Open Research Online

The Open University's repository of research publications
and other research outputs

Robust statistical properties of the size of large burst events in AE

Journal Item

For guidance on citations see FAQs.

Version: Version of Record

oro.open.ac.uk

Robust statistical properties of the size of large burst events in *AE*

P. Hush[1], S. C. Chapman[1], M. W. Dunlop[1,2,3], and N. W. Watkins[1,4,5,6]

[1]CFSA, Physics Department, University of Warwick, Warwick, UK, [2]RAL Space, Rutherford Appleton Laboratory, Didcot, UK, [3]Space Science Institute, School of Astronautics, Beihang University, Beijing, China, [4]CATS, London School of Economics and Political Science, London, UK, [5]Open University, Milton Keynes, UK, [6]Max Planck Institute for the Physics of Complex Systems, Dresden, Germany
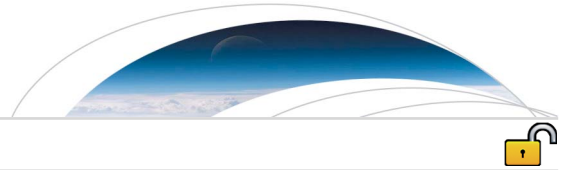
Abstract Geomagnetic indices provide a comprehensive data set with which to quantify space climate, that is, how the statistical likelihood of activity varies with the solar cycle. We characterize space climate by the *AE* index burst distribution. Burst sizes are constructed by thresholding the *AE* time series; a burst is the sum of the excess in the time series for each time interval over which the threshold is exceeded. The distribution of burst sizes is two component with a crossover in behavior at thresholds ≈1000 nT. Above this threshold, we find a range over which the mean burst size varies weakly with threshold for both solar maxima and minima. The burst size distribution of the largest events is exponential. The relative likelihood of these large events varies from one solar maximum and minimum to the next. Given the relative overall activity of a solar maximum/minimum, these results constrain the likelihood of extreme events of a given size.

1. Introduction

Quantifying the large-scale dynamic response of the magnetosphere to solar wind driving is central to our understanding of solar wind-magnetosphere coupling. This dynamic response is complex, with the "state" of the magnetosphere affecting its response to the driver. Auroral geomagnetic indices coupled with in situ solar wind monitors provide a comprehensive data set, spanning several solar cycles. We can characterize these observations in terms of "space climate" by quantifying how the statistical properties of ensembles of these observed variables vary between different phases of the solar cycle.

There has been a long-standing interest in characterizing space weather relevant time series in terms of the statistics of events or bursts, where a single burst is defined as the area under the curve of the segment of the time series that is continuously above a threshold.

Statistics are now robustly established for solar flares, where *Datlowe et al.* [1974], *Lin et al.* [1984], and *Dennis* [1985] were some of the first to perform frequency distributions on solar flare hard X-ray observations, and where their systematic characterization in terms of power laws has been well established [*Crosby et al.*, 1993; *Lee et al.*, 1993; *Crosby et al.*, 1998; *Georgoulis et al.*, 2001; *Aschwanden and Parnell*, 2002]. Solar surface magnetic fields seen in SOHO magnetograms also show power laws [*Parnell et al.*, 2009]. Heavy tailed statistics of fluctuations in solar wind magnetic energy density are seen by monitors at L1. Their distributions show solar cycle dependence [*Kiyani et al.*, 2007; *Hnat et al.*, 2007, 2011].

Geomagnetic indices such as *AE* show long memory and non-Gaussian statistics [*Takalo et al.*, 1993; *Consolini*, 1997; *Takalo and Timonen*, 1999; *Consolini and De Michelis*, 2002]. There have been studies to characterize bursts in the solar wind [*Moloney and Davidsen*, 2011, 2014] and to compare them with ground-based observations [*Wanliss and Weygand*, 2007; *Freeman et al.*, 2000b; *Uritsky et al.*, 2001; 2001; see also *Riley*, 2012].

While all of these studies characterize the statistics of bursts with power law regions, they recognize that there is a departure from this self-affine behavior for the largest events. One approach to modeling such distributions is that of Extreme Value Theory (EVT) which is formally applied to model the likelihood of the observables themselves [*Coles*, 2001] rather than bursts as such. EVT has so far been of limited application in space weather but notable pioneering exceptions include application to the *aa* index [*Siscoe*, 1976; *Silbergleit*, 1999] and the *Dst* index [*Silbergleit*, 1996; *Tsubouchi and Omura*, 2007]. Another application of EVT to energetic electron fluxes was by *Koons* [2001]. The latter paper, further developed by *O'Brien et al.* [2007], used the Generalized
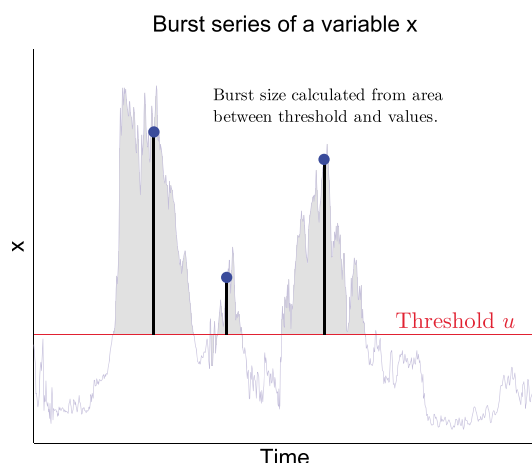
Figure 1. One day of the *AE* time series with an example threshold overplotted. The integrated areas above the threshold are used to create a new time series of "bursts." Each threshold has a corresponding set of bursts with its own mean value.

Extreme Value (GEV) distribution to establish an unexpected upper bound to the flux of relativistic "killer" electrons. Ground effects [*Thomson*, 2007] have only recently begun to be studied using the GEV framework [*Thomson et al.*, 2011; *Beggan et al.*, 2013], see also in the solar wind [*Moloney and Davidsen*, 2010].

The success of these studies suggests that even the largest events may have statistics that are robustly definable from the observations. In this letter we explore the properties of the burst distribution of a well-studied geomagnetic index, *AE*, with emphasis on the largest events, and its variation with the solar cycle. We find that there are clear, robust trends in how the burst size distribution of *AE* depends on the threshold used to define the bursts. At all phases of the solar cycle there is a clear transition to a single, approximately exponential burst size distribution of larger events. These large events follow the same distribution but a greater fraction fall into this category at solar maximum compared to solar minimum. These model-independent results provide constraints on the upper limits of activity above a threshold.

This letter identifies robust aspects of the distribution of bursts in *AE*. The size of a burst event is defined as the integrated area under the curve of an excursion of the time series above a threshold. In geomagnetic indices, the size of a burst characterizes the overall geomagnetic effect of an event. We find that plots of the mean burst size as a function of threshold readily identify the well-known [e.g., *Freeman et al.*, 2000b; *Uritsky et al.*, 2001] multicomponent nature of *AE*. Above a threshold of ≈1000 nT, the mean burst size becomes weakly dependent of threshold. We see the same behavior at both solar maxima and minima; at the maxima, and at more active minima, a larger fraction of burst events are found in this >1000 nT population. We find that these large burst events (hereafter LBEs) have a probability distribution of burst sizes that is independent of threshold and is exponential for LBE size >2000 nT min. This can inform statistical prediction of the largest, space weather relevant events.

2. Constructing Bursts From *AE*

We perform statistical analysis of the full *AE* data set as derived by the WDC [*Davis and Sugiura*, 1966]. The *AE* geomagnetic index is sampled every minute and the full data set is available from January 1978 to June 2014. From this data set we will also draw yearlong samples centered on solar maxima and solar minima, the date of each solar maximum/minimum is determined from the monthly smoothed sunspot number. The dates of the solar maxima are as follows: cycle 21, December 1979; cycle 22, July 1989; and cycle 23, March 2000. The dates for the solar minima are as follows: cycle 22, September 1986; cycle 23, May 1996; and cycle 24, January 2008.

In order to characterize the overall impact of a given event, we will focus on the statistics of burst sizes rather than of the raw data set. Events are defined in the time series as the contiguous group of data samples that exceed a given threshold. Aggregating these samples then suppresses observational uncertainties associated with individual data points while quantifying the overall impact of the event. In particular, *AE* is constructed from the data from the station with maximum response at any given time. Since the stations are spatially sparse and nonuniform, the peak of *AE* can significantly undersample a given event peak.

The method for constructing the burst time series is shown in Figure 1. The size of each burst is the integrated area between the segment of the time series above threshold and the threshold value. Thus, for the *AE* time series (nT) sampled every minute, the natural unit of a burst is nT min. For any given threshold, the original time series gives an ensemble of burst sizes *S* which we can quantify statistically. We have tested the robustness of all the results presented here against varying the (yearlong) time window within which an ensemble of bursts is constructed; we recalculated all quantities having offset the time window by 1 month. In general,
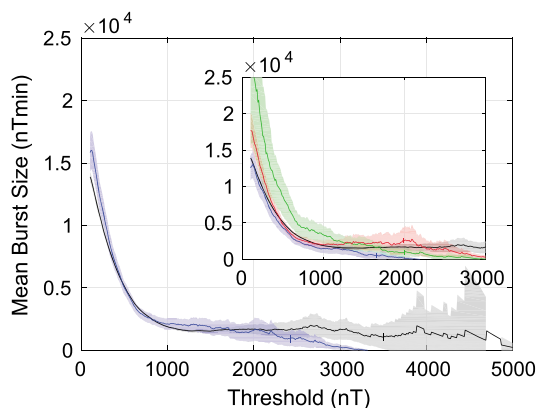
Figure 2. Mean size of burst above threshold is plotted versus threshold for (black) The entire *AE* data set (1978 to mid-2014) and (blue) 1 year intervals for each of the three years that span solar maxima of solar cycles 21–23 aggregated into a single data set. Inset is in the same format and plots the three 1 year intervals separately: cycle 21 (red), cycle 22 (green), and cycle 23 (blue). Two distinct regions of behavior can be seen, with a crossover at around 1000 nT. The shading indicates statistical uncertainty.

the statistics of burst size will depend on the threshold. An isolated burst that has a single, well-defined peak will decrease in size as the threshold increases so that if all bursts were of this type and of roughly the same peak height, the mean size would decrease with threshold. However, there is a variation in peak heights of the bursts so that as the threshold increases successively, smaller bursts with lower peaks tend to zero size first and drop out of the average. Also, the geomagnetic index time series are multipeaked; a single storm event can encompass several large peaks. In this case a single large burst at low threshold can divide into several, smaller bursts as the threshold is increased. At very high thresholds, when there are only a few isolated bursts remaining, one can see a fall off, but here the statistics are not sufficient to meaningfully discuss the mean. We exclude bursts of duration shorter than 5 min from the analysis as these bursts are not sufficiently statistically well resolved.

We plot mean burst size $\bar{S}$ as a function of the threshold T in Figures 2 and 3. Each point on these plots is an estimate of the mean $\bar{S}(T)$, based on a sample size $N(T)$, N decreases as T increases. The main panel plots this for the entire *AE* data set (black line) 1978–2014 and for an ensemble formed from the three solar maximum intervals, that is the three years of data centered in December 1979, July 1989, and March 2000. In the inset we show the same plot except that the three solar maxima are now plotted as separate, yearlong ensembles. The error on the mean is indicated by the error bars on these plots; this is given by $\bar{S} \pm F \times \sigma(S)/\sqrt{N}$ where $\sigma(T)$ is the standard deviation of the $N(T)$ samples associated with the threshold T, and F is the inverse of the t distribution at the 2.5% and 97.5% quantiles. We immediately see a clear crossover at a threshold of $\approx$1000 nT between a roughly linear decrease, and a flattening out of these curves with increasing threshold. Events which contribute to the mean burst size estimate for thresholds >1000 nT are events which at their peak amplitude are >1000 nT, these are the large burst events (LBEs). As we move to higher thresholds the number of bursts used to estimate the mean decreases, and the statistical uncertainty can be seen to increase. Within these uncertainties for the full *AE* data set, the LBEs have mean burst size that is essentially independent of the threshold.
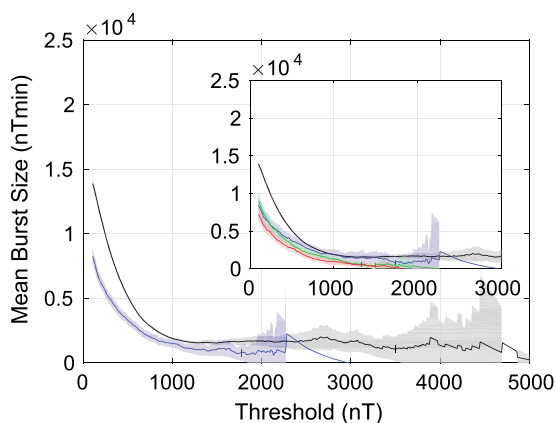

Figure 3. Mean size of burst above threshold is plotted versus threshold in the same format as Figure 2 for the minima of solar cycles 22–24 (inset: red, green, and blue, respectively).

Figure 3 also plots mean burst size as a function of threshold in the same format as Figure 2 but now for the years centered on solar minima (September 1986, May 1996, and January 2008). The same crossover in behavior and flattening out of the curves is seen. The main difference in these curves is the extent of the region >1000 nT, that is, the range of LBE peak amplitudes that occur. As we might expect, there are larger amplitude events occurring at solar maximum compared to solar minimum. Comparing Figures 2 and 3 looking at the ensembles over three successive maxima (main panel), the largest 20 events at solar maximum are at >2420 nT, whereas at solar minimum they are at >1770 nT (for the entire distribution over the full data set, the largest 20 events occur
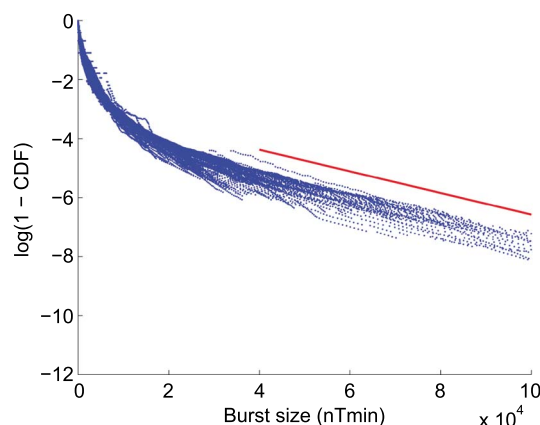
Figure 4. The CDF of burst sizes is plotted for thresholds between 1000 nT and 5000 nT. We plot $\log(1 - \text{CDF})$, to expand the tail of the distributions. For each threshold, the CDF falls on a similar curve. It is roughly exponential at burst sizes greater than 3×10^4 nT min. An exponential function $\exp(-S/S_0)$ for $S_0 = 2.73 \times 10^4$ nT min is plotted (offset, red line) for comparison.

above 3500 nT). However there is considerable variation between one solar maximum/minimum and the next, so that the level of activity can overlap. The largest 20 events at the maximum of cycle 21 are at thresholds > 1650 nT, whereas the largest 20 events of the minimum of cycle 22 are at thresholds > 1740 nT.

In summary, above 1000 nT threshold, the mean burst size of (i) the entire AE data set does not vary with threshold, it is constant within the errors; (ii) solar max for cycles 21–23 taken together (main panel of Figure 2) is constant up to $\approx 2000 - 2500$ nT after which it falls off, there is no data beyond 3000 nT; and (iii) solar minimum for cycles 22–24 taken together (main panel of Figure 3) is constant within the errors. Note that at very high thresholds, when there are only a few isolated bursts remaining, one can see a fall off with increasing threshold, but here the statistics are not sufficient to meaningfully discuss the mean. Smaller ensembles such as those for individual solar maxima and minima (shown inset in Figures 2 and 3) will more clearly show this fall off.

3. Statistical Properties of Bursts

We now consider whether the full probability distribution of burst sizes is independent of the threshold. Figure 4 plots the burst size complementary cumulative distribution function (ccdf); for a variable x, the ccdf(X) is the probability that $x > X$, it is 1-cdf(X). We have calculated the ccdf of the LBE burst size using thresholds between 1000 nT and 5000 nT, the resulting multiple curves are overplotted in Figure 4. We can see that within the uncertainties there is good collapse of these curves onto a single envelope, suggesting that there is a single function for the LBEs that is independent of threshold that captures the burst size probability. Varying the threshold varies the size of a given burst so that different thresholds use the same data to map out different regions of the same curve. The plot is on semilog axes and for the larger events is roughly linear. We overplot a straight line fit and estimate that for burst sizes > 4000 nT min the ccdf(X)$\sim\exp(-X/X_0)$ with $X_0 = 2.73 \times 10^4$ nT, which is a characteristic burst size of an LBE. The burst duration τ varies monotonically with the burst size S, such that $S \sim \tau^\alpha$ where $\alpha \approx 1.5$ for thresholds lower than 3000 nT. Above this, the number of events is sufficiently small that we cannot accurately estimate the exponent α. This relationship between burst size and duration is consistent with previous results (see, e.g., the larger, longer-duration events in *Uritsky et al.* [2001])

Burst size statistics are usually presented as a probability density function (pdf) and this is (minus) the derivative of the ccdf. Since the LBE ccdf tends to an exponential for large events, the pdf will fall off exponentially also. This is the characteristic roll off often seen at large values of burst size pdfs reported to have a power law range of dependence [*Freeman et al.*, 2000a; *Riley*, 2012].

4. Conclusions

The long, well-curated time series of geomagnetic indices provide a comprehensive data set with which to quantify space climate, that is, how the statistical likelihood of activity of a given intensity varies with the

solar cycle. We have characterized space weather events by their burst size as seen in the *AE* index. The statistical distribution of burst sizes is two component with a crossover in behavior at a threshold ≈1000 nT. For events above this threshold, the burst distribution varies only weakly as the threshold is increased. This behavior is robust for both solar maxima and minima. The likelihood of events occurring at these high thresholds (>1000 nT) is larger at solar maximum compared to solar minimum but does vary from one solar cycle to the next. The full distribution of burst sizes of the largest event is exponential.

Importantly, we have shown here that this exponential roll off region has an envelope which has a functional form that is independent of the threshold used to construct the bursts. This functional form could then be used as a basis to characterize the likelihood of these large bursts. As it is an exponential roll off, it suggests that estimates based on power law extrapolations to the largest events will tend to overestimate the burst size and hence the severity of these largest events.

Taking these results together could assist in the quantification of space climate, as it varies within a solar cycle, and across one solar cycle to the next. Above ≈1000 nT, the mean burst size becomes weakly dependent on threshold within uncertainties. The range of thresholds at which the largest events occur varies both between solar maximum and solar minimum, and from one solar maximum (or minimum) to the next. The full distribution of bursts also varies weakly with threshold. This could assist in the prediction of space climate for a given solar maximum or minimum. Once it is determined that solar activity is sufficient for events to be seen above a given threshold, the occurrence likelihood of the burst size of all events below that threshold (and >1000 nT) is known for that solar maximum or minimum.

Finally, the procedure used here is related to the method of mean residual life plots, where one plots the mean excess, that is, the mean of the values of the data points above a threshold. For data that follows generalized extreme value statistics, the mean excess varies linearly with threshold. We constructed mean residual life plots for yearlong samples of *AE* but found that the mean excess of the raw data is not a robust quantity, it fluctuates significantly, by as much as 125% if one varies the time interval of the data by 1 month. However, the mean excess is just the mean burst size/mean burst duration [*Lawrance and Kottegoda*, 1977; Watkins et al., manuscript in preparation] and we have found that both mean burst size and duration statistics are robust against varying the time interval of data. This then offers a robust methodology for estimating the mean excess.

References

Aschwanden, M. J., and C. E. Parnell (2002), Nanoflare statistics from first principles: Fractal geometry and temperature synthesis, *Astron. Astrophys.*, *572*, 1048, doi:10.1086/340385.

Beggan, C. D., D. Beamish, A. Richards, G. S. Kelly, and A. W. P. Thomson (2013), Prediction of extreme geomagnetically induced currents in the UK high-voltage network, *Space Weather*, *11*, 407–419, doi:10.1002/swe.20065.

Coles, S. (2001), *An Introduction to Statistical Modeling of Extreme Values*, 208 pp., Springer, London.

Consolini, G. (1997), Sandpile cellular automata and magnetospheric dynamics, in *8th GIFCO Conference, Cosmic Physics in the Year 2000*, vol. 58, edited by S. Aiello et al., pp. 123–126, Societa Italia di Fisica, Bologna, Italy.

Consolini, G., and P. De Michelis (2002), Fractal time statistics of AE-index burst waiting times: Evidence of metastability, *Nonlinear Processes Geophys.*, *9*, 419–423, doi:10.5194/npg-9-419-2002.

Crosby, N., N. Vilmer, N. Lund, and R. Sunyaev (1998), Deka-keV X-ray observations of solar bursts with WATCH/GRANAT: Frequency distributions of burst parameters, *Astron. Astrophys.*, *334*, 299–313.

Crosby, N. B., M. J. Aschwanden, and B. R. Dennis (1993), Frequency distributions and correlations of solar X-ray flare parameters, *Sol. Phys.*, *143*(2), 275–299, doi:10.1007/BF00646488.

Datlowe, D. W., M. J. Elcan, and H. S. Hudson (1974), OSO-7 observations of solar X-rays in the energy range 10–100 keV, *Sol. Phys.*, *39*(1), 155–174, doi:10.1007/BF00154978.

Davis, T. N., and M. Sugiura (1966), Auroral electrojet activity index *AE* and its universal time variations, *J. Geophys. Res.*, *71*(3), 785–801, doi:10.1029/JZ071i003p00785.

Dennis, B. R. (1985), Solar hard X-ray bursts, *Sol. Phys.*, *100*(1981), 465–490.

Freeman, M., N. Watkins, and D. Riley (2000a), Power law distributions of burst duration and interburst interval in the solar wind: Turbulence or dissipative self-organized criticality?, *Phys. Rev. E.*, *62*(6), 8794–8797.

Freeman, M. P., N. W. Watkins, and D. J. Riley (2000b), Evidence for a solar wind origin of the power law burst lifetime distribution of the *AE* indices, *Geophys. Res. Lett.*, *27*(8), 1087–1090, doi:10.1029/1999GL010742.

Georgoulis, M. K., N. Vilmer, and N. B. Crosby (2001), A comparison between statistical properties of solar X-ray flares and avalanche predictions in cellular automata statistical flare models, *Astron. Astrophys.*, *367*, 326–338, doi:10.1051/0004-6361:20000422.

Hnat, B., S. C. Chapman, K. Kiyani, G. Rowlands, and N. W. Watkins (2007), On the fractal nature of the magnetic field energy density in the solar wind, *Geophys. Res. Lett.*, *34*, L15108, doi:10.1029/2007GL029531.

Hnat, B., S. C. Chapman, G. Gogoberidze, and R. T. Wicks (2011), Scale-free texture of the fast solar wind, *Phys. Rev. E*, *84*(6), 065401, doi:10.1103/PhysRevE.84.065401.

Kiyani, K., S. C. Chapman, B. Hnat, and R. M. Nicol (2007), Self-similar signature of the active solar corona within the inertial range of solar-wind turbulence, *Phys. Rev. Lett.*, *98*(21), 211101, doi:10.1103/PhysRevLett.98.211101.

Koons, H. (2001), Statistical analysis of extreme values in space science, *J. Geophys. Res.*, *106*, 915–922.

Lawrance, A., and N. Kottegoda (1977), Stochastic modelling of riverflow time series, *J. R. Stat. Soc., Ser A*, *140*(1), 1–47.

Lee, T. T., V. Petrosian, and J. M. Mctiernan (1993), The distribution of flare parameters and implications for coronal heating, *Astrophys. J.*, *412*, 401–409, doi:10.1086/172929.

Lin, R. P., R. A. Schwartz, S. R. Kane, R. M. Pelling, and K. C. Hurley (1984), Solar hard X-ray microflares, *Astrophys. J.*, *283*, 421–425, doi:10.1086/162321.

Moloney, N. R., and J. Davidsen (2010), Extreme value statistics in the solar wind: An application to correlated Lévy processes, *J. Geophys. Res.*, *115*, A10114, doi:10.1029/2009JA015114.

Moloney, N. R., and J. Davidsen (2011), Extreme bursts in the solar wind, *Geophys. Res. Lett.*, *38*, L14111, doi:10.1029/2011GL048245.

Moloney, N. R., and J. Davidsen (2014), Stationarity of extreme bursts in the solar wind, *Phys. Rev. E*, *89*(5), 052812, doi:10.1103/PhysRevE.89.052812.

O'Brien, T. P., J. F. Fennell, J. L. Roeder, and G. D. Reeves (2007), Extreme electron fluxes in the outer zone, *Space Weather*, *5*, S01001, doi:10.1029/2006SW000240.

Parnell, C. E., C. E. DeForest, H. J. Hagenaar, B. A. Johnston, D. A. Lamb, and B. T. Welsch (2009), A power law distribution of solar magnetic fields over more than five decades in flux, *Astrophys. J.*, *698*, 75–82, doi:10.1088/0004-637X/698/1/75.

Riley, P. (2012), On the probability of occurrence of extreme space weather events, *Space Weather*, *10*, S02012, doi:10.1029/2011SW000734.

Silbergleit, V. M. (1996), On the occurrence of geomagnetic storms with sudden commencements, *J. Geomagn. Geoelectr.*, *48*, 1011–1016, doi:10.5636/jgg.48.1011.

Silbergleit, V. M. (1999), Forecast of the most geomagnetically disturbed days, *Earth Planets Space*, *51*, 19–22, doi:10.1186/BF03352205.

Siscoe, G. L. (1976), On the statistics of the largest geomagnetic storms per solar cycle, *J. Geophys. Res.*, *81*(25), 4782–4784.

Takalo, J., and J. Timonen (1999), Neural network prediction of the *AE* index from the PC index, *Phys. Chem. Earth, Part C*, *24*(1–3), 89–92.

Takalo, J., J. Timonen, and H. Koskinen (1993), Correlation dimension and affinity of *AE* data and bicolored noise, *Geophys. Res. Lett.*, *20*(15), 1527–1530, doi:10.1029/93GL01596.

Thomson, A. W. (2007), Geomagnetic hazards, in *Encyclopedia of Geomagnetism and Paleomagnetism*, edited by D. Gubbins and E. Herrero-Bervera, pp. 316–319, Springer, Dordrecht, Netherlands.

Thomson, A. W., E. B. Dawson, and S. J. Reay (2011), Quantifying extreme behaviour in geomagnetic activity, *Space Weather*, *9*, S10001, doi:10.1029/2011SW000696.

Tsubouchi, K., and Y. Omura (2007), Long-term occurrence probabilities of intense geomagnetic storm events, *Space Weather*, *5*, S12003, doi:10.1029/2007SW000329.

Uritsky, V. M., A. J. Klimas, and D. Vassiliadis (2001), Comparative study of dynamical critical scaling in the auroral electrojet index versus solar wind fluctuations, *Geophys. Res. Lett.*, *28*(19), 3809–3812, doi:10.1029/2001GL013026.

Wanliss, J. A., and J. M. Weygand (2007), Power law burst lifetime distribution of the *SYM-H* index, *Geophys. Res. Lett.*, *34*, L04107, doi:10.1029/2006GL028235.