(a_{\max}) = 1.26$, i.e. $a_{\max} = 3.53$, see Figure 3. Figure 2(b) displays the MH wavelet coefficients at the *characteristic scale* $a_{\max}$, that is $\mathcal{W}_{I_{304}}(3.53, \vec{b})$. As explained in Section 2, the wavelet coefficient of a function f will be maximum when the corresponding wavelet follows as closely as possible the shape of f. This is exactly what happens in Figure 2: the image given by the CWT at the characteristic scale reveals precisely the contour of the network cells seen in the original image. Below we define (somewhat loosely) the *network size* as the average between the cell diameters

(the cell interior appears in dark in EIT images like Figure 2(a)), and the cell junction thickness (these junctions are shown in bright in EIT images). With this definition, we can say that the characteristic scale gives an indication of the network size.

The essential support diameter of the MH wavelet being $4a_{\max}$, and the EIT pixel representing 1.800 Mm at the Sun (Auchère and Artzner, 2004), we can estimate the network size diameter as $4 \times a_{\max} \times 1.8$ Mm, and its surface as $\pi\ (2 \times a_{\max} \times 1.8)^2$ Mm^2. With the above value $a_{\max} = 3.53$, we obtain a diameter of 25.4 Mm (and a surface of 507 Mm^2). Those values are in good agreement with published values (Simon and Leighton, 1964; Meunier, 2003; Del Moro et al., 2004; DeRosa and Toomre, 2004) of network cell dimensions. For example, Simon and Leighton (1964) found cell diameters that range between 20Mm and 50Mm, whereas Del Moro *et al.* (2004) reports a mean cell diameter of 27Mm for some images taken in 1999.



(a) $0.5R_\odot$ solar disk (b) CWT at characteristic scale

Figure 2. (a) $0.5R_\odot$ solar disk observed on January 2, 1997 in the 30.4 nm passband. (b) Mexican Hat Continuous Wavelet Transform (CWT) of the left image computed at the characteristic scale $a_{\max} = 3.53$ pixels. The CWT is maximum at that scale because the corresponding wavelets have a size close to the one of the network.

3.2. E

We now analyze EIT 30.4 nm images taken during a period of 7 years, from January 1, 1997 up till September 5, 2004, at a cadence of one image every 5 days. All 595 images have 1024×1024 pixels and do not contain any missing block. Our interest lies in studying the time evolution of the network size over the solar cycle by means of the characteristic scale determined in the previous section.

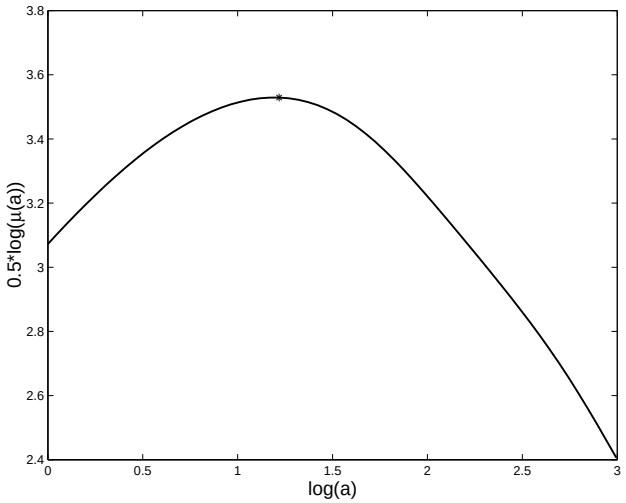


Figure 3. Scale measure of the image on Figure 2(a) in log-log scale. The maximum value of $\mu(a)$ is reached at $a_{\max} = 3.53$, leading to a typical network cell area of $\sim 500\ Mm^2$. See Section 3.1

Before computing the scale measure, it is necessary to reduce the effect of the Active Regions as seen in 30.4 nm. Indeed, these large bright structures produce a linear trend in the scale measure that can wipe out the characteristic scale. Hence in each image, the pixel values that are higher than a given constant threshold $T^\star$ are replaced by $T^\star$. This means that Active Regions are replaced by a constant value for which the corresponding CWT is zero, and thus their contribution to the scale measure is null. The value $T^\star$ was chosen to reduce the impact of active region(s) on $\mu(a)$, while leaving untouched the Solar Minimum images. The optimum was found to be 400 DN s^{-1} in photometrically normalized images, and it is the 0.86−th quantile of the last image in our data set.

After this preprocessing, we compute the Mexican Hat CWT of the whole image, using 60 scales a with values between 1 and 20, regularly spaced on a logarithmic scale. We compute the scale measure $\mu(a)$ as in Equation (3) using as domain Ω the solar disk of radius $0.5R_\odot$.

We are interested in a characteristic scale $a_{\max}$ that lies within the range of values [1.5; 7.5] pixels, that is, that corresponds to a network size belonging to [10.8 Mm;54 Mm]. However, several images do not show any characteristic scale within this range. This happens for example when a large AR covers most of the domain Ω. In this case, it is not possible anymore to extract an indication of the network size. Consequently, our algorithm ignores these images. Out of the 595 images treated, 45 do not exhibit any characteristic

scale within the given range. Most of these images dates from 2001, during the solar maximum.

Figure 4 represents the evolution of the network size diameter over a period of 7 years, together with a nonparametric fit (Friedman, 1984) of this evolution. We displayed here only the 550 diameters that falls within the range prescribed above. Note that in this analysis we did not correct for the varying distance between the SoHO mission and the Sun. Indeed, the corresponding approximation error is of a few percent, whereas our observed quantity (the characteristic scale) varies by a factor 2.

Figure 4 shows clearly that the network cell diameter increases from solar minimum to solar maximum, with peaks in 2000 and 2002. After the solar maximum, this diameter decreases. Note that the two peaks seen in the nonparametric fit of Figure 4 should not be taken in too strict a sense. Indeed, in 2001 several images were excluded from the analysis because no characteristic scales were found within the desired range. Hence the nonparametric fit may slightly underestimate the actual values in 2001.

These results are in concordance with the analysis of Meunier on MDI/SoHO magnetograms (Meunier, 2003), which also shows that the network characteristic size increases as the Sun reaches its period of maximal activity.

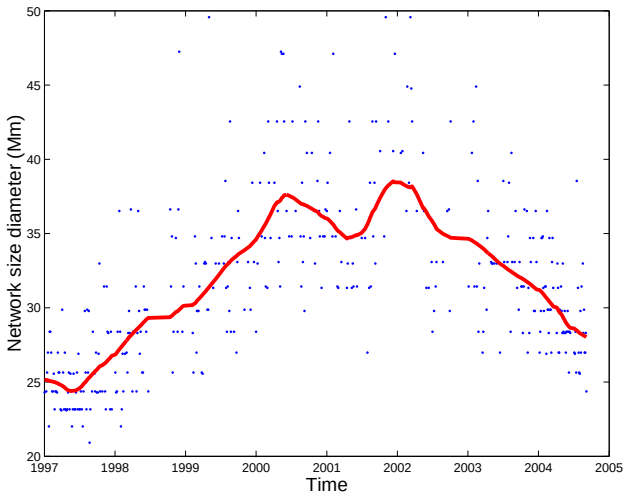


Figure 4. Time evolution of the estimated network size diameter in Mm. This diameter is computed as $a_{\max} \times 4 \times 1.8$, where $a_{\max}$ is the characteristic scale computed from the scale measure. The network size increases from solar minima to solar maxima.

4. Global scale measure analysis for detection of flares

We now consider a coronal data set constituted of a higher cadence sequence taken by the EIT telescope in the 19.5 nm passband (dominated by Fe XII emission) from 1 May 1998 up till 6 May 1998. As before, image format is 1024×1024, with no missing blocks. The sampling time is in general of 12 minutes. However, if an image contains some missing blocks, it is omitted from the analysis, hence it occurs that the sampling time is more than 12 minutes. During these six days there were two periods that collected 512×512 rebinned images. Hence no 1024×1024 images have been recorded on May 1 between 12:25 and 16:53. There is again such a gap between 18:24 on May 2 and 16:34 on May 3.

This particular data set was chosen for its high solar activity content: it contains two X-flares, four M-flares, 25 C-flares and 16 B-flares as recorded by the soft X-ray flux monitor (Bornmann et al., 1996) aboard the GOES-8 satellite. Our aim here is to analyze the wavelet spectrum when it is affected by the presence of bright features. Before proceeding to the wavelet analysis, we found it useful to first pre-process the images in order to remove most of the cosmic rays hits.

4.1. D

To remove cosmic ray hits (hereafter CRH), we modified the algorithm proposed in (Hochedez et al., 2002b; Antoine et al., 2002; Jacques, 2004), so as to decrease the probability of removing solar features (such as brightenings or loop footpoints).

The method detailed in Jacques (2004) proceeds as follows. In a first stage, the noise level in the image is estimated locally using a median filter. CRHs will belong to the set of points (denoted Σ) that exceed several times this noise level. As such, the set Σ contains many solar features. In a second stage, a local Hölder regularity analysis selects points that are uncorrelated with their neighbors. Those are less likely to represent solar structures. The algorithm considers that these pixels have been hit by cosmic rays, and replaces their values by a median over a small neighborhood of size 3×3. The method iterates over these two former stages.

The above algorithm correctly detects most CRHs, but still triggers on some solar features. Indeed, given the good Point Spread Function (PSF) of the EIT optics, point like solar structures can be uncorrelated with their neighbors, and therefore mistakenly categorized as CRHs. This happens especially in the ARs or close to the limb. To prevent the algorithm from removing solar features, we assume that a CRH will rarely appear at the same location on two consecutive images. On the contrary, small solar objects may live longer than 12 minutes at a same location. Our modification consists in looking at

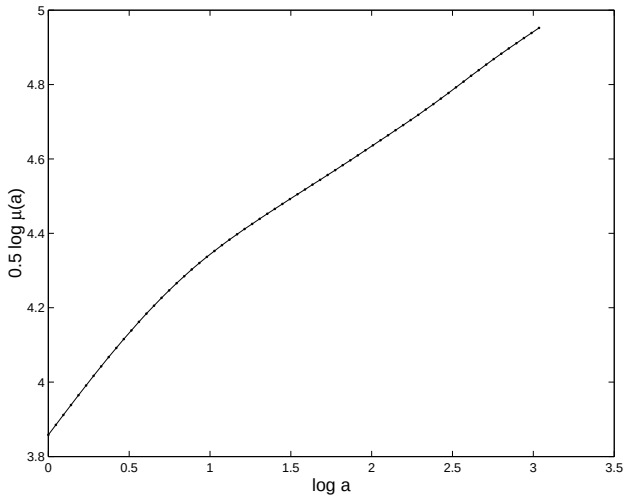


Figure 5. Example of a linear trend of $\log \mu(a)$ corresponding to the image taken on May 4, 1998 at 03:03.

two consecutive images (taken 12 minutes apart). If a detected CRH appears in two successive images, then it is considered as a *false* CRH and is not removed from the original image.

4.2. D 19.5

In this section, we analyze the *on-disc* part, denoted $\Omega^\star$, of the EIT images in the 19.5 nm channel that represents the solar corona. The on-disc part is defined as the disk centered on the Sun and having a radius equal to $0.95R_\odot$ (in order to avoid edge effect at the limb). The scale measure is computed by integration over $\Omega^\star$, and we now study its behavior in detail.

4.2.1. *Log-linear behavior of the scale measure*

In the absence of flares, the wavelet spectrum of a solar corona image typically behaves approximately like a power law (Hochedez et al., 2002a), that is:

$$\frac{1}{2}\log \mu(a) = \beta_0 + \beta_1 \log a + \epsilon\,, \tag{7}$$

where ϵ represents the statistical error in the model. This error accounts for the fact that the values of $\{\mu(a)\}$ computed from the images do not lie exactly on a straight line when represented on a log-log scale. Figure 5 shows an example of a data set $\{\log(a), \log(\mu(a))\}$ that follows approximately a straight line. It gives an indication of the *autosimilarity* shown in the solar corona, in the sense that the wavelet spectrum is always steadily increasing.

We now look at a subset of 44 images taken on May 4, between 03:00 and 15:00. During this period, no flare was recorded. For each image, we compute the wavelet spectrum $\mu(a)$ and we estimate the parameters (β_0, β_1) of its log-linear fit (7) using the least squares (LS) method. Let $(\widehat{\beta_0}, \widehat{\beta_1})$ denote the resulting LS estimate. We also evaluate the goodness of fit by computing the R^2-statistics, which is defined as the ratio between the variance explained by the model $\frac{1}{2}\log\mu(a) = \widehat{\beta_0} + \widehat{\beta_1}\log a$ and the total variation of the original data set. An $R^2 = 1$ means that the model fits the data perfectly (that is, in case of model (7), that the points $\{\log(a), \log(\mu(a))\}$ lie exactly on a straight line), and $R^2 = 0$ means the model is totally inappropriate (Chatterjee and Hadi, 1986).

Figure 6 displays the time evolution of the intercept $\widehat{\beta_0}$, the slope $\widehat{\beta_1}$, and the goodness of fit R^2 for these 44 images. Note that the R^2 is always larger than 0.95, indicating a good linear fit. It is interesting to note that $\widehat{\beta_0}$ and $\widehat{\beta_1}$ evolve in opposite ways: the larger the intercept, the smaller the slope. Figure 7(a) identify more precisely this relationship between intercept and slope, that is, it shows the linear least squares estimate that goes through the points $\{(\widehat{\beta_1}, \widehat{\beta_0})\}$. This linear fit on Figure 7(a) follows the equation

$$\widehat{\beta_0} = 4.87 - 2.69\widehat{\beta_1}, \quad \text{with} \quad R^2 - \text{statistics} = 0.87 \tag{8}$$

By putting equation (8) into the model $0.5\log(\mu(a)) = \widehat{\beta_0} + \widehat{\beta_1}\log a$, we see that at the scale $\log(a^\star) = 2.69$, we have $\frac{1}{2}\log(\mu(a^\star)) = 4.87$. In other words, the energy at the scale $a^\star = \exp(2.69) \approx 14.4$ pixels ~ 26.5Mm remains more or less constant across time, even though the energy level at small scales is subject to variations. Figure 7(b) illustrates the fact that $a^\star$ can be seen as a "fixed point" scale. The origin of this fixed scale is not yet understood, and is a subject for future research.

4.2.2. *Log-quadratic behavior of the scale measure*

In the presence of flares the dynamics of $\mu(a)$ changes: when the flare dominates the image from a signal energy point of view, it creates a characteristic scale: $\mu(a)$ exhibits a maximum *within* (and not at the boundary of) the range of values of $a = 1, \ldots, 20$. The data $(\log a, \frac{1}{2}\log\mu(a))$ are then best fitted using least squares method for the quadratic model

$$\frac{1}{2}\log\mu(a) = \gamma_0 + \gamma_1\log a + \gamma_2(\log a)^2 + \epsilon\,, \tag{9}$$

where $(\gamma_0, \gamma_1, \gamma_2)$ are the parameters of the model, and ϵ is the statistical error.

In fact, a small quadratic component is always present in the evolution of $\log\mu(a)$, as can be noticed from Figure 7(b). However, we need a tool to decide whether this quadratic component may be neglected or not. Our proposed method is as follows. We fit the linear model (7) to the data $(\log(a), \log(\mu(a)))$ using the least squares method, and we compute the resulting R^2-statistics,

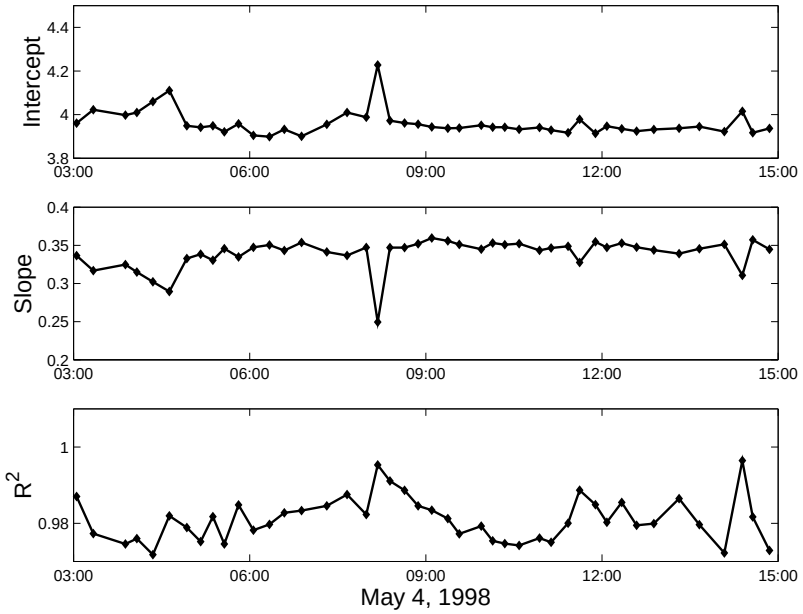


Figure 6. Estimated intercept $\widehat{\beta_0}$ and slope $\widehat{\beta_1}$ together with the R^2-statistics for the 19.5 nm images taken on May 4, 1998. Note how the intercept and slope evolves in opposite directions.

denoted R^2_{lin}. If $R^2_{lin} \geq 0.95$, we consider that the model is linear (quiet situation). On the contrary, if $R^2_{lin} < 0.95$, we compute the parameters of the quadratic model (9). If the R^2-statistic for this quadratic model is superior to 0.9, we consider the model (9) to be satisfactory, otherwise we conclude that

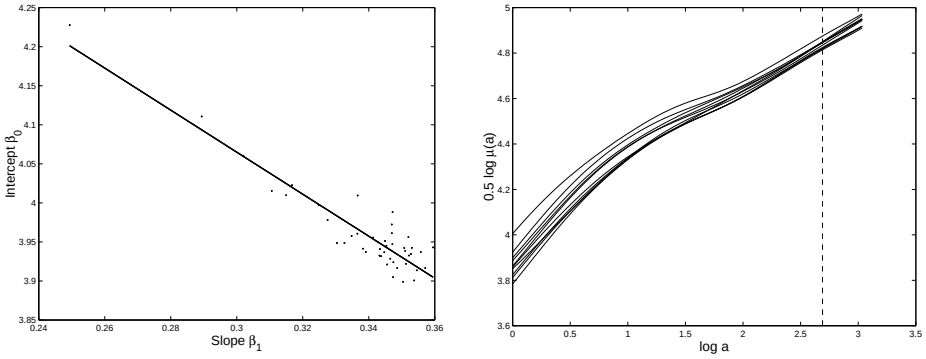


Figure 7. (a) Intercept $\widehat{\beta_0}$ as a linear function of the slope $\widehat{\beta_1}$ (b) First 10 scale measures of the May 4, 1998 series. The vertical line indicates the scale $a^\star$ where all the images have a similar value of $\mu(a)$.

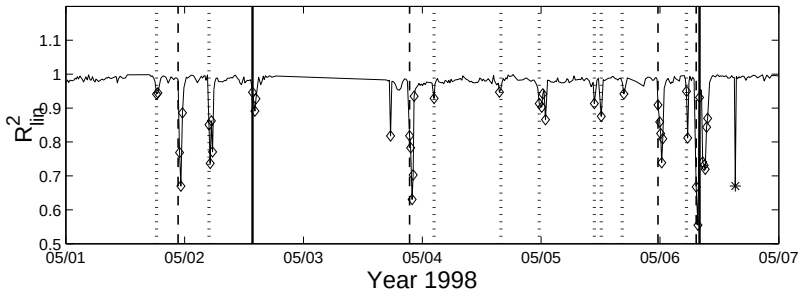


Figure 8. Values of R^2_{lin} for the series of May 1998 in the 19.5 nm passband of EIT. Diamond '$\diamond$' indicates that $R^2_{lin} < 0.95$, whereas a star '$*$' stands for cubic shape. The dashed (resp. continuous) lines represent the beginning of an M-flare (resp. X-flare). The dotted lines represent B- or C-flare activity (flare itself or post-flare eruption) visually seen in EIT image.

the data follows a model of higher order. This selection of a model (linear, quadratic, higher order) based on the value of the R^2-statistics is appropriate to our situation since the data set $\{\log(a), \log(\mu(a))\}$ contains little noise.

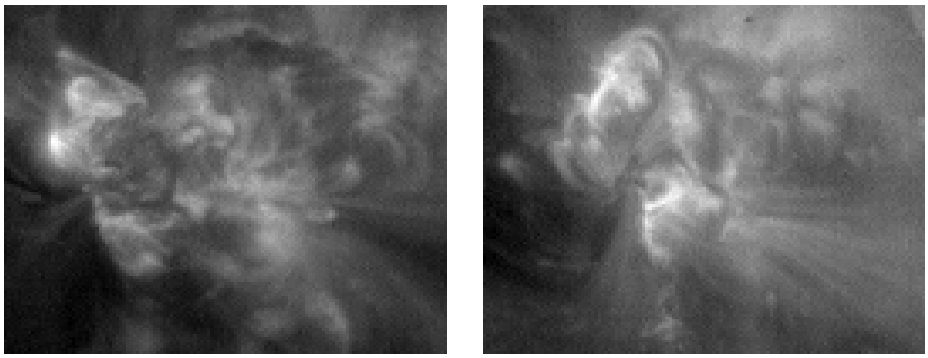
Figure 8 represents the R^2_{lin} values for our data set. Therein, the detection of a quadratic $\mu(a)$ is indicated by a diamond ('$\diamond$'). Only one image had a cubic scale measure; it is represented by a star ('$*$'). We look at the correspondence between a quadratic behavior of the scale measure in EIT, and the occurrence of flares recorded by the GOES X-ray sensor. We discover that M-flares and above are always related to a quadratic behavior of the scale measure.

First, the vertical dashed (resp. continuous) lines on Figure 8 represent the beginning of an M-flare (resp. X-flare). Every time there is an M- or X-flare, it is detected by our method. Indeed, these strong energetic events produce a clear characteristic scale in $\mu(a)$. Second, fainter events are also detected. A visual inspection of the images having a quadratic behavior of $\mu(a)$ shows either the presence of a flare, or of a post-eruption activity. The dotted vertical lines in Figure 8 indicate the time of events (either flares or post-flare eruptions) that correspond to GOES X-ray events detected by our method.

Figure 9(a) shows the snapshot of a post-flaring activity recorded on May 1 at 18:36. From GOES records, the corresponding C-flare ended at 18:17. After the flare there was probably a magnetic reconnection grabbed by our technique, followed by the eruption of a filament. The detail of the region that produces a B-flare on May 5 at 00:20 is represented in Figure 9(b). It is this region that produces the characteristic scale, and hence the quadratic behavior in $\mu(a)$. Note how the scale measure is sensitive to localized events like flaring. This is due to the fact that the CWT itself is localized, and aims at detecting discontinuities in the signal. Hence a bright region results in a

high value of the wavelet coefficients present at the location and scale of the bright event. In other words, due to the localized nature of the CWT, the scale measure is more reactive to local features than, e.g. the total sum or total variation time series.

Figure 8 shows two anomalous images, i.e. false positives detection of flares. The image recorded on May 3 at 17:35 shows a quadratic component without any flaring event: here the quadratic component is in fact created by a cosmic ray that was not filtered out by our preprocessing. Finally, the cubic behavior detected on May 6 at 15:13 is due to a small bright region in a magnetic loop that perturbs the linear trend of the scale measure.



(a) Post-flare (b) Flare

Figure 9. (a) Post-flaring activity recorded on May 1 at 18:36. (b) End of a B-flare recorded on May 5 at 00:20

All the B- or C-flares recorded by GOES do not necessarily have a counterpart in our detection using the global scale measure. This is due partly to the limited cadence of EIT images, and partly to the difference in wavelength ranges (EUV versus XUV) of both instruments. Indeed, the study of Berghmans *et al.* (2001) shows that there is not always a one to one correspondence between events observed in the XUV (such as the SXT telescope on board Yohkoh) and in the EUV.

Since nearly all the scale measures show a good quadratic fit, Figure 10 summarizes the six days time series using the estimated LS parameters $(\widehat{\gamma}_0, \widehat{\gamma}_1, \widehat{\gamma}_2)$ of the quadratic model (9). A bright solar event (flare or post flare eruption) will create a characteristic scale, that is, a maximum in $\mu(a)$, and the corresponding value of $\widehat{\gamma}_2$ will be negative. On the other hand, a cosmic ray will induce a lot of energy at small scale, and may create a quadratic behavior but with a positive square component, as it is the case for the image taken on May 3 at 17:35. Note also the value of the slope $\widehat{\gamma}_1$, which is close to 0.5: $\mu(a)$ is essentially proportional to a.

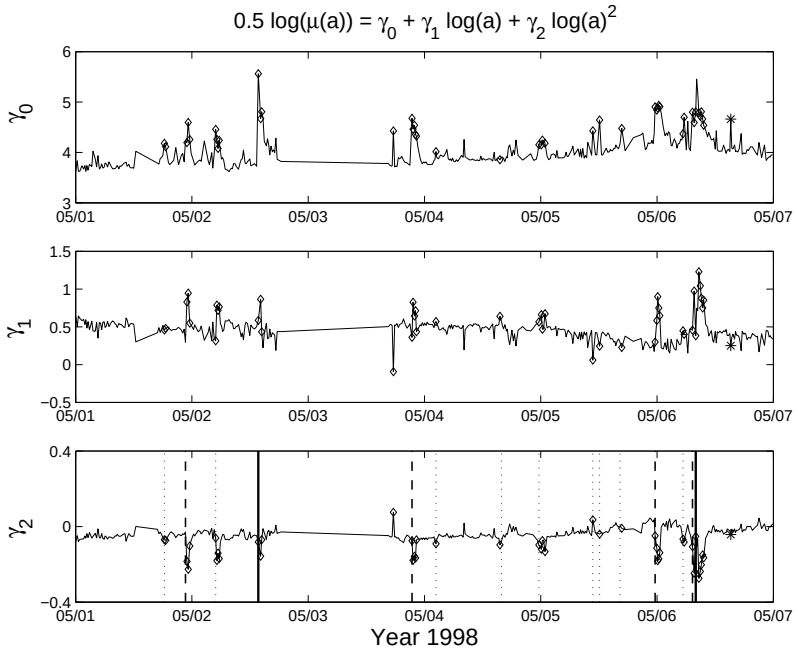


Figure 10. Intercept γ_1, slope component γ_2, and quadratic component γ_3 for the 19.5 nm series. Diamond '$\diamond$' indicates that $R^2_{\text{lin}} < 0.95$, whereas a star '$*$' stands for cubic shape. The continuous (resp. dashed) lines represent the beginning of an X-flare (resp. M-flare). The dotted lines represent B- or C-flare activity (flare itself or post-flare eruption) visually seen in EIT image.

5. Local scale measure for image segmentation

In this section, we propose a method to segment the solar disk based on *local* scale measures. Other works on segmentation include Worden *et al.* (1999), where the authors present a simple method for segmenting 30.4 nm images. The aim is to study the respective contributions of the different structures (plage, enhanced network, active network, quiet chromosphere) to the solar He II irradiance. Although the authors present a good motivation for doing segmentation, their method is somewhat ad hoc; in particular the parameters used for segmentation are derived by trial and error. Similarly, Veselovsky and Zhukov (Veselovsky et al., 2001; Zhukov et al., 2002) studied the contribution of AR, Coronal Holes and Bright Points to the total intensity seen in EIT images.

Our method is based on the fact that AR and Quiet Sun have different local scale measures. A clustering method allows to connect regions having a similar evolution of the local scale measure, and therefore is able to separate

regions of different activity. We first present our algorithm, and thereafter show its performance on two particular examples.

5.1. A

Let $T = 33$ be the diameter pixel size of the circular region used for computation of the local scale measure. Our treated images have 1024×1024 pixels. For each location $\vec{u}$ situated at least at a distance $T/2$ from the border of the image I, we perform the following operations:

1. Construct a small circular region $D^T_{\vec{u}}$ of diameter T centered at $\vec{u}$.
2. For a range of scales $a \in [1, 3]$, compute the local scale measure as
$$\mu_T(\vec{u}, a) = \int_{D^T_{\vec{u}}} \left| \mathcal{W}_I(\vec{b}, a) \right|^2 d^2\vec{b}$$
3. Estimate the parameters $(\beta^T_{0,\vec{u}}, \beta^T_{1,\vec{u}})$ and compute the R^2-statistics (denoted $(R^2)^T_{\vec{u}}$) for the linear model
$$\log \mu_T(\vec{u}, a) = \beta^T_{0,\vec{u}} + \beta^T_{1,\vec{u}} \log a + \epsilon.$$
Let $\widehat{\beta}^T_{0,\vec{u}}$ and $\widehat{\beta}^T_{1,\vec{u}}$ denote these least squares estimates.

Given the diameter $T = 33$ pixels, the maximum scale $a^{\max} = 3$ is chosen so that the circular region $D^T_{\vec{u}}$ may contain the essential support of four wavelets at scale $a^{\max}$ without overlapping.

As a result of the operations described above, we obtain 3 matrices of size 992×992. On these matrices we apply a mask in order to retain only the *on-disc* part, similarly to what is done in Section 4. For a given location $\vec{u}$, the behavior of $(\widehat{\beta}^T_{0,\vec{u}}, \widehat{\beta}^T_{1,\vec{u}}, (R^2)^T_{\vec{u}})$ will be different depending on whether $\vec{u}$ lies within an Active Region, inside the Quiet Sun, or close to a cosmic ray hit.

Our proposed method consists in grouping the values of $(\widehat{\beta}^T_{0,\vec{u}}, \widehat{\beta}^T_{1,\vec{u}}, (R^2)^T_{\vec{u}})$ into homogeneous clusters. Note that the data set used for clustering is quite large: it consists of $\approx 4.10^5$ locations $\vec{u}$ (or 'observations' in statistical terms) having each 3 components $(\widehat{\beta}^T_{0,\vec{u}}, \widehat{\beta}^T_{1,\vec{u}}$, and $(R^2)^T_{\vec{u}})$. Hence, amongst the set of clustering methods proposed in the statistical literature, we used the CLARA method of Kaufman and Rousseeuw (Kaufman and Rousseeuw, 1990; Struyf et al., 1997), since it is precisely designed to deal with large data sets. The CLARA method uses a partitioning of the database into a set of k clusters, where k needs to be specified by the user. Each cluster is represented by one of the objects in the cluster (called 'medoids'), and the algorithm aims at finding the partition that minimizes the sum of distance between a point

and its medoids. The procedure also gives as output a quality index (called the ‘isolation’ number) that reflects the strength of the clustering. Hence in the figures, the clusters are numbered as follows: the cluster number 1 has the highest quality index, the cluster number two the second highest quality index, and so on.

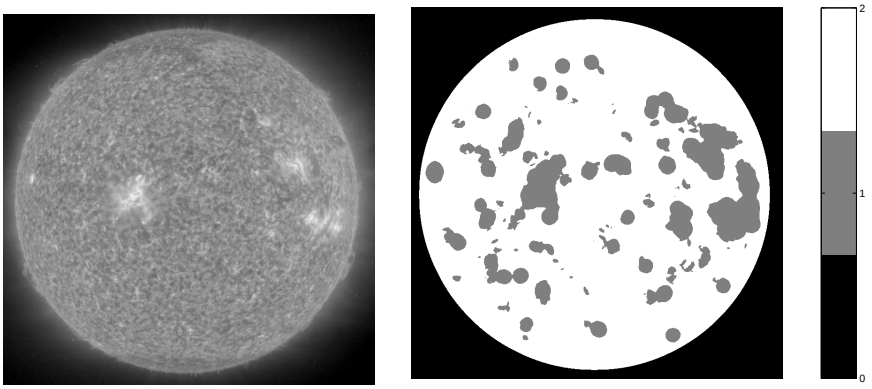
5.2. S

We first tested the segmentation procedure on two 30.4 nm images, one being taken in 1996 in period of solar minima, and the second in 2000 during maximal activity of the Sun. We tried the partitioning method using 2, 3, and 4 clusters. As expected, the 3rd and 4th clusters are negligible, only two large sets of points are detected by the method: one corresponds to the AR seen in 30.4 nm, and the second to the network, see Figure 11. The comparison between Figures 11(b) and 11(d) shows that the proportion of AR is higher in period of maximal activity. Hence this method could be used as a basis for automatically monitoring the proportion of the solar disk that contain active regions in the 30.4 nm passband. Similarly to the work of Worden *et al.* (1999), the aim would be to understand how the different structures seen in He II contribute to the solar irradiance and its variability.

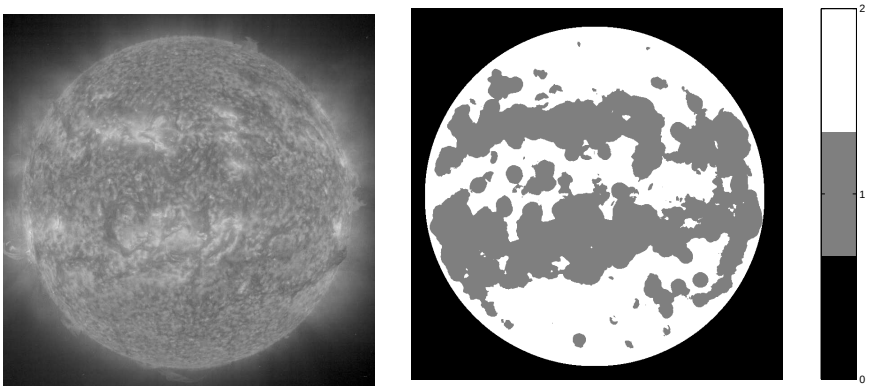
5.3. S

The images of He II could rather clearly be decomposed into network and active regions. The situation is more complex in the case of solar corona images, since these contain Active Regions, Quiet Sun, and Coronal Holes that do not have clear mutual boundaries. As before we performed the clustering method using 2, 3 and 4 clusters. We found out that the Quiet Sun and the Coronal Holes are grouped into one cluster. As we increase the number of clusters in the segmentation, the CLARA algorithm divides the bright regions and leaves untouched the Quiet Sun and Coronal Holes. Figure 12 shows the partitioning of two 19.5 nm (Fe XII) images using 3 clusters. One image was taken during solar minimal activity, the other in period of solar maxima. In both images, the first two clusters corresponds to bright regions: the cluster 2 contains the Active Regions and magnetic loops, and the cluster 1 groups the points close to the Active Regions that are relatively bright. The cluster number 3 represents the Quiet Sun and Coronal Holes, and has a lower quality index (i.e. is less well separated) than the first two clusters.

The method proposed in our paper could be used to *autonomously* evaluate the surface covered by active regions, and to also address EUV irradiance issues. Moreover, image segmentation procedures allow the development of new tools for long term studies equivalent to butterfly diagrams. A pre-processing that removes the strongest cosmic rays such as the one described in Section 4.1 would probably improve the performance. However, it



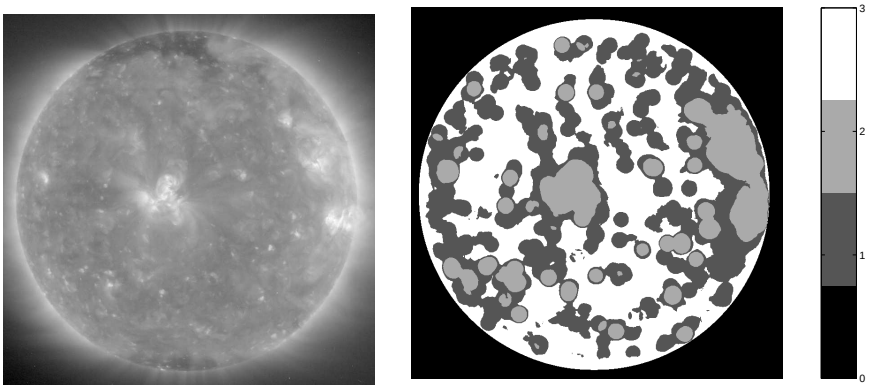
(a) 30.4 nm image in 1996 (b) Segmentation of (a) in two clusters



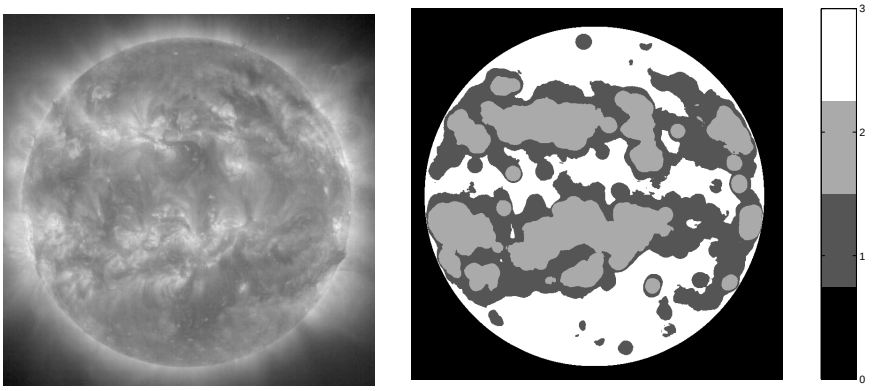
(c)30.4 nm image in 2000 (d) Segmentation of (c) in two clusters

Figure 11. (a) On-disc part of an 30.4 nm EIT image recorded on December 22,1996. (b) Result of the CLARA partitioning on the values of the linear fit and R^2-statistics of local scale measure. The two clusters shows the Active Region (Cluster 1) and the chromospheric network (Cluster 2). (c) On-disc part of an 30.4 nm EIT image recorded on August 03, 2000. (d) Result of the partitioning method of the image in (c) using two clusters. Again, Cluster 1 identifies the AR and Cluster 2 the network.

is important to keep solar features untouched during this operation. The fact that the Coronal Holes cannot easily be separated from the Quiet Sun may be due to the fact that both structures are relatively homogenous compared to ARs. In order to improve the performance of the segmentation, quadratic components of the local scale measure will be used in future research.



(a) 19.5 nm image in 1996 (b) Segmentation of (a) in three clusters



(c)19.5 nm image in 2000 (d) Segmentation of (c) in three clusters

Figure 12. (a) On-disc part of an 19.5 nm EIT image recorded on December 22, 1996. (b) Result of the partitioning method on the values of the intercept, slope and R^2 statistics computed from the local scale measure of the image. Cluster number 1 reveals the relatively bright areas close to the AR, whereas the cluster 2 shows the active regions together with regions that contains cosmic ray hits or brightenings. The quality index of these two clusters are close. The cluster 3 groups together the Quiet Sun and Coronal holes. (c) On-disc part of an 19.5 nm EIT image recorded on August 03, 2000. (d) Partitioning of the image in (c) in 3 clusters: Cluster number 1 reveals the areas close to the Active Regions that are brighter than the Quiet Sun, cluster 2 represents the active regions and some cosmic ray hits, and the Cluster 3 delimits the Quiet Sun together with the Coronal Holes

6. Discussion

In this paper, we have shown three applications of the wavelet spectrum in the analysis of EIT/SoHO images. The wavelet spectrum allows to follow the evolution of the chromospheric network size, to detect flares, and to separate, e.g, Quiet Sun and Coronal Holes from the Active Regions. These promising

results stimulate further investigations. We now discuss two possible extensions of our work.

In the 30.4 nm passband, we are working towards an approximation of the network size *distribution* using the scale measure. It will be of interest to see how this distribution evolves over the solar cycle.

In the 19.5 nm passband, a higher cadence of images would make it possible to *forecast* eruptions. Indeed, we empirically observed a small increase in scale measure at the smallest scale before a large eruption occurred. We conjecture that the scale measure reveals unresolved phenomena that could be interpreted in the framework of resistive MHD .

Finally, we only showed two applications of the segmentation of an image. In order to separate more precisely an image into Quiet Sun, Active Regions, Coronal Holes, Filaments/Prominences, etc,..., it might be necessary to use the local scale measure on data acquired simultaneously at various temperatures. The AIA instruments of the SDO mission will offer such a possibility.

Acknowledgement

The authors acknowledge the EIT consortium for building and operating the instrument. V. D, J. D.P. and J.-F. H. acknowledge the support from the Belgian Federal Science Policy Office through the ESA-PRODEX programme. The authors would like to thank A. Arneodo, D. Berghmans, F. Clette, E. Podladchikova, and A. Zhukov for helpful discussions, and J. Newmark for all his valuable and precise comments. Computer support from L. Wauters, S. Willems and A. Somerhausen is also acknoweldged. SoHO is a project of international cooperation between ESA and NASA to study the Sun.

References

Aletti, V., M. Velli, K. Bocchialini, G. Einaudi, M. Georgoulis, and J.-C. Vial: 2000, 'Microscale Structures on the Quiet Sun and Coronal Heating'. ApJ **544**, 550–557.

Antoine, J.-P., L. Demanet, J.-F. Hochedez, L. Jacques, R. Terrier, and E. Verwichte: 2002, 'Application of the 2-D wavelet transform to astrophysical images'. *Physicalia Magazine* **24**, 93–116.

Antoine, J.-P., R. Murenzi, and P. Vandergheynst: 1998, 'Directional wavelets revisited: Cauchy wavelets and symmetry detection in patterns'. *Appl. Comp. Harm. Anal.*

Antoine, J.-P., R. Murenzi, P. Vandergheynst, and S. T. Ali: 2004, *Two-dimensional wavelets and their relatvies*. Cambridge University Press.

Auchère, F. and G. E. Artzner: 2004, 'EIT Observations of the 15 November 1999 Mercury Transit'. Sol. Phys. **219**, 217–230.

Berghmans, D., D. McKenzie, and F. Clette: 2001, 'Active region transient brightenings. A simultaneous view by SXT, EIT and TRACE'. A&A **369**, 291–304.

Bornmann, P. L., D. Speich, J. Hirman, L. Matheson, R. Grubb, H. Garcia, and R. Viereck: 1996, 'GOES x-ray sensor and its use in predicting solar-terrestrial disturbances'. In: *Proc. SPIE Vol. 2812, p. 291-298, GOES-8 and Beyond, Edward R. Washwell; Ed.* pp. 291–298.

Chatterjee, S. and A. Hadi: 1986, 'Influential Observations, High Leverage Points, and Outliers in Linear Regression'. *Statistical Science* pp. 379– 416.

Del Moro, D., F. Berrilli, T. L. Duvall, and A. G. Kosovichev: 2004, 'Dynamics and Structure of Supergranulation'. *Solar Physics* **221**, 23–32.

DeRosa, M. L. and J. Toomre: 2004, 'Evolution of Solar Supergranulation'. ApJ **616**, 1242–1260.

Friedman, J. H.: 1984, 'A Variable Span Smoother'. Technical Report 5, Dept. of Statistics, Stanford Univ., California.

Hagenaar, H. J., C. J. Schrijver, and A. M. Title: 1997, 'The Distribution of Cell Sizes of the Solar Chromospheric Network'. ApJ **481**, 988–+.

Hochedez, J.-F., L. Jacques, E. Verwichte, D. Berghmans, L. Wauters, F. Clette, and P. Cugnon: 2002a, 'Multiscale activity observed by EIT/SoHO'. In: *ESA SP-477: Solspa 2001, Proceedings of the Second Solar Cycle and Space Weather Euroconference*. pp. 115–118.

Hochedez, J.-F., L. Jacques, A. Zhukov, F. Clette, and J.-P. Antoine: 2002b, 'Small features in the EIT-SOHO images'. In: *ESA SP-508: From Solar Min to Max: Half a Solar Cycle with SOHO*. pp. 295–298.

Jacques, L.: 2004, 'Ondelettes, repères et couronne solaire'. Ph.D. thesis, Université catholique de Louvain.

Kaufman, L. and P. Rousseeuw: 1990, *Finding Groups in Data: An Introduction to Cluster Analysis*. New York: Wiley.

Kittel, C.: 1958, *Elementary Statistical Physics*. New York: Wiley.

Meunier, N.: 2003, 'Statistical properties of magnetic structures: Their dependence on scale and solar activity'. A&A **405**, 1107–1120.

Portier-Fozzani, F., B. Vandame, A. Bijaoui, A. J. Maucherat, and E. Team: 2001, 'A Multi-scale Vision Model applied to analyze EIT images of the solar corona'. Sol. Phys. **201**, 271–287.

Reeves, E. M.: 1976, 'The EUV chromospheric network in the quiet sun'. Sol. Phys. **46**, 53–72.

Sanz, J. L., D. Herranz, and E. Martínez-Gónzalez: 2001, 'Optimal Detection of Sources on a Homogeneous and Isotropic Background'. ApJ **552**, 484–492.

Simon, G. W. and R. B. Leighton: 1964, 'Velocity Fields in the Solar Atmosphere. III. Large-Scale Motions, the Chromospheric Network, and Magnetic Fields.'. ApJ **140**, 1120–+.

Struyf, A., M. Hubert, and P. J. Rousseeuw: 1997, 'Integrating robust clustering techniques in S-PLUS'. *Comput. Stat. Data Anal.* **26**(1), 17–37.

Torrésani, B.: 1997, 'Analyse Temps-Fréquence et Signaux Non-Stationnaires'. Lectures notes, Montpellier.

Veselovsky, I. S., A. N. Zhukov, A. V. Dmitriev, M. V. Tarsina, F. Clette, P. Cugnon, and J. F. Hochedez: 2001, 'Global Asymmetry of the Sun Observed in the Extreme Ultraviolet Radiation'. Sol. Phys. **201**, 27–36.

Worden, J., T. N. Woods, W. M. Neupert, and J. Delaboudinière: 1999, 'Evolution of Chromospheric Structures: How Chromospheric Structures Contribute to the Solar HE II 30.4 Nanometer Irradiance and Variability'. ApJ **511**, 965–975.

Zhukov, A. N., I. S. Veselovsky, J.-F. Hochedez, F. Clette, O. A. Panasenco, and P. Cugnon: 2002, 'Global asymmetry of the Sun observed in the extreme ultraviolet radiation'. In: *ESA SP-508: From Solar Min to Max: Half a Solar Cycle with SOHO*. pp. 189–192.