STATISTICAL ANALYSIS OF THE MAXIMUM ENERGY IN SOLAR X-RAY FLARE ACTIVITY*

Krzysztof Burnecki

Hugo Steinhaus Center
and
Institute of Mathematics and Computer Science
Wrocław University of Technology
Wyb. Wyspiańskiego 27, 50–370 Wrocław, Poland
Krzysztof.Burnecki@pwr.wroc.pl

Aleksander Stanislavsky

Institute of Radio Astronomy, National Academy of Sciences of Ukraine
4 Chervonopraporna St., 61002 Kharkov, Ukraine
alexstan@ri.kharkov.ua

Karina Weron

Institute of Physics, Wrocław University of Technology
Wyb. Wyspiańskiego 27, 50–370 Wrocław, Poland
Karina.Weron@pwr.wroc.pl

(Received April 20, 2009)

We analyze the time series of soft X-ray emission registered in the years 1974–2007 by the GOES spacecrafts. We show that in the periods of high solar activity, namely 1977–1981, 1988–1992, 1999–2003 the maximum energy of soft X-ray solar flares exhibits both heavy tails and long-range dependence. We investigate the presence of long-range dependence by means of different self-similarity estimators. This analysis gives a promising start to model the appearance of such solar events during solar maxima.

PACS numbers: 02.50.Fz, 05.40.Fb, 96.60.–j

1. Introduction

Understanding the long-term solar variability is one of the most important problems in solar physics. Solar activity has clearly a periodic character, but solar cycles (taken separately) are different in form, amplitude and

* Presented at the XXI Marian Smoluchowski Symposium on Statistical Physics Zakopane, Poland, September 13–18, 2008.

length. Solar flare events are random in time and strength. Their prediction is an actual problem as these disturbances can pose serious threats to man-made spacecrafts, disrupt electronic communication channels and can even set up huge electrical currents in power grids [1]. It is enough to remind about serious problems with the spacecrafts GOES, Deutsche Telecom, Telstars 401, *etc.* Satellite operators would be glad to escape the unhappy surprise, and mission planners are compelled to take into account the future space weather forecast. The cost to the airline industry arose as planes were re-routed to lower altitudes, burning more fuel in force of atmospheric drag. Now the space weather forecast has the considerable importance as well as the weather forecast on the Earth.

As the geological records show, the Earth's climate has always been changing. The reasons for such changes, however, have always been subject to continuous discussions and are still not well understood. In addition to natural climate changes the risk of human influence on climate is seriously considered too. Any factor that alters the radiation received from the Sun or lost to Space will affect climate. So, Mann, Bradley and Hughes [2] have detected clearly a significant correlation between solar irradiance and reconstructed Northern Hemisphere temperature. The statistics indicates that during "Maunder Minimum" of solar activity the climate was especially cold, but when the intensity of solar radiance again increased from early nineteenth century through to the mid-twentieth century, the period coincides with the general warming. This, however, would either imply unrealistically large variations in total solar irradiance or a higher climate sensitivity to radiative forcing than normally accepted. Therefore other mechanisms have to be invoked. The most promising candidate is a change in cloud formation because clouds have a very strong impact on the radiation balance and because only little energy is needed to change their formation process. According to satellite records from 1979 to 1992, during solar minima, Earth was 3% cloudier than at solar maxima [3]. One of the ways to influence cloud formation might be through the cosmic ray flux that is strongly modulated by the varying solar activity.

The aim of this paper is to present a statistical analysis of soft X-ray (SXR) solar flare activity during the last three cycles (1974–2007 years). The paper is organized as follows. The random features of solar activity is outlined in Section 2. The data set is described in Section 3. In Section 4 we present an essence of our statistical investigation of power statistics of SXR flares. Interrelations among statistical flare parameters, such as tail index α and self-similarity exponent H for X-ray flux and their evolution during solar cycles, are analyzed. In the period of strong maxima the self-similarity exponent is almost constant. This feature can be used for predicting the power of SXR emission in future. Next we give a summary and discussion of the main results. Finally, the conclusions are drawn in Section 5.

2. Randomness in nature of solar activity

The solar 11-year cycle is driven by Sun's magnetic field. The Sun's magnetic field is produced by a hydromagnetic dynamo process underneath the solar surface and is cyclic in nature. This fundamental theoretical idea was established by Parker [4]. However, only within the last few years, theoretical models of the solar dynamo have become sophisticated enough to explain deep-laid aspects of solar activity. So, recently Dikrati and Gilman [5] have made an attempt of using a theoretical dynamo model to predict the strength of the upcoming cycle 24. Their analysis is based on fairly regular processes like the magnetic field advection and toroidal field generation by differential rotation during the rising phase of a cycle from a minimum to a maximum. However Choudhuri, Chatterjee and Jiang [6] exposed shortcomings of this study. The key problem in this dispute lies in the Babcock–Leighton mechanism (dominant in the declining phase) for which involves randomness (primary cause of solar cycle fluctuations) [7]. That is why, although active regions appear in a latitude belt at a certain phase of the solar cycle, where exactly within this belt the active regions appear seems random.

The other feature of solar activity is that there is a "magnetic persistence" between the surface polar fields and spot-producing toroidal fields, generating by differential rotation shearing [8]. This means that the Sun retains a memory of its magnetic field for a long time (about 20 years or so). The solar cycle prediction is similar to that employed in global atmospheric dynamics over the last decades. Such models predict changes in certain global characteristics of a cycle, without attempting to reproduce details that occur on smaller spatial scales and shorter time scales. The interrelation between global characteristics and small scale processes is an open problem, and meanwhile some effects of smaller scales are included in a parametric form.

Thus, randomness plays an important role in solar activity. Probably, the success of space weather forecast will rest on statistical methods. In this paper, we are going to analyze the solar flare activity from the SXR time series. We try to answer question, what are the characteristics of the underlying process, without searching any correlations with clearly solar parameters such as properties of spot growth, sunspot structure, magnetic topology, *etc.*

3. X-ray flare observations

Solar activity includes flares, prominence eruptions, coronal mass ejections, solar energetic particles, various radio bursts, high-speed solar wind streaming from coronal holes, *etc.* Solar flares are the most energetic and violent events occurring in the solar atmosphere. The energy release in a flare ranges from 10^{26} to 3×10^{32} ergs.

Observations of solar flare phenomena in X-rays became possible in the 1960s with the availability of space-borne instrumentation. Since 1974 broad-band SXR emission of the Sun has been measured almost continuously by the meteorology satellites operated by National Oceanic and Atmospheric Administration (NOAA) so as the Synchronous Meteorological Satellite (SMS) and the Geostationary Operational Environment Satellite (GOES). The first GOES was launched by NASA in 1975, and the GOES series extends to the currently operational GOES 11 and GOES 12. From 1974 to 1986 the SXR records are obtained by at least one GOES-type satellite; since 1983, data from two and even three co-operating GOES are generally available. The X-ray sensor, part of the space environment monitor system aboard GOES, consists of two ion chamber detectors, which provide whole-sun X-ray fluxes in the 0.05–0.3 and 0.1–0.8 nm wavelength bands.

Solar SXR flares are classified according to their peak burst intensity measured in the 0.1–0.8 nm wavelength band by GOES. The letters (A, B, C, M, X) denote the order of magnitude of the peak flux on a logarithmic scale, and the number following the letter gives the multiplicative factor, *i.e.*, A$n = n \times 10^{-8}$, B$n = n \times 10^{-7}$, C$n = n \times 10^{-6}$, M$n = n \times 10^{-5}$ and X$n = n \times 10^{-4}$ W/m^2. In general, n is given as a float number with one decimal (prior to 1980, n was listed as an integer). Now the data is widely available from the NOAA Space Environment Center site (`http://goes.ngdc.noaa.gov/data/avg/`).

In the meantime a wealth of data has been accumulated, which makes worthwhile re-investigating the temporal and spatial features of SXR flares on a more extensive statistical basis. In particular, the distribution of the X-ray flares (M $\geq$ 1) from 1987 to 1992 with respect to heliografic longitude was studied in [9], where it is shown that the flares were not uniformly distributed in longitude. The temporal analysis of X-flare statistics presented in [10–14] concerns basically the waiting-time distribution. In contrast, our studies will be devoted to the energy statistics of SXR flares in time.

The preliminary analysis of SXR data shows that they are heavy-tailed, namely the right tail of the empirical distribution function $F_n(x)$ can be approximated for large x by the power function $x^{-\alpha}$, where $0 < \alpha < 2$. The estimation of the power-law tail index in such data has a long history in statistics because of its practical importance [15–18]. However, the availability of huge amount of various data poses a set of new challenges for estimating the tail index. Many data, including the time series of SXR solar emission [19], are contaminated by different oscillations and noises complicating their analysis [20–22]. Moreover, the presence of dependence can also distort the results of different estimators of the Pareto-type tail index [23]. Therefore, the tail index should be tested as accurately as possible.

4. Solar flare statistics

We concentrate our attention upon the temporal intervals when the solar activity was strong (1977–1981, 1988–1992 and 1999–2003), see Fig. 1. Their analysis is interesting because in such periods the strongest (and dangerous) flares occur very often. Using the X-ray flare data from GOES satellite, that contain information about time of appearance and energy of solar flares (from `ftp://ftp.ngdc.noaa.gov/STP/SOLAR_DATA/SOLAR FLARES/XRAY_FLARES`), as an example, we consider the SXR time series from the 1st of January 2000 to the 31st of December 2003. The time series is presented in Fig. 2. We study the maximum energy emitted daily. This is in contrast to [24], where energy values aggregated on a daily basis were studied.

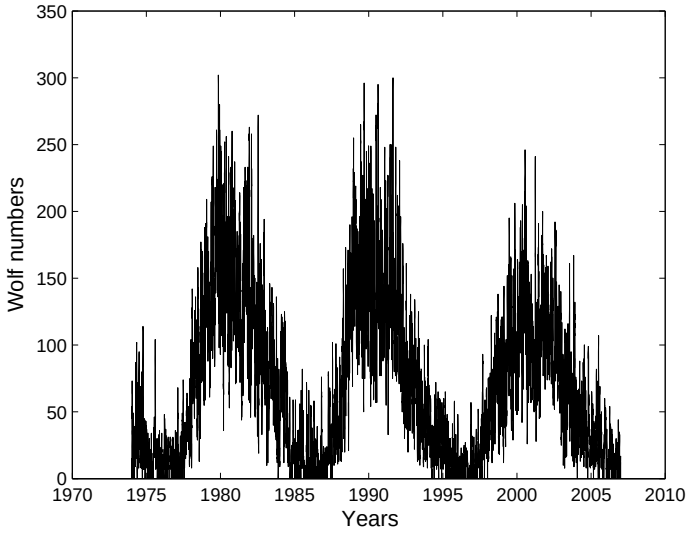

Fig. 1. Wolf numbers (characterizing solar activity) during the last solar cycles 1974–2006.

The analysis of the data shows that the tails of the underlying energy distribution obey the power-law. This implies that the distribution belongs to the domain of attraction of Lévy stable law with α equal to the tail index. We applied Hill [25] and Meerschaert–Scheffler estimators [26] to find the value of the tail index. The classic Hill estimator (HE) is known to be asymptotically normal and consistent. Moreover, it is scale invariant. However, it is not robust to shifting and contains a cutoff parameter that is usually determined in a qualitative fashion. The Meerschaert–Scheffler estimator (MSE) is consistent, shift and cutoff independent, and performs in many situations better than the Hill estimator, but it is not invariant to changes of the scale. In order to avoid problems with shift and scale dependence we center the data by median, and then rescale by the median

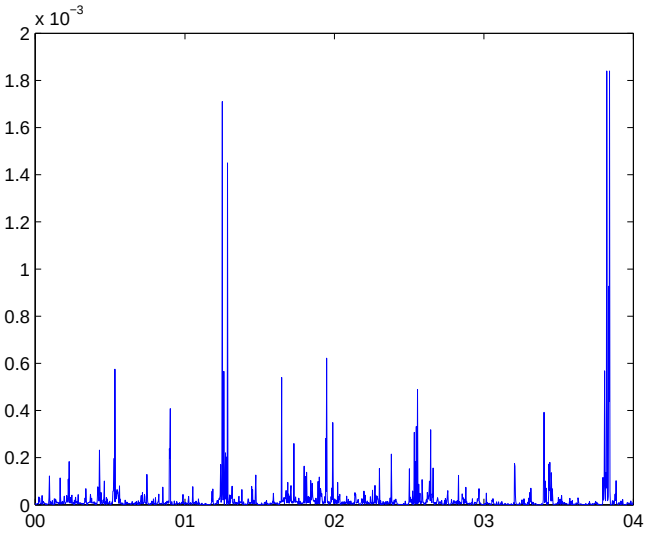

Fig. 2. Maximum energy-time series of solar flares during from 2000–01–01 to 2003–12–31.

of the absolute deviations from that center. This gives a consistent shift and scale invariant estimator of the tail index [26]. The value of $\alpha_{\rm MSE}$ equals 1.0938 for the Meerschaert–Scheffler estimator, and reaches $\alpha_{\rm HE} = 1.1$ for the Hill estimator. We can see that the values are pretty much alike.

We also note that analysis presented in [24] extended the tail index study on some cycles and spoke surely that the index tendency is kept at least during three solar cycles. It also showed a clear correlation between solar activity and both the self-similarity exponent and the tail index.

The self-similarity parameter H can be estimated by the finite impulse response transformation (FIRT). The FIRT estimator involves an array of coefficients made out of finite impulse response coefficients $d_{\rm FIRT}(n,k)$. The estimator $H_{\rm FIRT}$ is obtained from a log-linear regression on the coefficients and measuring the slope [27]. It is important that the estimator $H_{\rm FIRT}$ is unbiased for all $0 < \alpha < 2$, and for $1 < \alpha < 2$, under certain technical conditions, the estimator is consistent and asymptotically normal. The method applied for the original solar flare data is depicted in Fig. 3. We can see that the values of the statistic form a line indicating presence of self-similarity. We included a line indicating the best fit to the data, whose slope provides the estimate of H. The FIRT method is considered to be one of the most effective for detection of the self-similarity exponent for the data that statistically resemble fractional Lévy stable or Brownian motions and fractional autoregressive integrated moving average (FARIMA) models with finite and infinite variance innovations [27].

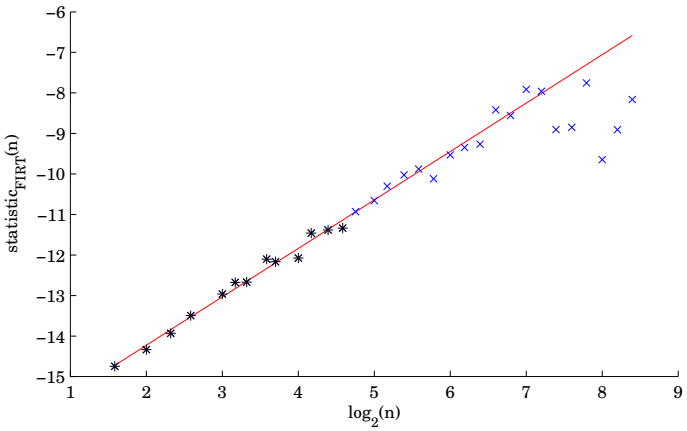

Fig. 3. Finite impulse response transformation (FIRT) for the solar flare data. The estimator H_{FIRT} is obtained by fitting a least-squares line to the values of the FIRT coefficients on a log-linear scale. The points at the high end (marked by 'x') are not used to fit the line, see [27]. The estimated slope of the line is $H_{\mathrm{FIRT}} = 1.1951$.

To estimate the self-similarity parameter H of a time series, we also use the method based on the variance of residuals (VR) [28]. In this approach the series is divided into blocks of size m, and within each block the partial sums of the series are calculated. A least-squares line is fitted to the partial sums within each block, and the sample variance of the residuals is computed. The variance of residuals is proportional to m^{2H}. This variance of residuals is computed for each block, and the median is computed over the blocks. A log–log plot *versus* m should follow a straight line with a slope of $2H$. The method applied for the original solar flare data is depicted in Fig. 4. We can see that the values of the statistic form a straight line, which indicates the presence of self-similarity. We included a line indicating the best fit to the data, whose slope provides the estimate of $2H$.

The self-similarity exponent H is related to the notion of long-memory (or long-range dependence) [29]. The typical way of defining long memory in the time domain is in terms of the rate of decay of the covariance function. If the autocovariance function $r(n)$ tends to 0 so slowly that the sum $\sum_{n=0}^{\infty} |r(n)|$ diverges, we say that in this case the process exhibits long-range dependence. This is true, *e.g.*, for the increment process of any H-self similar process with stationary increments with finite second moment and $H > 1/2$. The major difficulty occurs, when one tries to employ this definition to the processes with infinite second moment, *e.g.* heavy-tailed ones. Since in this case for $\alpha < 2$ the covariance is not defined, one needs to find an alternate measure of dependence, *e.g.* the codifference $\tau(n)$ which equals the covariance when $\alpha = 2$ [20] and is well defined for Lévy stable processes. We say

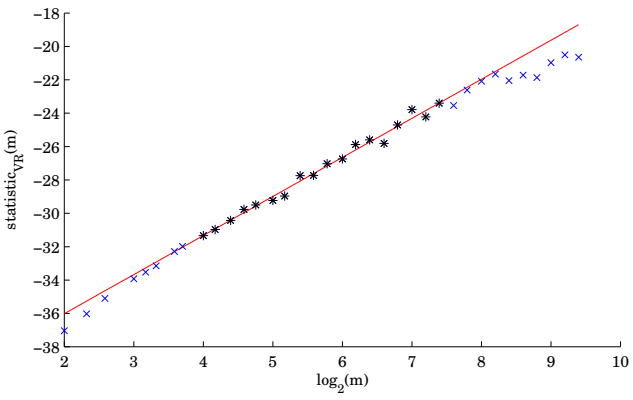

Fig. 4. Variance of residuals (VR) method for the solar flare data. The estimator $H_{\rm VR}$ is obtained by fitting a least-squares line to the median of the sample variance of residuals within blocks of size m on a log–log scale. The points at the very low and high ends (marked by 'x') are not used to fit the line, see [28]. The estimated slope of the line is $2H_{\rm VR} = 2.3402$.

that a stationary Lévy stable process has long memory if its codifference satisfies $\sum_{n=0}^{\infty} |\tau(n)| = \infty$. By analogy with the case $\alpha = 2$, we say that the process has long-range dependence when $H > 1/\alpha$. The classical example of a Lévy stable process with long memory is the fractional autoregressive integrated moving average (FARIMA) time series [30].

For the finite variance cases, the interpretation of the FIRT and VR estimators is very similar to the Hurst exponent. If only short-range correlations (or no correlations at all) exist in the studied series, then $H_{\rm FIRT} = H_{\rm VR} = 1/2$. If there is a correlation then $H_{\rm FIRT} = H_{\rm VR} \neq 1/2$. Moreover, if the estimator $H_{\rm FIRT} = H_{\rm VR}$ is greater than $1/2$, the time series is persistent and if $H_{\rm FIRT} = H_{\rm VR} < 1/2$ then the time series is not persistent. However, the estimators give an information on memory and not on distribution of the process.

To recover both the self-similarity exponent H and the tail index α applying an arbitrary H estimator we can use the BMW^2 computer test [29], based on analysis of both the original and surrogate data. The surrogate data are obtained here by a random shuffling of the original data positions, which breaks both short- and long-range dependence. If the original data set is Gaussian [29], then the values of the estimator should change to $1/2$ for the surrogate data independently on the initial values. If, however, the data belongs to the domain of attraction of α-stable distribution for $\alpha < 2$ [29], then the values of the estimator should change to $1/\alpha$ for the surrogate data independently on the initial values. Let us notice, that the test provides also a double-check of the values of the tail index obtained earlier with the help of Hill and Meerschaert–Scheffler estimators.

The FIRT and VR estimators for the shuffled data are depicted in Figs. 5 and 6, respectively. Therefore, the corresponding estimates for the parameter $1/\alpha$ are: $1/\alpha_{\mathrm{FIRT}} = H_{\mathrm{FIRT}} = 0.9736$ and $1/\alpha_{\mathrm{VR}} = H_{\mathrm{VR}} = 0.9395$.

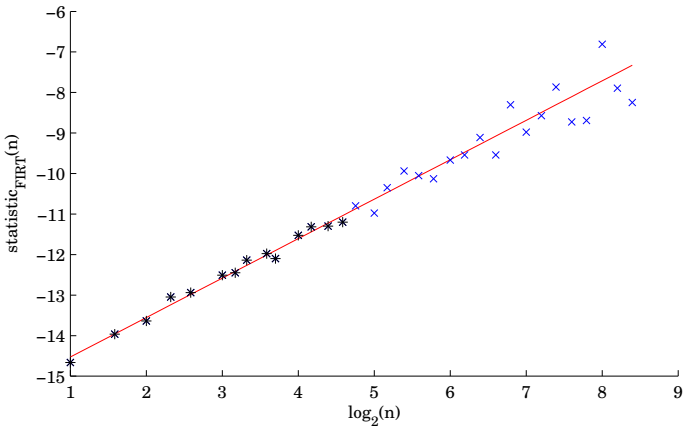

Fig. 5. Finite impulse response transformation (FIRT) for the shuffled solar flare data. The estimator H_{FIRT} is obtained by fitting a least-squares line to the values of the FIRT coefficients on a log-linear scale. The points at the high end (marked by 'x') are not used to fit the line, see [27]. The estimated slope of the line is $H_{\mathrm{FIRT}} = 0.9736$.

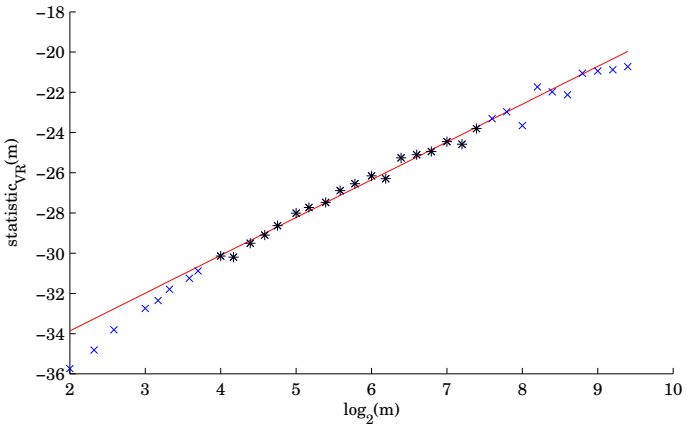

Fig. 6. Variance of residuals (VR) method for the shuffled solar flare data. The estimator H_{VR} is obtained by fitting a least-squares line to the median of the sample variance of residuals within blocks of size m on a log–log scale. The points at the very low and high ends (marked by 'x') are not used to fit the line, see [28]. The estimated slope of the line is $2H_{\mathrm{VR}} = 1.8790$.

We observe that the estimators are not far to the ones estimated by means of the Hill and Meerschaert–Scheffler estimators: $1/\alpha_{\rm HE} = 0.9091$, $1/\alpha_{\rm MSE} = 0.9142$.

5. Conclusions

In this paper we have shown that the maximum energy of solar SXR time series exhibits both heavy-tailed and long-range dependence effects. The first creates a random number of strong bursts on a background, and the second forms their persistence between each other. Comparing the values of the different estimators for the original data series and for the shuffled data we have estimated the self-similarity exponent H and the tail index α. For the moment, we cannot answer the question why H and α have such values, but we hope that launching new space-based systems and installing big astronomical instruments on the far side of the Moon will be very useful for the X-ray study of flaring stars. Probably, this will allow one to get new X-ray data (long-term time series) for these stars to compare their H and α with the solar parameters found for the Sun.

Our findings help one to construct a statistical model for the solar SXR activity. FARIMA model driven by Pareto noise for the analysis of solar SXR data was analyzed in [24]. The FARIMA approach seems plausible because it offers a lot of flexibility in modeling heavy tails, long- and short-range dependences. This is especially important in the case, when it is difficult to propose any clear physical model.

A.S. is grateful to the Institute of Physics and the Hugo Steinhaus Center for pleasant hospitality during his visit in Wrocław University of Technology. The work of A.S. was partly realized in the framework of the project INTAS-03-5727. The GOES X-ray light curve was made available courtesy of the NOAA Space Environment Center, Boulder, CO.

REFERENCES

[1] S. Clark, *Nature* **441**, 402 (2006).
[2] M.E. Mann, R.S. Bradley, M.K. Hughes, *Nature* **392**, 779 (1998).
[3] H. Svensmark, E. Friis-Christensen, *Sol. Terr. Phys.* **59**, 1225 (1997).
[4] E.N. Parker, *Astrophys. J.* **122**, 293 (1955).
[5] M. Dikpati, P.A. Gilman, *Astrophys. J.* **649**, 498 (2006).
[6] A.R. Choudhuri, P. Chatterjee, J. Jiang, *Phys. Rev. Lett.* **98**, 131103 (2007).
[7] D. Nandy, A.R. Choudhuri, *Science* **296**, 1671 (2002).
[8] M. Dikpati, G. de Toma, P.A. Gilman, *Geophys. Res. Lett.* **33**, L05102 (2006).

[9] K.-J. Li, B. Schmieder, Q.-Sh. Li, *Astron. Astrophys. Suppl. Ser.* **131**, 99 (1998).

[10] G. Boffetta, V. Carbone, P. Giuliani, P. Veltri, A. Vulpiani, *Phys. Rev. Lett.* **83**, 4662 (1999).

[11] F. Lepreti, C. Carbone, P. Veltri, *Astrophys. J.* **555**, L133 (2001).

[12] Y.-J. Moon, G.S. Choe, H.S. Yun, Y.D. Park, *J. Geophys. Res.* **106** (A12), 29951 (2001).

[13] M.S. Wheatland, *Sol. Phys.* **208**, 33 (2002).

[14] A. Veronig, M. Temmer, A. Hanslmeier, W. Otruba, M. Messerotti, *Astron. Astrophys.* **382**, 1070 (2002).

[15] G. Zipf, *Selective Studies and Principle of Relative Frequency in Language*, Harvard University Press, 1932.

[16] B.B. Mandelbrot, *Int. Econ. Rev.* **1**, 79 (1960).

[17] R. Adler, R. Feldman, M.S. Taqqu, *A Practical Guide to Heavy Tails: Statistical Techniques and Application*, Birkhäuser, Boston 1998.

[18] K. Park, W. Willinger, *Self-Similar Network Traffic and Performance Evaluation*, J. Wiley& Sons Inc., New York 2000.

[19] M. Baiesi, M. Paczuski, A.L. Stella, *Phys. Rev. Lett.* **96**, 051103(2006).

[20] A. Janicki, A. Weron, *A Simulation and Chaotic Behavior of α-Stable Stochastic Processes*, Dekker, New York 1994.

[21] V.E. Lynch, B.A. Carreras, R. Sanchez, B. LaBombard, B.Ph. van Milligen, *Phys. Plasmas* **12**, 052304 (2005).

[22] V.Yu. Gonchar, A.V. Chechkin, E.L. Sorokovoi, V.V. Chechkin, L.I. Grigor'eva, E.D. Volkov, *Plasma Phys. Rep.* **29**, 380 (2003).

[23] S. Stoev, G. Michailidis, Technical Report **447**, Department of Statistics, University of Michigan, 2006: `http://www.stat.lsa.umich.edu/~sstoev/max-spectrum-dep.pdf`

[24] S. Stanislavsky, K. Burnecki, M. Magdziarz, A. Weron, K. Weron, *Astrophys. J.* **693**, 1877 (2009).

[25] B.M. Hill, *Ann. Stat.* **3**, 1163 (1975).

[26] M.M. Meerschaert, H. Scheffler, *J. Stat. Plan. Infer.* **71**, 19 (1998); K. Bianchi, M.M. Meerschaert, working paper (2004).

[27] S. Stoev, V. Pipiras, M.S. Taqqu, *Signal Processing* **82**, 1873 (2002).

[28] M.S. Taqqu, V. Teverovsky, in *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*, eds. R. Adler, R. Feldman, M.S. Taqqu, Birkhäuser, Boston 1998, p. 177.

[29] A. Weron, K. Burnecki, Sz. Mercik, K. Weron, *Phys. Rev.* **E71**, 016113 (2005).

[30] J. Beran, *Statistics for Long-Memory Processes*, Chapman & Hall, New York 1994.