Comparisons between two wavelet functions in extracting coherent structures from solar wind time series

M. J. A. Bolzan* and F. L. Guarnieri
Instituto de Pesquisa e Desenvolvimento, Universidade do Vale do Paraíba (UNIVAP), São José dos Campos, Brazil

Paulo Cesar Vieira
Instituto de Pesquisa e Desenvolvimento, Universidade do Vale do Paraíba (UNIVAP), São José dos Campos, Brazil
(Received on 6 March, 2008)

Nowadays, wavelet analysis of turbulent flows have become increasingly popular. However, the study of geometric characteristics from wavelet functions is still poorly explored. In this work we compare the performance of two wavelet functions in extracting the coherent structures from solar wind velocity time series. The data series are from years 1996 to 2002 (except 1998 and 1999). The wavelet algorithm decomposes the annual time-series in two components: the coherent part and non-coherent one, using the daubechies-4 and haar wavelet function. The threshold assumed is based on a percentage of maximum variance found in each dyadic scale. After the extracting procedure, we applied the power spectral density on the original time series and coherent time series to obtain spectral indices. The results from spectral indices show higher values for the coherent part obtained by daubechies-4 than those obtained by the haar wavelet function. Using the kurtosis statistical parameter, on coherent and non-coherent time series, it was possible to conjecture that the differences found between two wavelet functions may be associated with their geometric forms.

Keywords: Wavelet Analysis; Coherent Structures; Solar Variables; Turbulence

1. INTRODUCTION

The turbulence is still an open problem in Physics and it constitutes a multiscale phenomena. An approach to turbulence like Direct Numerical Simulation (DNS) is impracticable due the computational obstacle. So, Larg-Eddy Simulations (LES) become a good approach for computational calculations of complex turbulent flows [7]. Recently, Farge et al. [8] developed the Coherent Vortex Simulation (CVS) which use the wavelet approach to deterministically simulate the time evolution of the coherent vortex. They used the assumption that coherent vortices are responsible for the non-Gaussianity of the Probability Density Function (PDF) of vorticity. Thus, the study of the statistical characteristics of the turbulence is an important tool to understand the physical mechanisms acting in the energy transfer between scales. In particular, the statistical study of Magnetohydrodynamical (MHD) turbulence is important to understand the energy transfers within the Solar-Terrestrial system. The consequences of solar disturbances on the magnetosphere-ionosphere system from Earth, have been analyzed in several studies, showing the importance of the intermittence and Coherent Structures (CS) present in the MHD flow [2, 16, 18]. According to Hussain [13], CS are turbulent variables (velocity, temperature, density and others) that have high self-correlation or with another variables in the determined time scale. Furthermore, Burlaga and Mish [4] have argued that the -2 spectral index found in the solar wind velocity may arise if coherent power is present in large-amplitude, low-frequency fluctuations. These large-amplitude low-frequency fluctuations are related to shocks evolving in the solar wind. The structures are strongly associated with energy dissipation of the turbulent flows and also a source of instability, at least in some scales.

Due to the problem related to the presence of CS in time series, several works have been developing methodologies to detect and extract the CS from time series [9, 12, 21] using the Wavelet Transform (WT). However, there is still the problem related to the choice of the wavelet function to perform the CS extraction. In fact, only a few studies have focused the geometric properties of the wavelet and coherent structures [3, 19]. Thus, in the present work, a comparison between results from extraction of CS performed by two wavelet functions from solar wind velocity time series is presented. The power spectral index and statistical analysis are used to quantify the possible differences. The WT was used to separate the solar wind velocity in two components: one that contains only the coherent (CS) part removed from original time series, and the other component containing information about the non-coherent (NC) part.

2. DATA

Solar wind data from SOHO satellite obtained in the public internet address *CELIAS/MTOF http://umtof.umd.edu/pm* was employed in this study. Data were obtained between years 1996 to 2002, except 1998 and 1999, with 1-hour time resolution (approximately $2.78X10^{-4}$ Hz). Figure 1 show the solar wind velocity plots for the five years analyzed.

Based in the approach of Kovács et al. [16], we can conjecture that the fluctuating nature of solar wind variables can be interpreted, in the present context as the manifestation of turbulent phenomena that takes place within the MHD flows. It has long been accepted that turbulence evolves through cascade processes that involve a hierarchy of coherent vortex structures belonging to a wide range of spatial scales. Kolmogorov [15] proposed the inhomogeneous flow-down (cascade) of the energy from system-size scales to dissipative scales (Kolmogorov). The inhomogeneity involves the singular behavior of the energy distribution in physical space resulting in strong gradients, or intermittence, in the time-series of the energy related to physical quantities of the system, e.g. ve-

*Electronic address: bolzan@univap.br

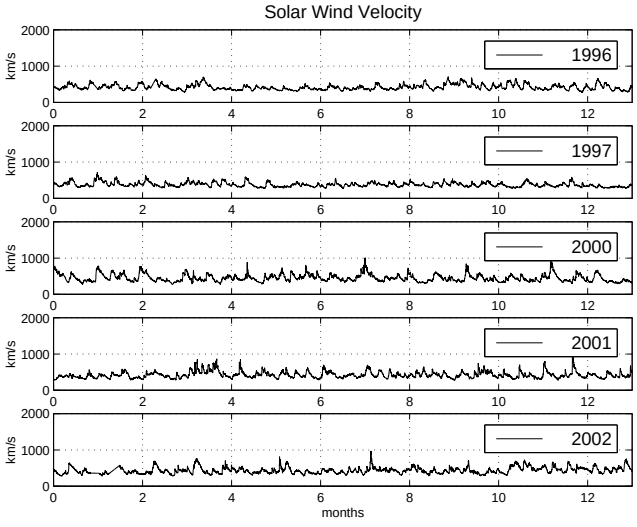

FIG. 1: Solar wind velocity time series for the five years shown in legend, measured by SOHO satellite.

locity [20], temperature [1] or magnetic fields [16, 18]. Thus, we used the approach of this Kolmogorovian scenario to study this data set.

3. THEORETICAL BACKGROUNDS

3.1. Wavelet Transform

The WT is a mathematical tool able to analyze any non-stationary time-series, showing the temporal variability of the power spectral density. The "wavelet" word indicates a set of functions with the form of small waves created by dilations, $\Psi(t) = \Psi(2t)$, and translations, $\Psi(t) = \Psi(t+1)$, applied on a simple generator function, $\Psi(t)$, which is called mother-wavelet. Mathematically, the wavelet function, with a scale a and at position b, is given by

$$\Psi_{a,b}(t) = a^{-1/2}\Psi\Big(\frac{t-b}{a}\Big), \tag{1}$$

where a and b are real and $a > 0$. The wavelet transform is defined by

$$W_\psi f(a,b) = \frac{1}{\sqrt{2}}\int f(t)\Psi_{a,b}(t)dt, \tag{2}$$

where the temporal function $f(t)$ is any time series.

There are two types of wavelet functions: orthogonal and non-orthogonal wavelets [6]. The most used orthogonal wavelet families are: haar, meyer, and daubechies, that are used for filtering/decomposition of the time series. The most famous non-orthogonal wavelet families are: morlet and the mexican-hat. For this study, we used two orthogonal wavelet functions: daubechies-4 and haar [5, 20]. The haar function is orthogonal, with dilation being dyadics in the form $a = 2^{-j}$ and its translations occur in discrete steps, in the form $b = 2^{-j}k$, where j and k are integers. This wavelet function is given by:

$$\Psi_{j,k}(t) = \begin{cases} 2^{j/2}, & 2^{-j}k \le t < (k+1/2) \\ -2^{j/2}, & 2^{-j}(k+1/2) \le t < (k+1) \\ 0, & \text{all other values of t} \end{cases}$$

The daubechies wavelet filters are the ones commonly used in image retrieval [17]. In general, db_n represents the family of daubechies wavelets, and n is the order. It is important to note that the family includes haar wavelet since it represents the same wavelet as db_1. The wavelets are built based in an small function $\phi(t)$, given by [17]:

$$\phi(t) = \sqrt{2}\sum_k l_k \phi(2t-k), \tag{3}$$

where ϕ is called scale function or scaling wavelet. The mother wavelet ψ is obtained by:

$$\psi(t) = \sqrt{2}\sum_k h_k \phi(2t-k), \tag{4}$$

where l_k and h_k are called coefficients of low-pass and high-pass filters, and are related by:

$$h_k = (-1)^k l_{1-k}. \tag{5}$$

For the db_4 used in this work, the coefficients for low-pass and high-pass filter are, respectively:

$$\left\{ \frac{\sqrt{3}+1}{4\sqrt{2}}, \frac{\sqrt{3}(\sqrt{3}+1)}{4\sqrt{2}}, \frac{\sqrt{3}-1}{4\sqrt{2}}, \frac{\sqrt{3}(\sqrt{3}-1)}{4\sqrt{2}} \right\}.$$

$$\left\{ -\frac{\sqrt{3}(\sqrt{3}-1)}{4\sqrt{2}}, \frac{\sqrt{3}-1}{4\sqrt{2}}, \frac{\sqrt{3}(\sqrt{3}+1)}{4\sqrt{2}}, -\frac{\sqrt{3}+1}{4\sqrt{2}} \right\}.$$

We chose the db_4 wavelet because this function is more regular than db_2, where the regularity increases with the order n. Figure 2 shows the mother and scaling wavelets of db_4 and haar for $j = 0$ and $k = 0$.

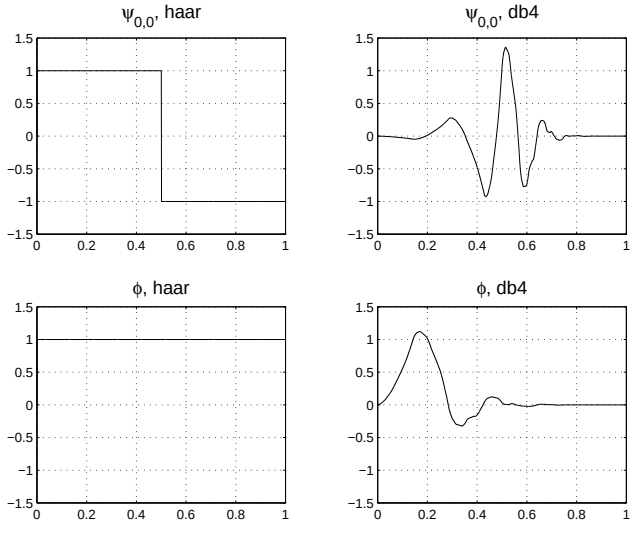

FIG. 2: Scaling (ϕ) and mother (ψ) functions for db_4 and haar wavelets.

3.2. Procedure for Coherent Structures Extraction

In order to study the performance of the two wavelet functions in extracting the CS, the time series was decomposed in all dyadic scales, i.e., if the time series is n data points long, it is possible to decompose it in j dyadic scales of $n = 2^j$. In the following step, we performed the variance in each dyadic scale to find the scale where the variance has maximum value. After this, we used a fixed threshold of 95% maximum variance found in a given scale and to separate the coherent (CS) and non-coherent (NC) wavelet coefficients. In this way, the wavelet coefficients with values higher than 95% maximum variance are considered due to CS, and wavelet coefficients below this threshold are considered the NC part. Finally, the CS and NC parts are then reconstructed by the WT inverse. Note that we use the same assumption used by Ruppert-Felsot et al. [21], i.e., the original time series can be represented by a few large amplitude wavelet coefficients, while the noise is contained by several remaining coefficients of small amplitude.

3.3. Statistical Analysis

The intermittent nature of the any physical quantity, can be investigated through the called Probability Density Function (PDF) of a set of two-point difference in time-series, $\overline{\delta H_r}(t) = \left(\delta H(t) - \langle \delta H(t) \rangle\right)/\sigma$, of the original field, H, where σ means the standard deviation of the differenced time-series. The parameter r represents a temporal increment. For ordinary HD or MHD fluids, intermittency appears in the heavy tails of the distribution functions at moderate scales implying non-Gaussian statistical behavior of the systems [10]. Frisch [10], proposed a way to quantify the degree of deviation from Gaussian distributions, i.e. the level of intermittency at different scales. We computed the kurtosis values of the two-points difference time-series defined as:

$$F = \frac{< \delta H_r(t)^4 >}{< \delta H_r(t)^2 >^2} \qquad (6)$$

where, again, the parameter r represents an increment scale. The kurtosis of a normally distributed process is equal to 3 [10]. Adding intermittent fluctuations to an originally Gaussian signal implies the spreading of its PDF and, consequently, the increase of its kurtosis value. We will calculate the kurtosis in 18 logarithmically spaced scales for Original, CS and NC parts obtained by haar and db_4.

4. RESULTS AND DISCUSSION

The process above mentioned was applied in five annual time series. Figure 3 shows the results of this filter process using the db_4, where the upper panel shows the original time series, the middle panel shows the CS information, and the bottom panel contains the NC part time series. We note that the CS signal (middle panel in Fig. 3) presents the same large structures observed in the original signal. This fact is an important point from db_4 wavelet function because it is necessary to extract the real information from original signal without introduce unreal information. Figure 4 shows the same time series

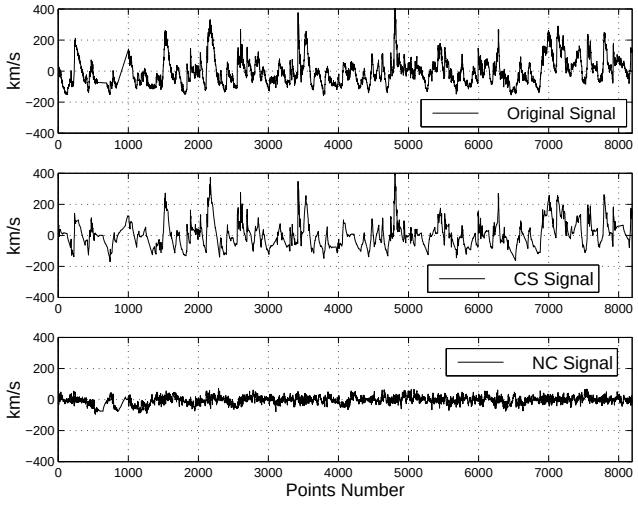

FIG. 3: Daubechies-4 filter process applied to solar wind velocity data for year 2002. Top panel shows the original time series, middle panel shows 95% of CS, and bottom panel shows the NC time series.

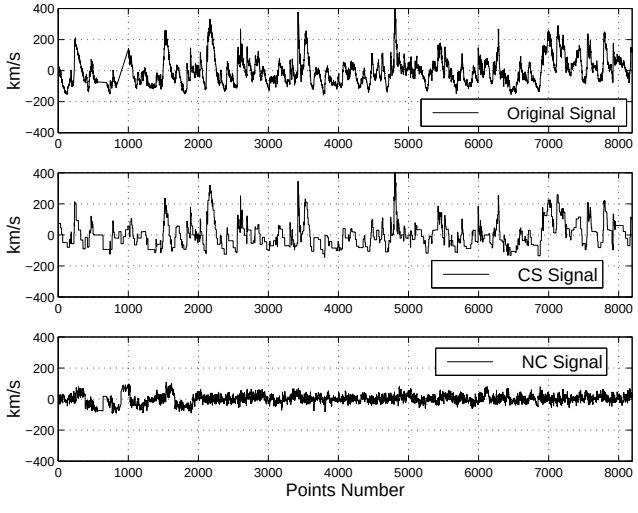

FIG. 4: Haar filter process applied on solar wind velocity data for year 2002. Top panel shows the original time series, middle panel shows 95% of CS, and bottom panel shows the NC time series.

but obtained by haar filtering process where, in this case, we note that the haar function creates false features such as rectangular forms on the CS time series shown in middle panel (Fig. 4). However, it is important point out that the NC part from both extractions shows evidence of coherent structures as well. In fact, we conjecture that this behavior may be due the non optimal choice of the threshold that do not permit to extract efficiently the all coherent structure presents in the time series. Even so, our objective here is to do a preliminary study about the two wavelet functions and to use a methodology based in statistical approach to compare these two wavelet functions.

We applied the Power Spectral Density (PSD) on all original time series and for the CS time series obtained from the filtering process to find out the power spectral index. These power spectral indexes were obtained through Minimum

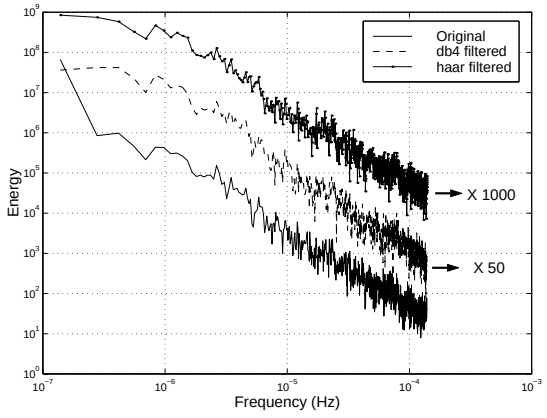

FIG. 5: Power spectral density (PSD) applied for the solar wind time series shown in Fig. 2. The continuous line represents the PSD for the original time series; dotted line represents the PSD for the coherent signal filtered by db_4; dashed line represents the PSD for the coherent signal filtered by haar function. The three PSD were vertically shifted for better visualization.

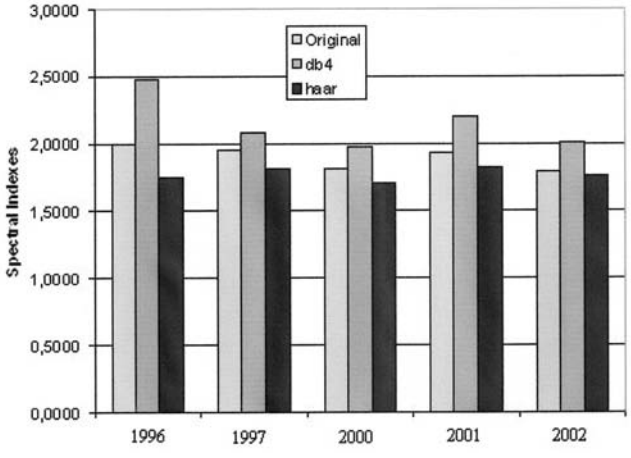

FIG. 6: Absolute values of spectral index for original time series and CS time series from db_4 and from haar wavelets.

Square Fitting applied on loglog plots of the PSD. Figure 5 shows an example of PSD applied on the same time series shown in Fig. 2.

The power spectral indices for all time series are shown in Fig. 6. We observe superior values of spectral indices for CS time series obtained by db_4 filtering process when compared with those CS for the time series obtained by haar filtering process. Note that we use the fact that the statistical theory of homogeneous turbulence suggests that the noise may have some correlation, which corresponds to a scaling law steeper than for a white noise, i.e. $k^{-5/3}$ in 3D. Thus, the presence of the CS on the signal promotes the elevation of the spectral indices. These preliminary results indicate that the db_4 wavelet was better in extracting the CS from original time series. The mean values of spectral indices from CS time series from db_4 was $-(2.16 \pm 0.20)$, in agreement with results from Ishizawa and Hattori [14], who also used the WT procedure.

In order to define the best wavelet function between db_4 and haar, we calculated the kurtosis parameter. The kurtosis is a statistical parameter useful to detect the intermittent phenomena present in the time series [10] and was performed in 18 logarithmically spaced scales. Figure 7 shows the results of the kurtosis for the original time series, CS, and NC time series obtained from db_4 filtering process and also for CS and NC time series obtained from haar. Note that the kurtosis mean values from NC time series for both wavelet functions are very close to a Gaussian distribution, i.e., kurtosis near to 3.

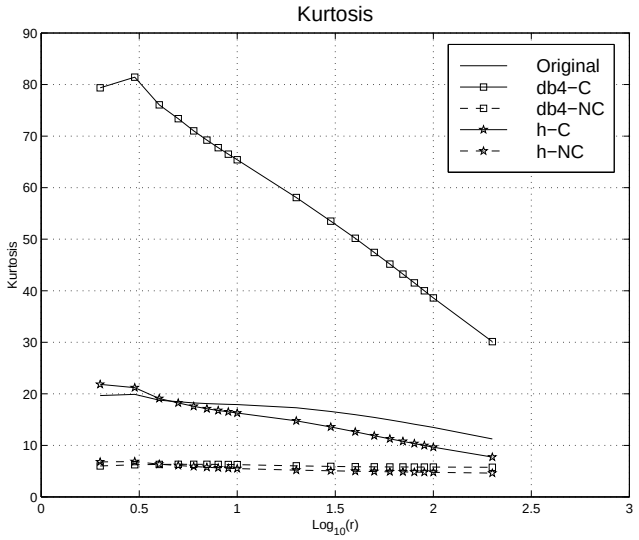

FIG. 7: Kurtosis parameter mean values for original time series (continuous line), coherent structure (square symbol) and non-coherent structure (square symbol and dotted line) time series from db_4, and coherent structure (star symbol) and non-coherent structure (star symbol and dotted line) time series from haar process filtering.

A rate between kurtosis mean values from CS and NC time series for both wavelet results showed above was evaluated. Figure 8 shows the mean values of this rate for kurtosis obtained from db_4 and haar filtering process. We observe that the results from db_4 is closer to the results from original time series than the results obtained with the haar wavelet. This fact shows that the db_4 filtering process is the best method to extract more coherent structures from original time series when compared with haar results.

The results above mentioned are probably associated with the geometric form of each wavelet. In this sense, we observe that the db_4 wavelet function have a geometric form much closer to the geometric characteristics of the time series. This fact of the matching between wavelet functions and time series explain also the difference of the spectral indices found in Fig. 6. We observed higher values of the spectral indices of CS time series from db_4 when compared with CS time series from haar, showing that the db_4 wavelet function is able to extract more physical information from original time series than haar wavelet function. The fact that db_4 wavelet function presents better results than haar wavelet may be explained by the db_4 anti-symmetric form. This idea was already pointed out by Hagelberg and Gamage [11], who said the following: *"We illustrate the dependence of the coherent structure detection mechanism on the choice of analyzing wavelet, demonstrating that anti-symmetric wavelets are better suited to de-*

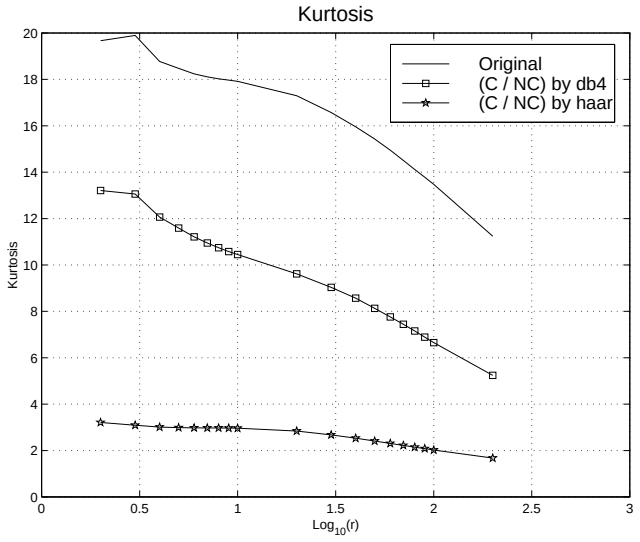

FIG. 8: Kurtosis values for original time series (continuous line), kurtosis of the rate between CS and NC time series obtained for db4 (square symbol) and for haar filtering process (star symbol).

tecting zones of concentrated shear, while symmetric wavelets result in detection of zones of concentrated curvature."

This is only a preliminary study since it is necessary to compare results obtained in different percentage values of coherent structure extraction. Further studies may focus on the comparison of these results for different solar cycle phases (such as a minimum, during year 1996, and a maximum around year 2001), and also estimate the CS detection/extraction algorithm efficiency. Anyway, the results here presented can give an orientation for the choice of an appropriated wavelet to be employed in the extraction of coherent structures from turbulent time series measured in hydrodynamic and magnetohydrodynamic environment.

5. CONCLUDING REMARKS

In this work, we performed a comparison from CS extraction results between two wavelet functions: db_4 and haar. For this, we used the annual time series of the solar wind velocity measured by the SOHO satellite since 1996 to 2002 years (except 1998 and 1999).

Results from power spectral shown that, the coherent part obtained from db_4 wavelet function present higher values of the spectral indices than the values obtained by haar wavelet. We applied the statistical parameter kurtosis, in CS and NC time series obtained by both wavelet functions and also for original time series. The non-coherent time series obtained by both wavelet functions showed Gaussianity characteristics (kurtosis near to 3). The results indicate that the both wavelet functions are able to extract coherent structures, however, the coherent time series showed that the db_4 wavelet function was able to extract more coherent structures than the haar wavelet. We conjecture that the good matching between the geometric form of the db_4 and the characteristics from original time series, lead to the best results. However, we noted that the NC part presented some evidence of coherent structures. We think that this fact is due to a non optimal choice of the threshold in the extraction process. Thus, in future works will study the behavior of these results as a function of changes in the threshold value.

6. ACKNOLEDGMENTS

The authors tanks to all people from CELIAS/MTOF experiment on the Solar Heliospheric Observatory (SOHO) spacecraft by a vailable data. M. J. A. Bolzan thanks to FAPESP (grant n. 2005/00511-9) and CNPq (grants n. 310431/2006-9) (Brazilian agencies) by financial support.

[1] M. J. A. Bolzan, F. M. Ramos, L. D. A. Sá, C. Rodrigues Neto, R. R. Rosa. Journal of Geophysical Research, 107(D20), doi:10.1029/2001JD000378, 2002.
[2] M. J. A. Bolzan. Brazilian Journal of Physics, 35(3A), 592-596, 2005.
[3] M. J. A. Bolzan, P. C. Vieira. Brazilian Journal of Physics, 36(4A), 1217-1222, 2006.
[4] L. F. Burlaga, W. H. Mish. Journal of Geophysical Research-Space Physics, 92(A2), 1261-1266, 1987.
[5] A. S. L. O. Campanharo, F. M. Ramos, E. E. N. Macau, R. R. Rosa, M. J. A. Bolzan, L. D. A. Sá. Philosophical Transactions of the Royal Society A-Mathematical Physical and Engineering Sciences, 366(1865), 579-589, 2008.
[6] I. Daubechies **Ten lectures on wavelets**, SIAM, Philadelphia, 1992, pp. 278-285.
[7] G. A. Degrazia, A. B. Nunes, P. Satyamurty, O. C. Acevedoa, H. F. C. Velho, U. Rizza, J. C. Carvalho. Atmospheric Environment, 41(33), 7059-7068, 2007.
[8] M. Farge, K. Schneider, N. Kevlahan. Physics of Fluids, 11(8), 2187-2201, 1999.
[9] M. Farge, N. K.-R. Kevlahan, V. Perrier, , K. Schneider. Turbulence analysis, modeling and computing using wavelets, in: J. C. van den Berg (ed.), **Wavelet in Physics**, Cambridge University Press, 117-190 (1999).
[10] U. Frisch, **Turbulence**. Cambridge: Cambridge University Press, UK, 1995.
[11] C. R. Hagelberg, N. K. K. Gamage. Boundary-Layer Meteorology, 70(3), 217-246, 1994.
[12] L. Hudgins, J. H. Kaspersen. Wavelets and detection of coherent structures in fluid turbulence, in: J. C. van den Berg (ed.), **Wavelet in Physics**, Cambridge University Press, 201-225 (1999).
[13] A. K. M. F. Hussain. Physics of Fluids, 26(10), 303-356, 1983.
[14] A. Ishizawa, Y. Hattori. Journal of Phys. Soc. Jpn, 67, 441, 1998.
[15] A. N. Kolmogorov. Journal of Fluid Mechanics, 13(1), 82-85, 1962. (see also in Proc. R. Soc. Lond. A, 434, 9-13, 1991).
[16] P. Kovács, V. Carbone, Z. Vörös. Planetary and Space Science, 49, 1219-1231, 2001.
[17] T. Li, M. Ogihara. IEEE Transactions on Multimedia, 8(3), 564-574, 2006.
[18] A. T. Y. Lui. Journal of Atmospheric and Solar-Terrestrial Physics, 64, 125-143, 2002.
[19] J. Ma, M. Y. Hussaini. Applied Physics Letters, 91, 184101, 2007.

[20] F. M. Ramos, M. J. A. Bolzan, R. R. Rosa, L. D. A. Sá. Physica D - Nonlinear Phenomena, 193, 278-291, 2004.
[21] J. E. Ruppert-Felsot, O. Praud, E. Sharon, H. L. Swinney. Physical Review E, 72(1), 1-14, 2005.