Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY
MNRAS **490,** 1879–1893 (2019) doi:10.1093/mnras/stz2668

Extreme value theory in the solar wind: the role of current sheets

Tiago F. P. Gomes,[1,2]★ Erico L. Rempel,[1,2] Fernando M. Ramos,[1] Suzana S. A. Silva[2] and Pablo R. Muñoz[3]

[1]*National Institute for Space Research – INPE, Av. dos Astronautas 1758, 12227-010 São José dos Campos, SP, Brazil*
[2]*Institute of Aeronautical Technology – ITA, Praça Marechal Eduardo Gomes 50, 12228-900 São José dos Campos, SP, Brazil*
[3]*Department of Physics and Astronomy, University of La Serena, Av. Juan Cisternas 1200, La Serena, Chile*

Accepted 2019 September 9. Received 2019 September 5; in original form 2019 May 5

ABSTRACT

This article provides observational evidence for the direct relation between current sheets, multifractality and fully developed turbulence in the solar wind. In order to study the role of current sheets in extreme-value statistics in the solar wind, the use of magnetic volatility is proposed. The statistical fits of extreme events are based on the peaks-over-threshold (POT) modelling of *Cluster 1* magnetic field data. The results reveal that current sheets are the main factor responsible for the behaviour of the tail of the magnetic volatility distributions. In the presence of current sheets, the distributions display a positive shape parameter, which means that the distribution is unbounded in the right tail. Thus the appearance of larger current sheets is to be expected and magnetic reconnection events are more likely to occur. The volatility analysis confirms that current sheets are responsible for the −5/3 Kolmogorov power spectra and the increase in multifractality and non-Gaussianity in solar wind statistics. In the absence of current sheets, the power spectra display a −3/2 Iroshnikov–Kraichnan law. The implications of these findings for the understanding of intermittent turbulence in the solar wind are discussed.

Key words: turbulence – methods: data analysis – methods: statistical – solar wind.

1 INTRODUCTION

The goal of the field of space weather is to monitor and predict events related to solar activity and the interplanetary medium and their impact on Earth (Hanslmeier 2007; Riley 2012). Those events include solar flares, coronal mass ejections and the solar wind and are responsible for phenomena such as magnetic storms and substorms (e.g. Gonzalez et al. 1994), which can lead to interference in transportation, energy, telecommunications and navigation systems (Hathaway 2015).

The solar wind is particularly important in space weather studies, due to its continuous interaction with the Earth's magnetosphere. It is known that the solar wind is permeated by coherent structures such as magnetic flux ropes and current sheets (Chian et al. 2016). The latter constitute the main factor responsible for multifractality and departure from Gaussianity in the statistics of magnetic field time series (Chian & Miranda 2009), as well as intermittent turbulence (Bruno & Carbone 2013). Such phenomena are closely related to the occurrence of extreme events, which are events with high amplitude in comparison with the distribution's mean. For instance, the non-linear interactions of the magnetic field in the solar wind at small scales lead to the creation of intense current sheets, which can be associated with magnetic reconnections (Boozer 2014; Osman et al. 2014; Greco et al. 2018). Such reconnections accelerate particles, causing high energy release in the space plasma (e.g. Gosling 2012). The statistical analysis of extreme events has been widely employed in studies of natural phenomena such as wind velocity (Fawcett & Walshaw 2006), droughts and floods (Engeland, Hisdal & Frigessi 2004), earthquakes (Ogata 1988) and diseases (Thomas et al. 2016). Extreme-event theory has also been used in space weather science, where it was applied to solar flares (Tsiftsi & De la Luz 2018), geomagnetic storms (Tsubouchi & Omura 2007), solar wind–magnetosphere coupling (Moloney & Davidsen 2010) and sunspot numbers (Acero et al. 2017, 2018). Extreme-value theory (Coles 2001) comprises a set of advanced techniques and statistical tools to model the distribution functions of time series displaying extreme events. The present work presents the application of extreme-value theory to the solar wind, focusing on the role of current sheets in determining the distribution shapes.

The relation of current sheets to the statistics of extreme events and turbulence in space plasmas can be better appreciated by an analysis of intermittency. We consider two models of intermittent turbulence, the Kolmogorov (K41) and the Iroshnikov–Kraichnan (IK). Originally, the K41 theory by Kolmogorov (1941) predicted universality and scale-invariance in a given frequency range in a turbulent spectrum, the inertial range. Inside this range, the

★ E-mail: tiagofpgomes@gmail.com

turbulence loses geometrical information about the boundary conditions related to the largest scales (integral scale), resulting in a statistical self-similarity between scales. This universality suggested by Kolmogorov was strongly criticized by Landau, for whom small-scale fluctuations could be affected by perturbations in the integral scale (Frisch 1995); furthermore, strong velocity gradients should be confined to very limited regions of the flow, reflecting the presence of intermittency and multifractality. In 1962, Kolmogorov added intermittency and multifractality to his theory (Kolmogorov 1962), refining his global-scale invariance hypothesis to a local-scale invariance.

The first model to describe turbulent processes in the presence of magnetic fields was presented by Iroshnikov (1964) and Kraichnan (1965). In it, an external isotropic magnetic field is considered and perturbations propagate as waves along field lines with the Alfvén speed (Alfvén 1942), where two waves can travel in the same or opposite directions. In the case of counter-propagation, instabilities and non-linear interactions can take place (Parker 1979), so that in the IK model turbulence is guided by the collision of Alfvén waves. Thus, while the energy cascade in the Kolmogorov model occurs due to vortex–vortex interactions, in the IK model the energy transfer is given by collisions of Alfvén wave packets. Despite the presence of a background magnetic field, the IK model is isotropic, which has led to some criticism (Matthaeus et al. 1996; Montgomery & Matthaeus 1995; Sridhar & Goldreich 1994). Alternative models have included anisotropy for the magnetic field (e.g. Goldreich & Sridhar 1997).

Both the K41 and IK power-spectrum laws are observed in solar wind data (Salem et al. 2009), but most works point to a Kolmogorov ($-5/3$) scaling law for energy spectra (Chian & Miranda 2009; Chian & Munoz 2011), while the IK scaling ($-3/2$) is seldom found (Li et al. 2011). This is probably due to the fact that current sheets are very frequent in the solar wind. There is evidence that current sheets are the coherent structures responsible for the K41 scaling law and in their absence the solar wind displays an IK scaling law (Li et al. 2011).

Due to the possible impacts of current sheets in space plasma dynamics, understanding the physics underlying the creation and evolution of such structures is an important step to describe solar wind turbulence. Several works have sought methods to detect discontinuities in the solar wind magnetic field caused by current sheets. For instance, in Chian & Miranda (2009) a time series of normalized magnetic-field differences was used, while in Greco et al. (2009, 2018) the partial variance of increments method was adopted. In the present work, we propose a new technique based on the volatility (Poon 2005), a statistical tool widely employed to detect and predict sudden changes in financial markets (Tsay 2010). This analysis is based on so-called return time series (Tsay 2010), which present interesting statistical properties such as stationarity and ergodicity. The measure of the bursts of those return time series is called volatility (Poon 2005) and displays the clustering (or memory) effect (Tsay 2010). In other words, the volatility at a certain time depends on the volatility at previous times. Therefore, it is common to model and even predict the volatility of financial series by using auto-regressive models for variance, such as, for example, the autoregressive conditional heteroskedasticity (ARCH) model (Engle 1982).

Financial time series present some key characteristics found in solar wind time series, namely chaoticity (Hołyst, Żebrowska & Urbanowicz 2001), multifractality, extreme events (Green, Hanan & Heffernan 2014) and turbulence (Voit 2005), which encourages the use of the same tools in the analysis of solar wind data. Loosely, the effects of current sheets and magnetic reconnections in the solar wind are qualitatively analogous to the effects of economic crises in financial time series.

In Section 3 we describe our solar wind data and the pre-processing steps necessary for statistical analysis. The methodology is presented in Section 2, where a description of the generalized Pareto distribution is given, as well as methods for choosing a threshold for extreme events and for estimating the GPD parameters. We also present Li's method for current-sheet detection and explain how to compute the zeta function for studying the multifractal nature of fully developed turbulence. The results of the time series analysis are presented in Section 4 and a discussion and conclusions are given in Section 5.

2 METHODOLOGY

2.1 Generalized Pareto distribution (GPD)

Two of the most popular techniques to model extreme values are block maxima and the peaks-over-threshold (POT) technique (Coles 2001). In the block maxima method, a time series is divided into a number of sub-blocks and the maxima of each block are chosen as the new random variables to compute the extreme event statistics, the distribution function of which is given by (Jenkinson 1995; Fisher & Tippett 1928)

$$G(z) = \exp\left\{-\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]^{\frac{-1}{\xi}}\right\}, \tag{1}$$

defined on $\{z: 1+\xi(z-\mu)/\sigma > 0\}$, where $-\infty < \mu < \infty$ is the location parameter, $\sigma > 0$ is the scale parameter and $-\infty < \xi < \infty$ is the shape parameter. Equation (1) is known as the generalized extreme values (GEV) distribution (Coles 2001). Although simple and robust, this type of modelling may leave out a number of extreme events that belong to the same block, since only the maximum is considered. Alternatively, the POT technique considers all events above a certain threshold, thus avoiding missing important peaks in the analysis. In this work, the POT is employed and its variables are modelled by the generalized Pareto distribution (GPD: Coles 2001).

Let $(X_i: i = 1, ..., n)$ be a sequence of independent and identically distributed (iid) random variables with a distribution function F. One can consider the events, X, from this set that are above a given threshold u $(X > u)$ and define a new random variable $Y = X - u$, known as the threshold excess, for which the conditional probability is given by

$$\begin{aligned}\Pr(X > u+y | X > u) &= \frac{\Pr[(X > u+y) \cap (X > u)]}{\Pr(X > u)} \\ &= \frac{\Pr[X > u+y]}{\Pr(X > u)} \\ &= \frac{1-F(u+y)}{1-F(u)} \\ &= 1 - H(y), \qquad y > 0,\end{aligned} \tag{2}$$

where $H(y)$ is the distribution function of $y = x - u$. If F were known, then the distribution function of the threshold excess would be easily determined, but generally that is not the case. According to Pickands (1975) and Balkema & de Haan (1974), if the block maxima time series converges into a GEV distribution, then the threshold excess time series, for u sufficiently large, must converge

to the following distribution:

$$H(y) = \begin{cases} 1 - \left(1 + \frac{\xi y}{\widetilde{\sigma}}\right)^{-1/\xi}, & \xi \neq 0, \\ 1 - \exp\left(-\frac{y}{\widetilde{\sigma}}\right), & \xi \to 0, \\ \frac{y}{\widetilde{\sigma}}, & \xi \to -1, \end{cases} \tag{3}$$

defined over $y > 0$ and $\left[1 + \left(\xi y / \widetilde{\sigma}\right)\right] > 0$, where

$$\widetilde{\sigma} = \sigma + \xi (u - \mu), \tag{4}$$

in which (σ, μ, ξ) is the same set of parameters as in the GEV distribution (Coles 2001).

The family of distributions defined by (3) is the generalized Pareto distribution family. The scale parameter $\widetilde{\sigma}$ measures the variability of the series, while the shape parameter ξ is responsible for characterizing the tail of the distribution. It is important to emphasize that ξ is the same for both GEV and GPD and, if $\xi > 0$, then the distribution does not have an upper bound, that is, it has infinite support given by $y =]0, \infty[$. Otherwise, if $\xi < 0$, the distribution has an upper bound and finite support $y = \left]0, u - \left(\widetilde{\sigma}/\xi\right)\right]$ (Coles 2001). In the cases $\xi \to -1$ and $\xi \to 0$, the GPD reduces, respectively, to the uniform distribution $U(0, \widetilde{\sigma})$ and the exponential distribution with mean equal to $\widetilde{\sigma}$, as in (3).

2.1.1 *Threshold selection*

The choice of threshold for the GPD must be made in such a way as to balance bias and variance. It should not be too low, as this could select non-extreme values, which would result in bias. Conversely, it should not be too high, as selecting very few extreme values increases the variance of the estimator. In classical extreme events theory, there are two techniques for threshold choice. The first one, introduced by Davison & Smith (1990), consists of a graphical method based on the linearity of mean threshold excesses, also called the mean excess (Coles 2001). Although useful, it may be difficult to interpret the mean residual life plot, thus hampering the choice of an adequate threshold. Therefore, a complementary method is also used, based on the stability of parameters. If the u_0 threshold excesses follow a GPD with parameters ξ and σ, then the excesses of a threshold $u > u_0$ will also follow a GPD with the same shape parameter ξ (Coles 2001). However, due to (4), the scale parameter is given by

$$\sigma_u = \sigma_{u_0} + \xi (u - u_0). \tag{5}$$

Therefore, for $\xi \neq 0$ the scale parameter varies with u. However, reparametrizing the scale parameter as $\sigma^* = \sigma_u - \xi u$ and considering equation (5), we obtain

$$\sigma^* = \sigma_{u_0} - \xi u_0, \tag{6}$$

which is constant. In this case, $\sigma^*(u)$ is called the reparametrized scale parameter. As a consequence, the estimates for σ^* and ξ must be approximately constant for a correct GPD modelling. As with the mean residual life, an adequate value for the threshold must be the least value inside a range of thresholds for which σ^* and ξ are approximately constant.

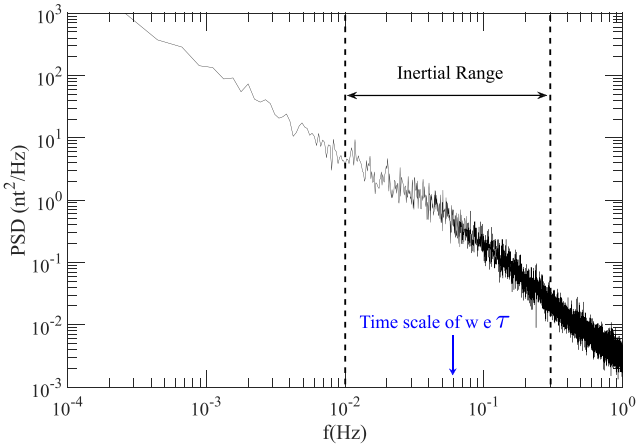

Figure 1. Power spectral density for 2008 March 9 time series. The approximately linear region of the PSD between dashed lines is the inertial range. The time-scale τ for the volatility calculation was chosen from the midpoint of the inertial range.

2.1.2 *Parameter estimation*

There are currently several techniques available for parameter estimation for the GPD (Hosking & Wallis 1987; Coles 2001; Luceño 2006; Zhang 2007, 2010; Park & Kim 2016). In the present work, we adopt maximum-likelihood estimation (MLE: Coles 2001), since it is the most popular, flexible and general method.

Let $(Y_i: i = 1, ..., n)$ be independent random variables, given by a u threshold excess $(Y_i = X_i - u)$ and following a GPD. The log-likelihood function is defined as

$$l(\sigma, \xi) = \begin{cases} -n \log \sigma - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^{n} \log\left(1 + \frac{\xi y_i}{\sigma}\right), & \xi \neq 0, \\ -n \log \sigma - \frac{1}{\sigma} \sum_{i=1}^{n} y_i, & \xi \to 0, \end{cases} \tag{7}$$

when $(1 + \xi y_i/\sigma) > 0$ for $i = 1, ..., n$ and $l(\sigma, \xi) = -\infty$ otherwise. In this case, the maximum-likelihood estimators $(\hat{\sigma}, \hat{\xi})$ of (σ, ξ) are defined as the values that maximize the corresponding log-likelihood function $l(\sigma, \xi)$ (Coles 2001).

2.1.3 *Return-level estimation*

Given a GPD with parameters ξ and σ modelling a random variable X exceeding a threshold u, from equations (2) and (3) we have (Coles 2001)

$$\Pr(X > x_m) = \frac{1}{m} = \begin{cases} \varsigma_u \left[1 + \frac{\xi(x_m - u)}{\sigma}\right]^{-1/\xi}, & \xi \neq 0, \\ \varsigma_u \exp\left[\frac{-(x_m - u)}{\sigma}\right], & \xi \to 0, \end{cases} \tag{8}$$

or, in terms of x_m,

$$x_m = \begin{cases} u + \frac{\sigma}{\xi}\left[(m\varsigma_u)^{\xi} - 1\right], & \xi \neq 0, \\ u + \sigma \ln(m\varsigma_u), & \xi \to 0, \end{cases} \tag{9}$$

where $\varsigma_u = Pr(X > u)$ and x_m is the m-quantile, which expresses the magnitude of an event that occurs with a probability $1/m$, i.e. that on average occurs every m time units. Such an x_m is called the return level and m is the return period.

Usually, one plots x_m as a function of the logarithm of m. According to equation (9), the plot should be approximately linear for $\xi = 0$, concave for $\xi > 0$ and convex for $\xi < 0$.

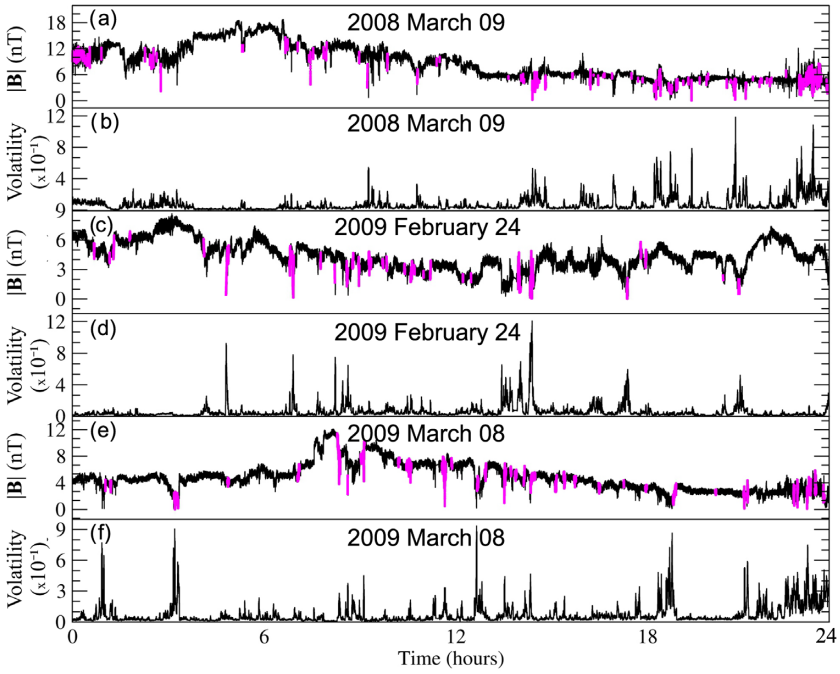

Figure 2. Time series of the magnitude of the magnetic field in the solar wind and the respective volatilities in the presence of current sheets (magenta/gray) for (a, b) 2008 March 9, (c, d) 2009 February 24 and (e, f) 2009 March 8.

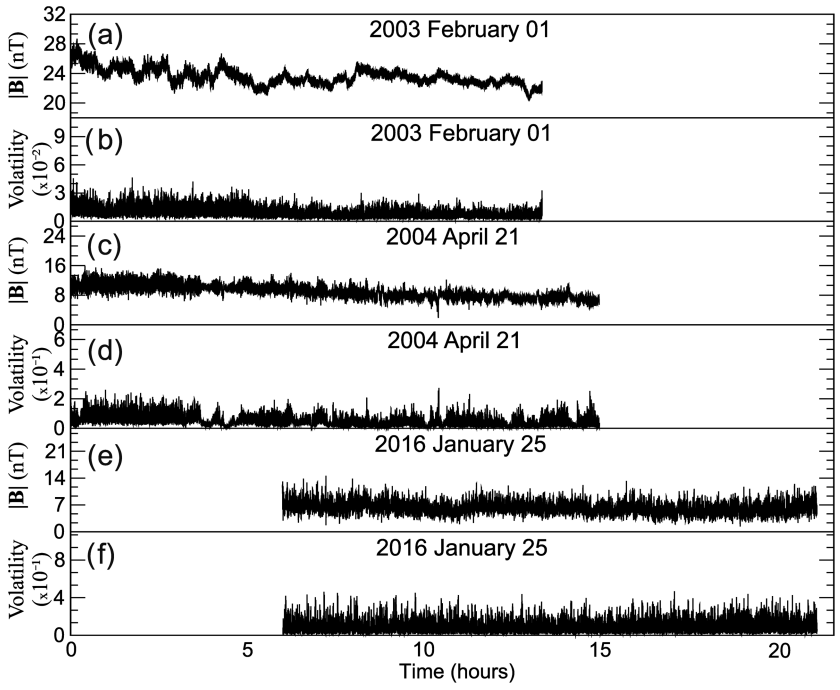

Figure 3. Time series of the magnitude of the magnetic field in the solar wind and the respective volatilities in the absence of current sheets for (a, b) 2003 February 1, (c, d) 2004 April 21 and (e, f) 2016 January 25.

2.2 Current-sheet detection

Current sheets are coherent structures located in approximately two-dimensional regions associated with a rotation of the magnetic field vectors. Li's method (Li 2008) is a robust and accurate way to detect such regions.

Consider a magnetic field time series, $\boldsymbol{B}(t)$. According to Li (2008), for a given time lag τ, the probability density of finding the rotation angle between vectors $\boldsymbol{B}(t+\tau)$ and $\boldsymbol{B}(t)$ in the interval $[\theta, \theta+\Delta\theta]$ is

$$f(\theta,\tau)\,\Delta\theta = \frac{N^{\tau}\left(\theta < \theta' < \theta+\Delta\theta\right)}{N^{\tau}\left(0 < \theta' < \pi\right)}, \tag{10}$$

where $N^{\tau}(\theta < \theta' < \theta+\Delta\theta)$ is the number of pairs of vectors $\boldsymbol{B}(t+\tau)$ and $\boldsymbol{B}(t)$ for which the rotation angles are between θ and

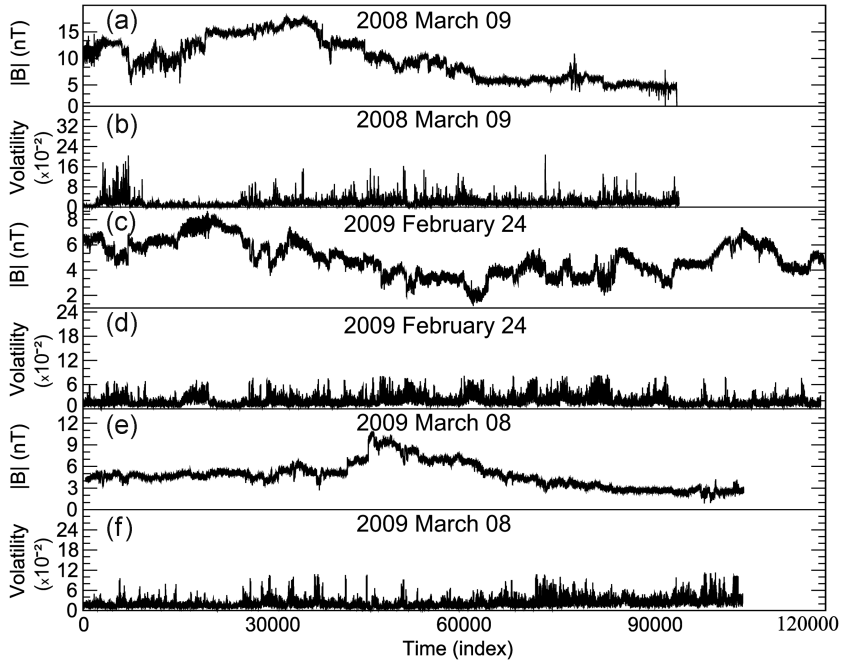

Figure 4. Time series of the magnitude of the magnetic field in the solar wind and the respective volatilities corresponding to Fig. 2, after removing the current sheets.

$\theta + \Delta\theta$ and $N^{\tau}(0 < \theta^{'} < \pi)$ is the total number of measurements. In this case, the distribution function given by

$$F(\theta, \tau) = \int_{\theta}^{\pi} \mathrm{d}\theta' f\left(\theta', \tau\right) = \frac{N^{\tau}\left(\theta < \theta' < \pi\right)}{N^{\tau}\left(0 < \theta' < \pi\right)} \tag{11}$$

is simply the frequency of occurrence of rotation angles greater than θ (Li 2008). If there are current sheets in the time series of $\boldsymbol{B}(t)$, then $F(\theta, \tau)$ must obey a linear scaling law with τ when θ is larger than a given critical angle θ_{c} (Li 2008), that is

$$F(\theta, N\tau) \sim NF(\theta, \tau) \quad \text{when} \quad \theta > \theta_{\mathrm{c}}. \tag{12}$$

In this case, for each window of length 2τ[1] centred at time T, we check the number of angles between $\boldsymbol{B}(t+\tau)$ and $\boldsymbol{B}(t)$ that are larger than θ.[2] If this quantity is larger than a certain threshold (L),[3] then a current sheet is found in T.

2.3 Zeta function

For a given dynamical variable $\phi(t)$, the structure function is defined as (Frisch 1995)

$$S_p(\tau) = \langle \delta_{\tau}^{p}(t) \rangle = \langle [\phi(t+\tau) - \phi(t)]^{p} \rangle, \tag{13}$$

where $\langle ... \rangle$ represents the time or space average, p are the statistical moments and τ is a time or space lag. Given Kolmogorov's similarity hypothesis and scale-invariance ($\delta\phi(\tau) \propto \tau^{h}$: Kolmogorov 1941, 1962), inside the inertial range the order p structure function follows a power law $S_p(\tau) \propto \tau^{\zeta(p)}$, where $\zeta(p) = ph$ is the scaling exponent or zeta function (Frisch 1995). In this case, $\zeta(p)$ is the slope of the $\log S_p(\tau) \times \log \tau$ plot. The set of ζ values computed for each statistical moment p[4] produces a zeta function curve that is related intimately to several monofractal and multifractal models of turbulence (Frisch, Sulem & Nelkin 1978; Burlaga 1991; She & Leveque 1994; Politano & Pouquet 1995), in which the linearity reflects the presence of the self-similarity or monofractality predicted by the K41 theory. Nonetheless, departures from this linearity are commonly observed in space missions (see e.g. Chian & Miranda 2009; Chian & Munoz 2011), indicating the presence of intermittency or multifractality in the turbulence.

3 ORIGINAL DATA AND PRE-PROCESSING

The solar wind data employed in this work consist of the magnitude of the magnetic field collected at 22 Hz by the Fluxgate Magnetometer (FGM) on board *Cluster-1*, near to the Earth's bow shock. The data were obtained from the European Space Agency (ESA) website[5] during periods when the spacecraft was outside the Earth's magnetosphere, where direct contact with the solar wind environment was possible. Our goal is to compare time series where current sheets are present with time series where current sheets are absent, so we checked all available data from 2001–2016. It is extremely difficult to find long time intervals where current sheets are absent in the solar wind; thus only three current sheet-free time series were detected: 2003 February 1, 2004 April 21 and 2016 January 25, all with a duration of more than 10 and less than 16 hours. We also selected three 24-hour time series with current sheets: 2008 March 9, 2009 February 24 and 2009 March 8.

To detect the current sheets, we verified a linear scaling of the distribution function $F(\theta, \tau)$ (equation 11) for time-scales between 30 and 240 s and also for critical angles between 60° and 120°, from which we adopt a time-scale $T = 120$ s and a critical angle $\theta = 60°$. Our data are pre-processed with a low-pass filter followed by a decimation process, which reduces noise and the length of the time

[1] In this work, $\tau = 200$ s.
[2] In this work, $\theta = 60°$.
[3] In this work, $L = 60$ per cent.
[4] In this work, $p = 1, ..., 6$.
[5] https://www.cosmos.esa.int/web/csa

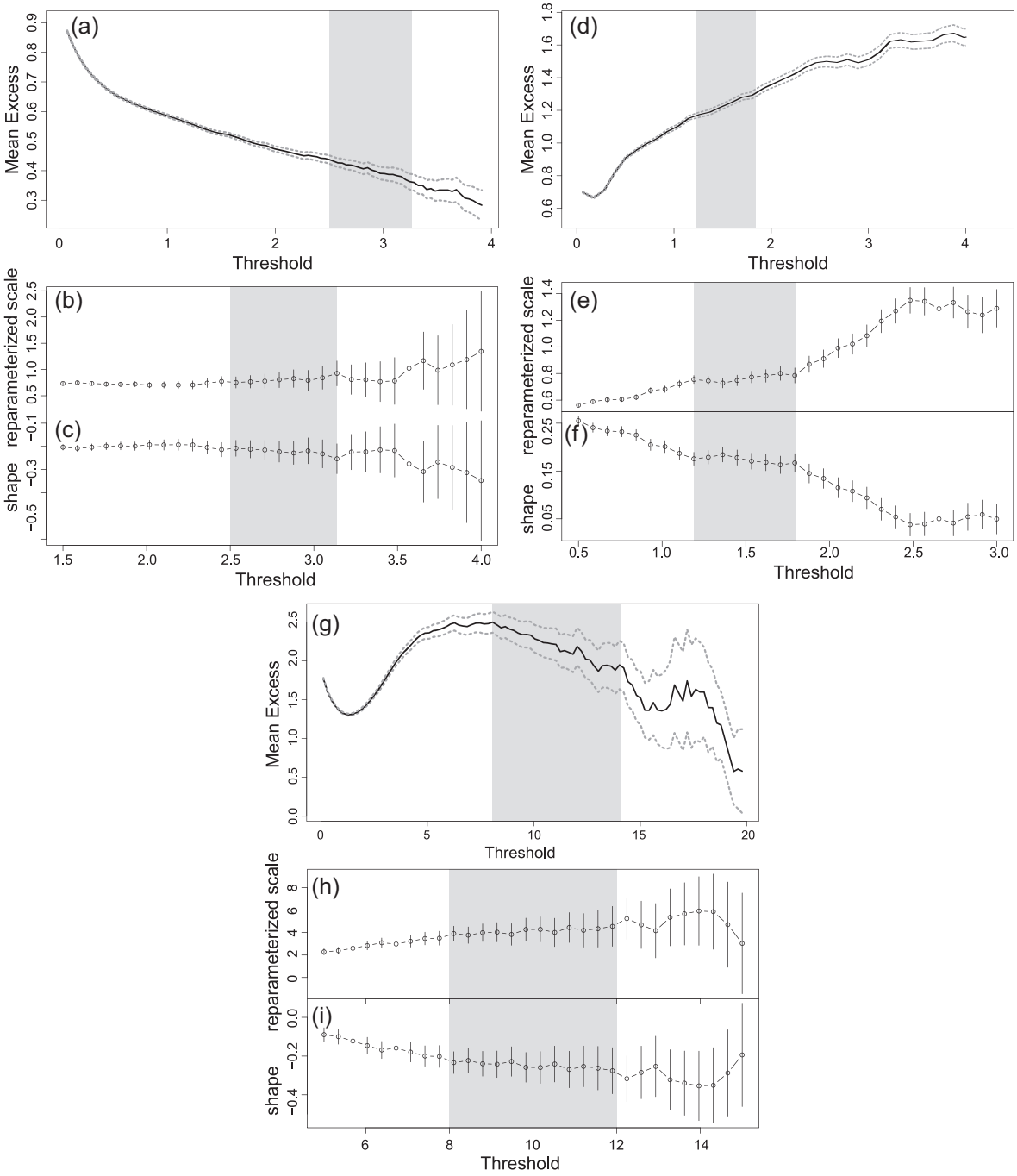

Figure 5. Threshold selection diagnostic plots. Mean excess plot, reparametrized scale parameter and shape parameter estimates against the threshold for volatility time series for (a, b, c) 2016 January 25, (d, e, f) 2008 March 9 and (g, h, i) 2008 March 9, after removing the current sheets. The dashed lines and solid bars correspond to 95 per cent confidence limits. The shaded regions mark the intervals for which the mean residual life is approximately linear and the shape parameter and the reparametrized scale are approximately constant.

series by a factor $M = 10$. We adopt MATLAB's *decimate* function, which implements a low-pass Chebyshev Type I infinite impulse response (IIR) filter (Oppenheim & Verghese 2017) of order 8 and cut-off frequency of $0.8/M$, which guarantees that we are free of aliasing effects.

Sometimes, current sheets are correlated with local minima in time series of the magnitude of the magnetic field. However, that is not always the case and a statistical analysis of extreme events based solely on the time series of $|\boldsymbol{B}|$ may be inappropriate. Hence, another method is needed to detect all current sheet regions without losing the basic statistical properties of the original series.

Since rotations of the magnetic field vector generate a current density, intense current sheets are usually associated with large rotations of $\boldsymbol{B}$, resulting in large fluctuations or *bursts* in time series of two-point differences of the form

$$r_{\mathrm{mag}}(t) = |\boldsymbol{B}(t+\tau)| - |\boldsymbol{B}(t)|\,, \tag{14}$$

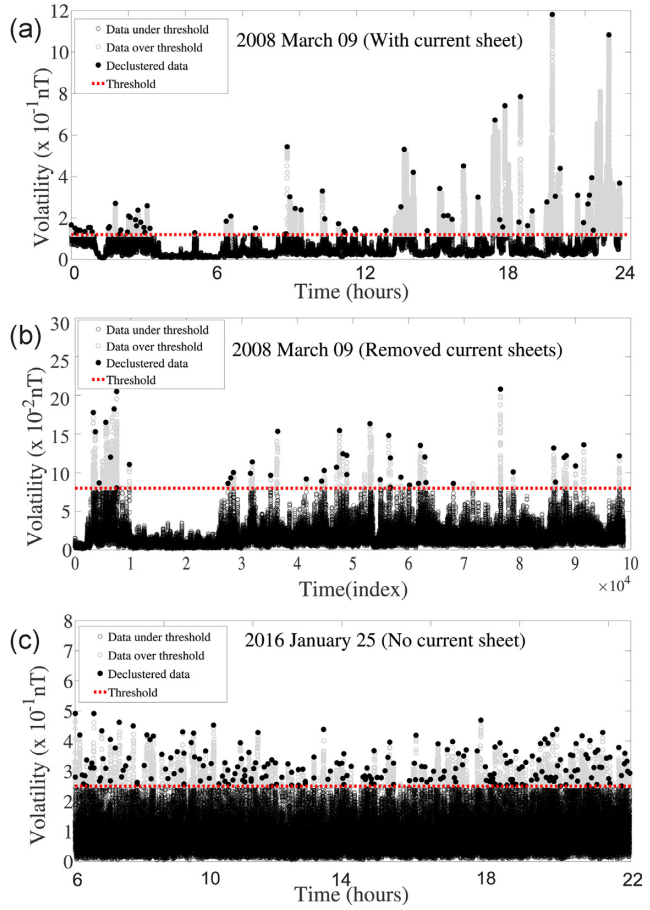

Figure 6. Declustered time series for the volatility of the magnitude of the magnetic field for (a) 2008 March 9, (b) 2008 March 9 after removing the current sheets and (c) 2016 January 25. The dashed lines indicate the threshold and the highlighted points above the threshold represent the new time series obtained after declustering.

where τ is a time lag. In the context of financial sciences, the bursts are known as *volatility* (Poon 2005; Tsay 2010), which is usually associated with violent changes in the financial market, such as a financial crisis. The two-point difference series is known as the financial return (Tsay 2010), therefore we call $r_{\rm mag}$ the *magnetic return*.

The volatility can be measured using the standard deviation or the sample variance. Usually, the standard deviation is more stable than the variance and has the same dimension as the sample mean (Poon 2005). For those reasons, we perform our volatility calculations using the standard deviation in this work. The magnetic volatility can be measured from the standard deviations in different periods of the time series of $r_{\rm mag}$ or, alternatively, from the standard deviations of the log magnetic return (Tsay 2010):

$$\Delta r_{\rm mag}\,(t) = \log\left(|\boldsymbol{B}\,(t+\tau)|/|\boldsymbol{B}\,(t)|\right). \tag{15}$$

In our case, the use of the log magnetic return is preferred, since the series of logarithms is better behaved and data visualization is easier with logarithms when the data present high variability.

We compute the standard deviations in a moving window of length w over the time series of $\Delta r_{\rm mag}$, then obtain the magnetic volatility for a time series with N samples as

$$vol_{\rm mag}\,(j) = \sqrt{\frac{1}{w-1}\sum_{i=j}^{w+j-1}\left(\Delta r_{\rm mag}(i) - \mu(j)\right)^2}, \tag{16}$$

where $j = 1, ..., N - w + 1$ controls the position of the moving window in the time series and $\mu(j)$ is the mean $\Delta r_{\rm mag}$ inside the window. The values of w and τ are chosen to be the same order of magnitude as the length of the high-energy current sheets, which are responsible for the extreme events in our data. Their length is estimated from the power-spectrum density (PSD), computed with the Welch method (Welch 1967) with a Hanning window (Paschmann & Daly 1998). In general, the main current sheets are located in the intermediate region of the PSD inertial range. Thus, we look for a time-scale in the middle of the inertial range. From Fig. 1, we choose $\tau = w = 50$ s, since $\tau = 1/f$.

4 RESULTS

4.1 Choice of threshold and declustering process

Fig. 2 shows time series of the magnitude of the magnetic field in the solar wind and the respective volatilities for 2008 March 9, 2009 February 24 and 2009 March 8. The magenta (gray) lines indicate regions where current sheets were detected by Li's method (Li 2008). There is a strong correlation between regions where current sheets were detected and the occurrence of extreme events in the volatility. This is to be expected, since strong current sheets are associated with sudden rotations of the magnetic field vector and the volatility is a parameter that measures abrupt variations in a time series. Thus, the volatility consists of an excellent filter to amplify extreme events.

In order to identify the role of current sheets in the statistics of extreme events in the solar wind, we selected three additional time series in which current sheets were not detected. Although we checked every time series from 2001–2016 in which *Cluster-1* was near to the bow shock, we could not find any 24-h time series in which current sheets were absent. Nonetheless, we found three time series free of current sheets (at the time-scale considered) with a duration of approximately 15 hours, as shown in Fig. 3. Although these time series are shorter than in Fig. 2, this was not problematic for the statistical analysis conducted below. Furthermore, three other time series were extracted from the data in Fig. 2 by removing all their current sheets. They are shown in Fig. 4. It can be seen from Figs 3 and 4 that, in the absence of current sheets, the volatility time series do not reveal extreme values.

Before conducting an extreme-events analysis by the POT technique, we need to establish the threshold that will define extreme events, conduct data declustering and then check for data independence.

The POT technique analyses all sample points that exceed a certain threshold u and the modelling of the probability distribution of the exceeding points can be done by the GPD. Here, we adopt the two methods mentioned in Section 2.1.1 to choose the threshold. The first method uses the linearity of the mean residual life (or mean excess) as a function of u, where one chooses the threshold from an interval in which the mean residual life displays an approximately linear behaviour. The second method is based on the stability of the shape parameter and the reparametrized scale, where the threshold value is taken from an interval of u in which the shape parameter and the reparametrized scale are approximately constant. These two methods are illustrated in Fig. 5 for the volatility of the time series

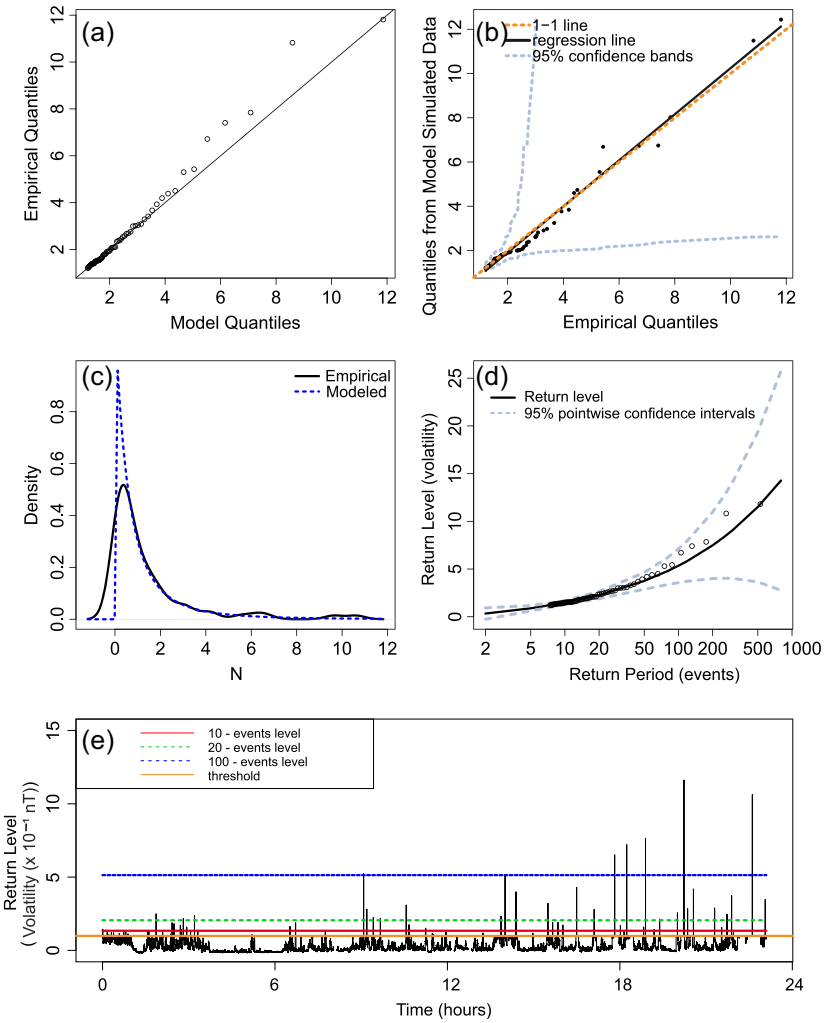

Figure 7. Diagnostic plots for GPD model fit for the volatility time series of 2008 March 9. (a) *QQ* plot of empirical data quantiles in relation to fitted data quantiles obtained from the GPD model; (b) *QQ* plot of random data quantiles, obtained from the GPD model in relation to empirical data quantiles. The dotted lines represent the 95 per cent confidence interval; the dashed line represents the 1–1 line; the solid line represents the linear regression; (c) probability density functions of empirical data (solid line) and the adjusted GPD model (dashed line); (d) return levels (solid lines) as a function of the return periods (in log scale) with 95 per cent pointwise confidence interval (dashed lines); (e) return levels and periods indicated in the time series. The solid orange (light gray) line represents the threshold; the solid red (dark gray) line indicates the level to be achieved once at each occurrence of two events; the dashed green line indicates the level to be achieved once at each occurrence of 20 events; the dotted blue line indicates the level to be achieved once at each occurrence of 100 events.

with current sheets of 2008 March 9 (see Fig. 2b), the time series without current sheets of 2016 January 25 (see Fig. 3f) and the 2008 March 9 time series after removing the current sheets (see Fig. 4b). The shaded regions mark the intervals for which the mean residual life is approximately linear and the shape parameter and the reparametrized scale are approximately constant. The threshold parameter is chosen as the smallest value in the shaded areas, thus $u = 2.5$ in Figs 5(a), (b) and (c), which corresponds to the 97.53th percentile of the volatility time series; $u = 1.2$ in Figs 5(d), (e) and (f), which corresponds to the 83.36th percentile, and $u = 8$ in Figs 5(g), (h) and (i), corresponding to the 98.88th percentile. We stress that the linearity of the mean residual life and the stability of the shape parameter are necessary but not sufficient conditions for an adequate choice of u. For instance, in Figs 5(a), (b) and (c), the interval $u = [2.0, 2.5]$ exhibits an approximately linear mean residual life and the shape parameter is approximately constant. Nonetheless, this choice of interval provides a threshold $u = 2.0$ that results in the inclusion of non-extreme values, leading to bias in the distribution. Therefore, the choice of u must take into account both aforementioned methods, while still avoiding bias by values of u that are too low or increased variance by values that are too high.

Volatility time series often display the formation of clusters of contiguous points that exceed the given threshold. The problem with such clusters is that they increase the dependence among data points, thus violating the basic assumption of data independence for the correct application of the statistical modelling of extreme events. Therefore, a process of declustering may be necessary to decrease the short-term memory of the series. One declustering method that can be used is 'runs declustering' (Leadbetter 1983), which consists of choosing a run length, r, such that values that exceed the threshold and are separated by a number of observations smaller than r that are considered to belong to the same cluster. Subsequently, a new sample is obtained by selecting the maximum value of each cluster. The choice of r must be large enough to avoid data dependence and small enough to avoid missing too many extreme values. Fig. 6 illustrates the declustering process for the

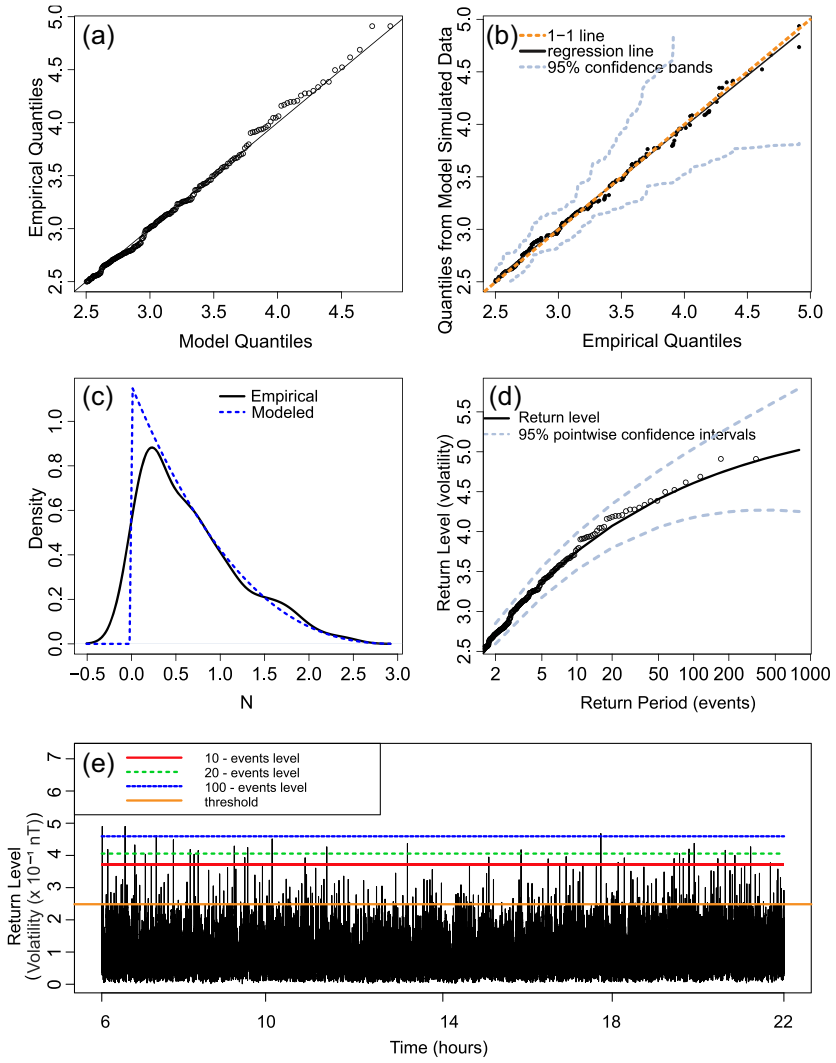

Figure 8. Diagnostic plots for GPD model fit for the volatility time series. (a)–(e) The same as in Fig. 7, but for 2016 January 25.

volatility of the three time series, (a) 2008 March 9, (b) 2008 March 9 after removing the current sheets and (c) 2016 January 25. The new samples correspond to the highlighted points above the chosen thresholds (dashed lines). The degree of dependence among points in a sample can be measured with the 'extremal index', θ (see Leadbetter 1983 and Coles 2001), where $0 \leq \theta \leq 1$ and $\theta = 1$ for fully independent time series.

Additionally, we also conducted the Ljung–Box Q-statistics test (Ljung & Box 1978), which is based on the hypothesis

$$\begin{cases} \mathbf{H_o} : \text{the data are not correlated;} \\ \mathbf{H_a} : \text{the data are correlated;} \end{cases}$$

for which the Q-statistics test is given by

$$Q = n(n+2)\sum_{k=1}^{m} \frac{\hat{\rho}_k^2}{n-k}, \tag{17}$$

where n is the length of the time series, $\hat{\rho}_k^2$ is the estimated autocorrelation at lag k and m is the number of lags being tested. We verified that the Ljung–Box Q-statistics test does not exceed the critical values for the significance level at 5 per cent. Therefore, according to the Ljung–Box Q-statistics test, the null hypothesis of non-correlation of the data after declustering is observed, confirming the efficiency of the declustering technique.

4.2 GPD analysis

The selected time series are data-independent, as confirmed by standard tests, and therefore we can proceed with application of the generalized Pareto distribution (GPD) modelling. Figs 7, 8 and 9 show the diagnostic plots of GPD models fitted on the declustered time series for the volatility of $|\boldsymbol{B}|$ on 2008 March 9, 2016 January 25 and 2008 March 9 (after removing the current sheets), respectively. The QQ plots of empirical data quantiles against GPD's theoretical quantiles are shown in Figs 7(a), 8(a) and 9(a), where the vertical axes represent the empirical quantiles and the horizontal axes show the quantiles from model-simulated data, obtained from equation (3). The QQ plots in Figs 7(a), 8(a) and 9(a) reveal an excellent fit of the time series by a GPD, as confirmed by the histograms in Figs 7(c), 8(c) and 9(c).

The QQ plot points in Fig. 7(a) still have a satisfactory fit, even though they are not as well aligned with the diagonal as in Figs 8(a) and 9(a). This departure from the diagonal can be explained by the fact that the tail of the distribution for the empirical data is heavier

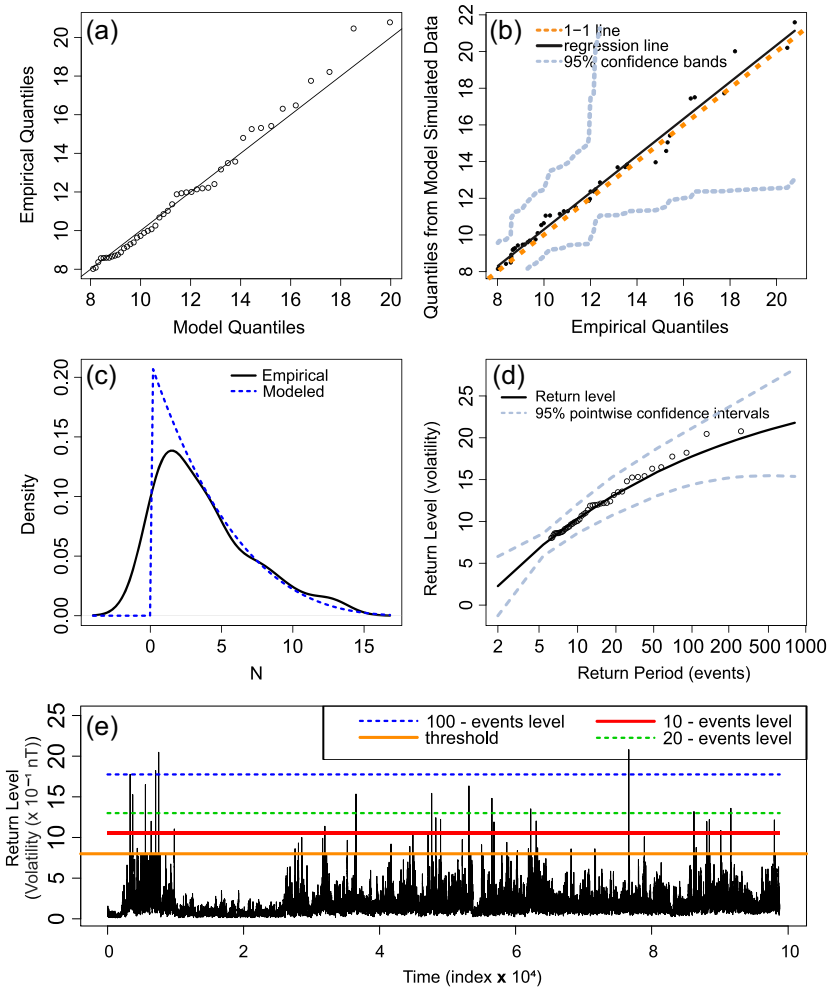

Figure 9. Diagnostic plots for GPD model fit for the volatility time series. (a)–(e) The same as in Fig. 7, but for 2008 March 9 after removing the current sheets.

than the tail predicted by the theoretical GPD. The histograms displayed in Fig. 7(c) support that explanation, as the empirical density plot (solid black line) has a heavier tail than the fitted GPD density (blue dashed line).

Figs 7(b), 8(b) and 9(b) display the *QQ* plots of the randomly generated data quantiles from the fitted GPD against the empirical data quantiles with 95 per cent confidence bands. The horizontal axes in Figs 7(b), 8(b) and 9(b) represent the observational data empirical quantiles and the vertical axes represent the quantiles of randomly generated GPD data. The almost exact superposition of the 1–1 lines and the linear regression lines demonstrates the quality of the statistical fits.

The return-level plots with 95 per cent pointwise normal approximation confidence intervals for the three time series are shown in Figs 7(d), 8(d) and 9(d). The horizontal axes display the return periods on a logarithmic scale, while the associated return levels are found in the vertical axes. The return-level curve in Fig. 7(d) is convex and therefore without an asymptotic limit, suggesting that the distribution of the time series with current sheets does not have an upper bound on the right tail. In this case, the shape parameter is positive ($\xi > 0$). In contrast, the current sheet-free time series of 2016 January 25 (Fig. 8d), as well as the time series obtained from removing the current sheets on 2008 March 9 (Fig. 9d), exhibit concave return-level curves and therefore present an asymptotic limit, suggesting that the distributions of time series without current sheets have an upper bound on the right tail. Their shape parameters are negative and the distribution is bounded in the right tail. Furthermore, in all aforementioned cases, the sample's representative points do not extrapolate their respective confidence intervals, evidencing satisfactory GPD fits.

The return-level values for each 10, 20 and 100 event occurrences and their respective 95 per cent confidence intervals are found in Table 1, where the results of all parameters obtained from the statistical diagnostics of volatility time series in Figs 2–4 are included. We see that the confidence intervals computed with different statistical techniques are in general agreement. For best visualization, all return-level plots are redone using the return-level time series shown in Figs 7(e), 8(e) and 9(e), where both the return periods and thresholds are represented by horizontal lines.

Table 1 lists the GPD fit parameters for the volatility time series of Figs 2, 3 and 4. All parameters were estimated by the maximum-likelihood estimate. Once more, the 95 per cent confidence interval for each parameter is estimated using three techniques: the normal approximation, the profile likelihood and bootstrapping with 500 replicates. It is worth mentioning that, in the choice of our GPD parameters, we used criteria that prioritize goodness of fit over excess

Table 1. Estimates of the GPD parameters (ξ and σ) and return levels with their respective 95 per cent confidence intervals (in parentheses) for volatility time series of Figs 2, 3 and 4 using the normal approximation technique.

Time series	Normal approximation									
	ξ		σ		10 events		20 events		100 events	
	Value	95% CI	Value	95% CI	Return Level	95% CI	Return Level	95% CI	Return Level	95% CI
2008 March 9 (with current sheet)	0.431	(0.083,0.780)	0.863	(0.511,1.214)	1.475	(1.270,1.679)	2.268	(1.806,3.078)	5.344	(3.574,7.114)
2009 February 24 (with current sheet)	0.409	(−0.302,1.120)	1.352	(0.297,2.406)	1.554	(0.712,2.397)	2.426	(1.951,2.902)	5.716	(3.961,7.471)
2009 March 8 (with current sheet)	0.225	(−0.301,0.751)	1.469	(0.583,3.555)	1.575	(1.061,2.088)	2.622	(2.078,3.166)	5.786	(4.270,7.301)
2008 March 9 (removed current sheet)	−0.230	(−0.551,0.092)	4.683	(2.682,6.683)	10.343	(8.599,12.086)	12.100	(10.717,15.281)	17.755	(14.379,21.130)
2009 February 24 (removed current sheet)	−0.234	(−0.483,0.014)	1.933	(1.292,2.575)	6.525	(5.868,7.181)	7.534	(6.727,8.340)	9.331	(8.125,10.536)
2009 March 8 (removed current sheet)	−0.225	(−0.373,−0.076)	2.084	(1.636,2.531)	6.935	(6.311,7.559)	7.849	(7.116,8.581)	9.495	(8.371,10.618)
2003 February 1 (no current sheet)	−0.115	(−0.290,0.059)	0.673	(0.509,0.837)	2.917	(2.713,3.120)	3.295	(3.050,3.540)	4.065	(3.635,4.494)
2004 April 21 (no current sheet)	−0.360	(−0.540,−0.180)	0.511	(0.371,0.651)	1.804	(1.644,1.963)	2.051	(1.839,2.261)	2.433	(2.080,2.785)
2016 January 25 (no current sheet)	−0.284	(−0.406,−0.161)	0.861	(0.710,1.013)	3.754	(3.523,3.983)	4.072	(3.798,4.346)	4.609	(4.177,5.040)

Table 2. The same results as Table 1, but using the profile-likelihood method.

Time series	Profile-likelihood approximation									
	ξ		σ		10 events		20 events		100 events	
	Value	95% CI	Value	95% CI	Return Level	95% CI	Return Level	95% CI	Return Level	95% CI
2008 March 9 (with current sheet)	0.432	(0.147,0.858)	0.863	(0.576,1.282)	1.475	(1.476,1.587)	2.268	(2.106,2.676)	5.344	(4.097,8.459)
2009 February 24 (with current sheet)	0.409	(−0.102,1.300)	1.352	(0.614,2.802)	1.554	(0.670,1.950)	2.426	(2.465,2.650)	5.716	(4.375,8.000)
2009 March 8 (with current sheet)	0.225	(−0.189,0.910)	1.470	(0.769,2.579)	1.575	(1.301,1.739)	2.622	(2.320,3.064)	5.786	(4.619,8.000)
2008 March 9 (removed current sheet)	−0.230	(−0.355,0.176)	4.683	(3.173,7.081)	10.343	(9.645,11.362)	13.000	(12.194,14.552)	17.755	(16.861,21.735)
2009 February 24 (removed current sheet)	−0.234	(−0.316,0.061)	1.934	(1.431,2.675)	6.525	(6.232,6.928)	7.541	(7.069,8.045)	9.213	(9.259,10.680)
2009 March 8 (removed current sheet)	−0.225	(−0.298,−0.055)	2.084	(1.677,2.572)	6.935	(6.642,7.253)	7.849	(7.541,8.235)	9.495	(9.331,10.702)
2003 February 1 (no current sheet)	−0.115	(−0.230,0.093)	0.673	(0.532,0.852)	2.917	(2.853,2.994)	3.295	(3.131,3.397)	4.065	(3.950,4.676)
2004 April 21 (no current sheet)	−0.360	(−0.387,−0.144)	0.512	(0.467,0.670)	1.804	(1.765,1.878)	2.051	(1.987,2.116)	2.433	(2.447,2.587)
2016 January 25 (no current sheet)	−0.284	(−0.357,−0.143)	0.862	(0.726,1.021)	3.754	(3.822,3.822)	4.072	(4.170,4.170)	4.609	(4.515,4.837)

Table 3. The same results as in Tables 1 and 2, but using the bootstrapping resampling technique.

Time series	ξ		σ		Bootstrapping approximation 10 events		20 events		100 events	
	Value	95% CI	Value	95% CI	Return Level	95% CI	Return Level	95% CI	Return Level	95% CI
2008 March 9 (with current sheet)	0.431	(0.035,0.732)	0.863	(0.554,1.310)	1.486	(1.393,1.597)	2.281	(1.957,2.702)	5.389	(3.809,7.429)
2009 February 24 (with current sheet)	0.410	(−0.224,0.830)	1.352	(0.780,2.627)	1.471	(0.681,1.863)	2.446	(2.325,2.632)	5.720	(4.340,7.842)
2009 March 8 (with current sheet)	0.225	(−0.260,0.577)	1.469	(0.900,2.445)	1.544	(1.308,1.708)	2.661	(2.358,3.078)	5.765	(4.570,7.422)
2008 March 9 (removed current sheet)	−0.223	(−0.663,0.006)	4.683	(3.303,7.591)	10.431	(9.707,11.353)	13.137	(11.736,14.725)	17.474	(15.140,19.766)
2009 February 24 (removed current sheet)	−0.234	(−0.520,−0.045)	1.933	(1.434,2.783)	6.560	(6.232,6.928)	7.541	(7.069,8.045)	9.213	(8.518,10.022)
2009 March 8 (removed current sheet)	−0.225	(−0.402,−0.111)	2.083	(1.684,2.562)	6.936	(6.541,7.355)	7.828	(7.365,8.275)	9.422	(8.742,10.163)
2003 February 1 (no current sheet)	−0.115	(−0.309,0.022)	0.673	(0.545,0.870)	2.921	(2.780,3.066)	3.295	(3.116,3.503)	4.019	(3.693,4.415)
2004 April 21 (no current sheet)	−0.360	(−0.667,−0.210)	0.511	(0.388,0.704)	1.813	(1.743,1.892)	2.057	(1.955,2.163)	2.416	(2.285,2.531)
2016 January 25 (no current sheet)	−0.284	(−0.451,−0.196)	0.861	(0.745,1.059)	3.751	(3.631,3.878)	4.065	(3.925,4.200)	4.583	(4.366,4.768)

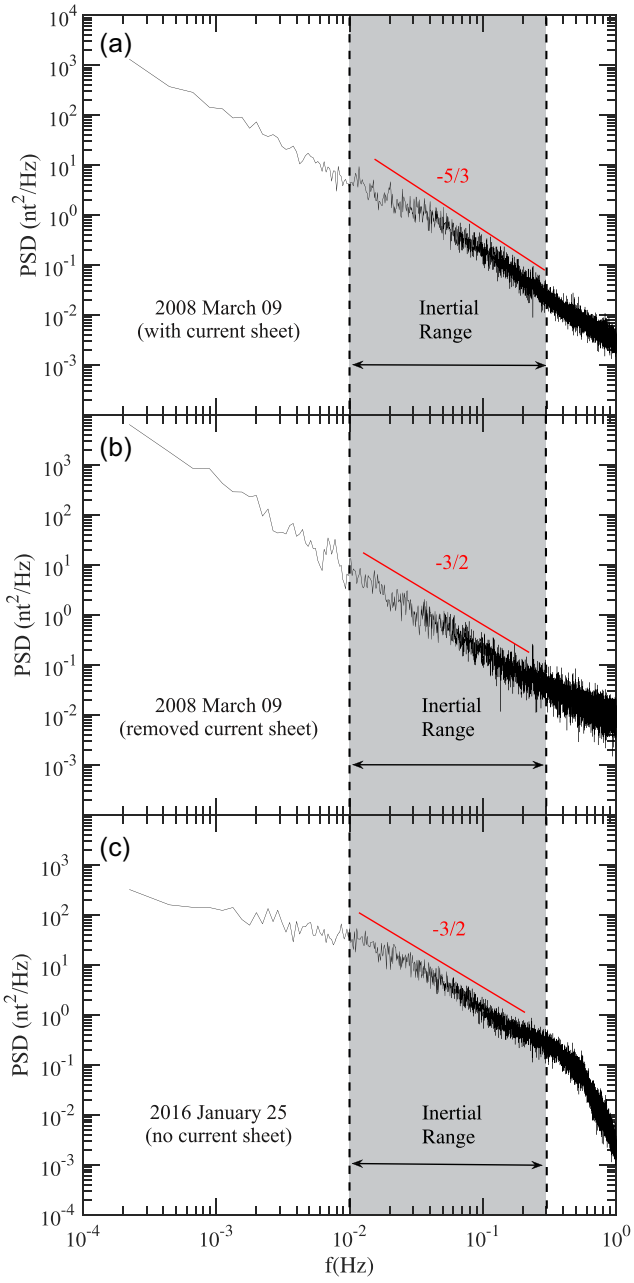

Figure 10. Power-spectrum density (PSD) of the time series of $|\boldsymbol{B}|$ for 2008 March 9, 2008 March 9 after removing current sheets and 2016 January 25. The shaded regions mark the inertial range. Slopes of the power spectra calculated by linear regression inside the inertial ranges are represented by solid lines.

of parameters, such as provided by the negative log-likelihood, Akaike's information criterion (AIC) and Bayesian information criterion (BIC) tests (Tsay 2010). The best models are those with the lowest values of 'negative log-likelihood', AIC and BIC (Tsay 2010). From Table 1, it can be seen that all volatility time series with current sheets (Fig. 2) display a positive shape parameter. Although the confidence intervals have some negative values, the intervals are predominantly positive.

Regarding the time series without current sheets (Fig. 3), the shape parameter assumes negative values, as seen in Table 1, which also shows the shape parameter for the time series with current

Table 4. Slopes of the power spectra calculated by linear regression inside the inertial ranges with their respective coefficients of determination R^2 for all time series shown in Figs 2, 3 and 4. Extremal indexes θ before and after declustering with their respective 95% confidence intervals (in parentheses) are also shown.

Time series	Inertial range		Extremal index after declustering		Extremal index before declustering	
	Slope	R^2	θ	95% CI	θ	95% CI
2008 March 9 (with current sheet)	-1.69 ± 0.01	0.945	0.91	(0.76,1.00)	0.0016	(0.0010,0.0029)
2009 February 24 (with current sheet)	-1.67 ± 0.01	0.949	0.69	(0.50,1.00)	0.0025	(0.0015,0.0047)
2009 March 8 (with current sheet)	-1.73 ± 0.01	0.957	0.83	(0.72,1.00)	0.0019	(0.0011,0.0044)
2008 March 9 (removed current sheet)	-1.49 ± 0.01	0.895	0.69	(0.52,1.00)	0.02	(0.01,0.08)
2009 February 24 (removed current sheet)	-1.50 ± 0.01	0.936	0.87	(0.77,1.00)	0.05	(0.04,0.07)
2009 March 8 (removed current sheet)	-1.52 ± 0.01	0.924	0.82	(0.64,1.00)	0.03	(0.02,0.06)
2003 February 1 (no current sheet)	-1.45 ± 0.02	0.926	0.60	(0.45,1.00)	0.01	(0.008,0.03)
2004 April 21 (no current sheet)	-1.48 ± 0.01	0.917	0.75	(0.62,1.00)	0.03	(0.02,0.07)
2016 January 25 (no current sheet)	-1.55 ± 0.01	0.950	0.99	(0.98,1.00)	0.11	(0.10,0.13)

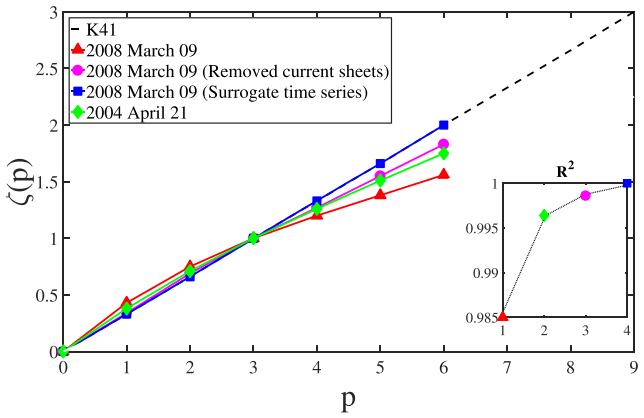

Figure 11. Zeta functions for the time series of magnitude of the magnetic field for 2008 March 9 (red triangles), 2008 March 9 after removing the current sheets (magenta/gray circles), 2008 March 9 after surrogating (blue square) and 2004 April 21 (green diamonds). Coefficients of determination R^2 for the linear regressions calculated from the zeta curves are also shown. The dashed line corresponds to K41 (self-similar) Kolmogorov scaling.

sheets removed (Fig. 4). Again, the shape parameter is negative, proving that the current sheets impact the extreme-event statistics in the solar wind directly. While time series without current sheets have bounded distributions, the presence of these coherent structures in the time series causes the distribution to become unbounded and therefore highly unpredictable.

The 95 per cent confidence intervals of all return-level plots and parameters were obtained by supposing that the distributions of the estimates are normal, which cannot be guaranteed. For that reason, we also computed the 95 per cent confidence intervals with two alternative techniques that do not assume a normal distribution of estimates: the profile-likelihood method (Royston 2007) and the bootstrapping resampling technique with 500 replicates (Efron 1979). Tables 2 and 3 show the same results as Table 1, but using the profile-likelihood and bootstrapping techniques, respectively.

4.3 Turbulence and multifractality

The power-spectrum density (PSD) plots of the magnitude of the magnetic field for 2008 March 9 (with current sheets), 2008 March 9 (removed current sheets) and 2016 January 25 (without current sheets) are shown in Fig. 10. As can be seen in Fig. 10(a), the PSD for a time series with current sheets exhibits an inertial range that follows a K41 scaling law (solid line with slope −5/3: Kolmogorov 1941), whereas the PSD for time series without current sheets (Figs 10b and c) exhibits an IK scaling (solid line with slope −3/2: Iroshnikov 1964; Kraichnan 1965). The slopes of the inertial ranges of the power spectra for all time series shown in Figs 2, 3 and 4 are listed in Table 4, which shows that all time series with current sheets present an excellent fit of the K41 scaling and all time series without current sheets or with current sheets removed display an IK law, corroborating the results of Li et al. (2011). Table 4 also shows that the linear regression coefficients of the inertial ranges are very close to one, confirming the linear behaviour of those regions.

Fig. 11 shows the zeta function as a function of the statistical moments $\zeta(p)$ for the same three time series as Fig. 10. It can be seen that $\zeta(p)$ is less linear for series with current sheets, as confirmed by the values of the coefficient of determination R^2, also plotted in Fig. 11. The coefficient of determination measures the goodness of fit and ranges from 0 to 1 (Wright 1921), where the value 1.0 indicates a perfect fit, whereas 0 indicates a very poor fit. In the case of a linear association between the variables, R^2 measures the accuracy of linear regression. We suggest that current sheets are responsible for enhancing multifractality in the time series, since such structures are known to be responsible for an increase in intermittency in the solar wind, thus enhancing the heterogeneity between time-scales.

A deeper analysis of multifractality in the presence of current sheets can be accomplished with the surrogate iterative amplitude-adjusted Fourier transform (IAAFT: Kantz & Schreiber 1997). The IAAFT consists of the reshuffle of a series by randomizing the phases of its Fourier modes. This type of surrogate breaks all non-linearities while preserving the PSD and, consequently, the autocorrelation function of the original series. The new series has an approximately Gaussian PDF, indicating that the coherent structures responsible for the intermittency are destroyed by the IAAFT. Fig. 11 shows that the post-surrogate series presents an almost perfectly linear ζ function, following a K41 law (i.e. monofractal). This is evidence that the main source of multifractality in this kind of series is the fat tails of the PDF, which is affected directly by the current sheets.

5 CONCLUSION

Several works have sought for methods to detect discontinuities in the solar wind magnetic field caused by current sheets. For instance, in Chian & Miranda (2009) a time series of the normalized magnetic field differences was used, while in Greco et al. (2009, 2018) the

partial variance of increments method was adopted. In this work, we successfully applied, for the first time, the magnetic volatility as a method to detect those discontinuities and to study extreme-value statistics in the solar wind. The method is easily implemented and resulted in time series with high peaks in current-sheet regions, which suggests that the magnetic volatility could be explored as an alternative way to detect those structures in future works.

Our results provide observational evidence for the direct relation between current sheets and multifractality and fully developed turbulence in the solar wind. Solar wind magnetic field series with current sheets have PSDs where the inertial range follows a −5/3 K41 scaling law, whereas series without current sheets display a −3/2 IK law. Thus, we confirm that current sheets are responsible for fully developed Kolmogorov turbulence in the solar wind, as previously proposed by Li et al. (2011). We conclude that the presence of current sheets significantly enhances the multifractality of the time series and those structures are the main factor responsible for the increase in autocorrelation (memory) and departure from Gaussianity in solar wind magnetic field data.

In the last years, the relation between current sheets and extreme events in magnetic field time series in the solar wind has been explored through intermittency or departure from Gaussianity (Chian & Munoz 2011; Wu et al. 2013; Greco et al. 2009, 2018; Servidio et al. 2011). Based on rigorous statistical analysis, the present work proposes a quantitative way to show this relation. By using extreme-value theory, we demonstrated that the magnetic volatility series is well fitted by a generalized Pareto distribution (GPD). We also showed that current sheets are the main structures responsible for the sign of the shape parameter ξ, the most important parameter in this type of modelling. In the presence of such structures, $\xi > 0$, revealing that the volatility time series are modelled by a GPD with infinite support. In the absence of current sheets, $\xi < 0$, resulting in a GPD with finite support. Thus, in series with current sheets, an upper bound for the magnetic volatility is not to be expected, implying that wide rotations of the magnetic field vector or a large shear angle are possible and, consequently, magnetic reconnection events are more likely to occur (Gosling et al. 2007).

In summary, we were able to model solar wind data using a generalized Pareto distribution. Current sheets were identified as the main factor responsible for the behaviour of the GPD shape parameter, which characterizes the tail of the distribution. In addition, we proposed the use of the magnetic volatility, a simple and robust way to analyse extreme events, intermittent turbulence and multifractality in solar wind data. We hope that our work will help future works in predicting the occurrence of extreme current sheets associated with magnetic reconnection.

ACKNOWLEDGEMENTS

The authors thank Laura Maria Gomes de Figueiredo for valuable participation in the edition of the figures present in this article. We thank the Cluster FGM instrument teams for open access to their data. The statistical analysis made use of the EXTREMES package of the free software R. This research is supported by Brazilian agencies CAPES, FAPESP and CNPq.

REFERENCES

Acero F. J., Carrasco V. M. S., Gallego M. C., García J. A., Vaquero J. M., 2017, ApJ, 839, 98
Acero F. J., Gallego M. C., García J. A., Usoskin I. G., Vaquero J. M., 2018, ApJ, 853, 80
Alfvén H., 1942, Nature, 150, 405
Balkema A. A., de Haan L., 1974, Ann. Probability, 2, 792
Boozer A. H., 2014, Phys. Plasmas, 21, 072907
Bruno R., Carbone V., 2013, Living Reviews in Solar Physics, 10, 2
Burlaga L. F., 1991, J. Geophys. Res., 96, 5847
Chian A. C.-L., Miranda R. A., 2009, Ann. Geophys., 27, 1789
Chian A. C.-L., Munoz P. R., 2011, ApJ, 733, L34
Chian A. C.-L., Feng H. Q., Hu Q., Loew M. H., Miranda R. A., Muñoz P. R., Sibeck D. G., Wu D. J., 2016, ApJ, 832, 179
Coles S., 2001, An Introduction to Statistical Modeling of Extreme Values, Springer Series in Statistics. Springer-Verlag, London
Davison A. C., Smith R. L., 1990, J. R. Statistical Soc., Ser. B (Methodological), 52, 393
Efron B., 1979, Ann. Statistics, 7, 1
Engeland K., Hisdal H., Frigessi A., 2004, Extremes, 7, 5
Engle R., 1982, Econometrica, 50, 987
Fawcett L., Walshaw D., 2006, Environmetrics, 17, 795
Fisher R. A., Tippett L. H. C., 1928, Math. Proc. Cambridge Philos. Soc., 24, 180
Frisch U., 1995, Turbulence: The Legacy of AN Kolmogorov. Cambridge Univ. Press, Cambridge
Frisch U., Sulem P.-L., Nelkin M., 1978, J. Fluid Mech., 87, 719
Goldreich P., Sridhar S., 1997, ApJ, 485, 680
Gonzalez W. D., Joselyn J. A., Kamide Y., Kroehl H. W., Rostoker G., Tsurutani B. T., Vasyliunas V. M., 1994, J. Geophys. Res.: Space Phys., 99, 5771
Gosling J. T., 2012, Space Sci. Rev., 172, 187
Gosling J. T., Phan T. D., Lin R. P., Szabo A., 2007, Geophys. Res. Lett., 34, L15110
Greco A., Matthaeus W. H., Servidio S., Chuychai P., Dmitruk P., 2009, ApJ, 691, L111
Greco A., Matthaeus W. H. P. S., Osman K. T., Servidio S., Wan M., Dmitruk P., 2018, Space Sci. Rev., 214, 1
Green E., Hanan W., Heffernan D., 2014, European Phys. J. B, 87, 129
Hanslmeier A., 2007, The Sun and Space Weather, Astrophysics and Space Science Library. Springer, Netherlands
Hathaway D. H., 2015, Living Rev. Solar Phys., 12, 4
Hołyst J., Żebrowska M., Urbanowicz K., 2001, European Phys. J. B – Condensed Matter and Complex Systems, 20, 531
Hosking J. R. M., Wallis J. F., 1987, Technometrics, 29, 339
Iroshnikov P., 1964, SvA, 7, 566
Jenkinson A. F., 1995, Q. J. R. Meteorol. Soc., 81, 158
Kantz H., Schreiber T., 1997, Nonlinear Time Series Analysis. Cambridge Univ. Press, New York, NY, USA
Kolmogorov A., 1941, Akademiia Nauk SSSR Doklady, 30, 301
Kolmogorov A. N., 1962, J. Fluid Mech., 13, 82
Kraichnan R. H., 1965, Phys. Fluids, 8, 1385
Leadbetter M. R., 1983, Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete, 65, 291
Li G., 2008, ApJ, 672, L65
Li G., Miao B., Hu Q., Qin G., 2011, Phys. Rev. Lett., 106, 125001
Ljung G. M., Box G. E. P., 1978, Biometrika, 65, 297
Luceño A., 2006, Comput. Stat. Data Anal., 51, 904
Matthaeus W. H., Ghosh S., Oughton S., Roberts D. A., 1996, J. Geophys. Res.: Space Phys., 101, 7619
Moloney N. R., Davidsen J., 2010, J. Geophys. Res.: Space Phys., 115, A10114
Montgomery D., Matthaeus W. H., 1995, ApJ, 447, 706
Ogata Y., 1988, J. Amer. Stat. Assoc., 83, 9
Oppenheim A., Verghese G., 2017, Signals, Systems and Inference, Global Edition. Pearson, Harlow
Osman K. T., Matthaeus W. H., Gosling J. T., Greco A., Servidio S., Hnat B., Chapman S. C., Phan T. D., 2014, Phys. Rev. Lett., 112, 215002
Parker E., 1979, Cosmical Magnetic Fields: Their Origin and their Activity. Oxford Univ. Press, Oxford, p. 858
Park M. H., Kim J. H., 2016, Comput. Stat. Data Anal., 98, 91

Paschmann G., Daly P., 1998, Analysis Methods for Multi-spacecraft Data, ISSI Scientific Report. International Space Science Institute, Noordwijk
Pickands J., 1975, Ann. Statistics, 3, 119
Politano H., Pouquet A., 1995, Phys. Rev. E, 52, 636
Poon S., 2005, A Practical Guide to Forecasting Financial Market Volatility. The Wiley Finance Series. Wiley, Chichester
Riley P., 2012, Space Weather, 10, S02012
Royston P., 2007, Stata J., 7, 376
Salem C., Mangeney A., Bale S. D., Veltri P., 2009, ApJ, 702, 537
Servidio S., Greco A., Matthaeus W. H., Osman K. T., Dmitruk P., 2011, J. Geophys. Res.: Space Phys., 116, A09102
She Z.-S., Leveque E., 1994, Phys. Rev. Lett., 72, 336
Sridhar S., Goldreich P., 1994, ApJ, 432, 612
Thomas M., Lemaitre M., Wilson M. L., Viboud C., Yordanov Y., Wackernagel H., Carrat F., 2016, PLOS ONE, 11, 1
Tsay R., 2010, Analysis of Financial Time Series, Wiley Series in Probability and Statistics. Wiley, Hoboken
Tsiftsi T., De la Luz V., 2018, Space Weather, 16, 1984
Tsubouchi K., Omura Y., 2007, Space Weather, 5, 1
Voit J., 2005, The Statistical Mechanics of Financial Markets, 3rd edn, Wiley Series in Probability and Statistics. Springer-Verlag, Berlin, Heidelberg
Welch P., 1967, IEEE Trans. Audio Electroacoust., 15, 70
Wright S., 1921, J. Agricultural Res., 20, 557
Wu P. et al., 2013, ApJ, 763, L30
Zhang J., 2007, Australian & New Zealand Journal of Statistics, 49, 69
Zhang J., 2010, Technometrics, 52, 335

This paper has been typeset from a TEX/LATEX file prepared by the author.