

# **S. C. Chapman, N. W. Watkins, and E. Tindale**

## **Reproducible aspects of the climate of space weather over the last five solar cycles**

**Article (Published version)**  
**(Refereed)**

**Original citation:**

Chapman, S. C. and Watkins, N. W. and Tindale, E. (2018) *Reproducible aspects of the climate of space weather over the last five solar cycles*. [Space Weather](#), 16 (8). pp. 1128-1142. ISSN 1542-7390

DOI: [10.1029/2018SW001884](https://doi.org/10.1029/2018SW001884)

Reuse of this item is permitted through licensing under the Creative Commons:

© 2018 The Authors  
CC BY 4.0

This version available at: <http://eprints.lse.ac.uk/90267/>

Available in LSE Research Online: September 2018

LSE has developed LSE Research Online so that users may access research output of the School. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. You may freely distribute the URL (<http://eprints.lse.ac.uk>) of the LSE Research Online website.

RESEARCH ARTICLE

10.1029/2018SW001884

Key Points:

- $F_{10.7}$ ,  $AE$ ,  $D_{ST}$  and solar wind dynamic pressure distributions are characterized for the last five solar maxima
- At solar maximum the tail of the distribution of return times follows a functional form that is solar cycle invariant
- If the mean and variance can be predicted, then the space climate of a given solar maximum is known

Correspondence to:

S. C. Chapman,  
S.C.Chapman@warwick.ac.uk

Citation:

Chapman, S. C., Watkins, N. W., & Tindale, E. (2018). Reproducible aspects of the climate of space weather over the last five solar cycles. *Space Weather*, 16, 1128–1142. <https://doi.org/10.1029/2018SW001884>

Received 6 APR 2018

Accepted 16 JUL 2018

Accepted article online 27 JUL 2018

Published online 29 AUG 2018

©2018. The Authors.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

# Reproducible Aspects of the Climate of Space Weather Over the Last Five Solar Cycles

S. C. Chapman<sup>1,2</sup> , N. W. Watkins<sup>1,2,3,4</sup> , and E. Tindale<sup>1</sup> 

<sup>1</sup>Centre for Fusion, Space and Astrophysics, Department of Physics, University of Warwick, Coventry, UK, <sup>2</sup>Center for Space Physics, Department of Astronomy, Boston University, Boston, MA, USA, <sup>3</sup>Centre for the Analysis of Time Series, London School of Economics and Political Science, London, UK, <sup>4</sup>School of Science, Technology, Engineering and Mathematics, Open University, Milton Keynes, UK

**Abstract** Each solar maximum interval has a different duration and peak activity level, which is reflected in the behavior of key physical variables that characterize solar and solar wind driving and magnetospheric response. The variation in the statistical distributions of the  $F_{10.7}$  index of solar coronal radio emissions, the dynamic pressure  $P_{\text{dyn}}$  and effective convection electric field  $E_y$  in the solar wind observed in situ upstream of Earth, the ring current index  $D_{ST}$ , and the high-latitude auroral activity index  $AE$  are tracked across the last five solar maxima. For each physical variable we find that the distribution tail (the exceedences above a threshold) can be rescaled onto a single master distribution using the mean and variance specific to each solar maximum interval. We provide generalized Pareto distribution fits to the different master distributions for each of the variables. If the mean and variance of the large-to-extreme observations can be predicted for a given solar maximum, then their full distribution is known.

**Plain Language Summary** Earth's near-space plasma environment is highly dynamic, with its own space weather. Space weather impacts include electrical power loss, aviation disruption, interrupted communications, and disturbance to satellite systems. The drivers of space weather, the sun and solar wind, and the response seen at Earth have now been almost continually monitored by ground- and space-based observations over the last five solar cycles (more than 50 years). Each of the last five solar maxima has a different duration and peak activity level and as a consequence the climate of Earth's space weather is also different at each solar maximum. We find that some aspects of the space weather climate are in fact reproducible; they can be inferred from that of previous solar maxima. This may help understand the behavior of future solar maxima.

## 1. Introduction

The space plasma environment near the Earth is highly dynamic, with its own space weather. Space weather impacts include electrical power loss, aviation disruption, interrupted communications, and disturbance to satellite systems (Baker & Lanzerotti, 2016; Hathaway, 2015). The state of the solar wind and the Earth's geomagnetic response have now been almost continually monitored by ground- and space-based observations at reasonably high resolution at over the last five solar cycles. Each of the last five solar cycles has a different duration and peak activity level for the interval around the maximum (hereafter solar maximum interval), which will be reflected in the intensity of space weather activity at Earth. This paper seeks to quantify how these different activity levels during each solar maximum interval are manifested in key space weather relevant physical parameters. It is not a study of solar wind-magnetosphere coupling per se (see Newell et al., 2007, for an example of such a study); however, statistical studies of this kind (Tindale & Chapman, 2016) can shed light on how these coupling parameters vary with solar cycle variation in the overall level of activity.

The solar corona and solar wind drivers, and Earth's geomagnetic response, are routinely characterized by space weather variables that are both multicomponent in their physical origin and can exhibit *long-tailed* statistics such that the largest events are significantly more likely than would be expected from a Gaussian distribution. Space weather events correspond to the tails of these distributions (Baker & Lanzerotti, 2016; Hathaway, 2015), and there is a considerable interest in both reconstructing past space climatology (Lockwood et al., 2017, 2018) and the possibility of quantitative prediction of the probability distribution of these more extreme values for the next solar maximum interval.

The solar corona shows solar cycle variation (Luhmann et al., 2002; Schwenn, 2006) leading in turn to variation in the rate of flares and coronal mass ejections (Cremades & St. Cyr, 2006; Wheatland, 2000) and their geoeffectiveness (Shen et al., 2014; Thomson et al., 2010). While it is well known that overall geomagnetic activity tracks that of the solar cycle (e.g., Baker & Lanzerotti, 2016; Hathaway, 2015; Richardson & Cane, 2012, and references therein), the behavior of the extremes do not necessarily track the time *average*, so that knowledge of the full distribution is needed. For example, the time average of the *aa* index always exceeds a baseline value, and this baseline increases linearly with averaged sunspot number; however, the average *aa* values above this baseline show significant scatter and can take relatively large values even when sunspot numbers are low (Feynman, 1982).

The detailed quantification of the time variation of the statistical distribution of key space weather relevant variables is thus a topical question. The distributions of geomagnetic activity (Tanskanen et al., 2017) and geomagnetic indices (Campbell, 1979; Lockwood et al., 2017, 2018) vary both within and between solar cycles (Tindale & Chapman, 2016, 2017). The distributions of solar wind variables can be non-Gaussian (Veselovsky et al., 2010, and references therein). The distributions of fluctuations in solar wind magnetic energy density (Hnat et al., 2007, 2011; Kiyani et al., 2007) are non-Gaussian and show solar cycle dependence. Fluctuations in geomagnetic indices (Hnat et al., 2002, 2003, 2005) track this behavior across the distribution. Solar wind-magnetosphere coupling parameters (Tindale & Chapman, 2016) also track the variation between and across solar cycles seen in distributions of solar wind magnetic field. There have been recent attempts to survey the statistics of extreme events from sun to solar wind to Earth assuming power law statistics and time stationarity (Riley, 2012). Recently, Hush et al. (2015) found that the underlying distribution functional form of extreme bursts in the *AE* index does not vary significantly with the solar cycle, with their relative intensity tracking the solar cycle, suggesting robust underlying processes that can be modeled on a physical basis. By assuming and fitting a lognormal form for the distribution, Lockwood et al. (2017, 2018) found evidence that the underlying distribution of geomagnetic indices and power input into the magnetosphere are approximately invariant in their underlying distribution functional form, which provided a powerful tool for past space climate reconstruction. However, as Lockwood et al. (2018) note, the lognormal does not necessarily provide the best fit across the full distribution for all variables of interest. Indeed, Tindale and Chapman (2017) found that the distribution of solar wind magnetic field can be far from lognormal. Here we will show how to test whether the functional form of the distribution of a given variable is invariant against solar cycle dependent changes, without assuming any specific functional form, lognormal or otherwise, for the underlying distribution. Only when this invariance is established do we then perform a fitting to specific distributions.

In this paper we focus on key physical variables that are used in the modeling and characterization of space weather impacts at Earth. The *F*10.7 index (Tapping, 2013) of solar coronal radio emissions is used by many operational space weather models as their prime solar input. It is correlated with the density of the upper atmosphere, which in turn has consequences for the design and operation of satellites in low Earth orbit (e.g., Vedder & Tabor, 1992). Solar wind drivers of geomagnetic space weather disturbances include solar wind dynamic pressure  $P_{\text{dyn}}$  and effective convection electric field  $E_y \sim vB_z$  observed in situ upstream of Earth. Two key indices that characterize magnetospheric response at relatively high time resolution as observed at the Earth's surface and are widely used to parameterize the intensity of space weather events are the enhancement of the ring current  $D_{\text{ST}}$ , and auroral activity at high latitudes, the *AE* index (Mayaud, 1980).

Quantile-quantile (QQ) plots (Gilchrist, 2000) compare the statistical distribution of one data sample against that of another. Using a QQ plot analysis over intervals of the maxima and preceding minima of solar cycles 23 and 24, we recently found (Tindale & Chapman, 2016, 2017) that samples of the largest events change in a different manner to the bulk of the distribution as we compared one solar maximum interval to the other. The distribution is multicomponent: It has (at least) two components—a core or bulk of smaller amplitude, more frequently occurring values, and a long tail of larger, less frequently occurring values. The QQ plot can be used to identify the threshold, that is, the value or quantile that is at the crossover between the core of the distribution and the long tail of extreme values. Importantly, the QQ plot can be used to identify this threshold without requiring any fitting of the functional form of the distribution. Looking across the past five solar cycles, we will use this method to identify the threshold at which there is a transition in the data, from the bulk of the distribution to the long tail, for each of the key physical variables *F*10.7,  $P_{\text{dyn}}$ ,  $E_y$ , *AE*, and  $D_{\text{ST}}$ . We then find that the observations in excess of this threshold, the distribution long tail, follow a distribution that only differs in its mean and variance from one solar cycle maximum to the next. The long tail (threshold exceedences) of distributions from a given solar maximum interval then rescale to any other when normalized

to the exceedence mean and standard deviation of that solar maximum interval. Each of the variables  $F10.7$ ,  $P_{\text{Dyn}}$ ,  $E_y$ ,  $AE$ , and  $D_{\text{ST}}$  has a distinct distribution long tail, and each of these for each solar cycle maximum can be described by its own single *master* distribution for that variable along with the exceedence mean and variance for that unique solar maximum interval.

Extreme value theory (EVT) is a well-developed approach to characterizing the statistics of large observations, which has found application in space weather (Thomson et al., 2011). It provides a statistical characterization in terms of generalized extreme value and generalized Pareto distributions (GPDs). Pioneering application to space weather includes the  $F10.7$  index (Vedder & Tabor, 1992), the half-daily  $aa$  index (Silbergleit, 1999; Siscoe, 1976), and the  $D_{\text{ST}}$  index (Silbergleit, 1996; Tsubouchi & Omura, 2007). Another application of EVT in the context of space weather has been to energetic electron fluxes (Koons, 2001; Meredith et al., 2015), which also established an unexpected upper bound to the flux of relativistic *killer* electrons (O'Brien et al., 2007). Important fine-scale details of geomagnetic ground effects have recently been studied using this framework (Beggan et al., 2013; Thomson et al., 2011). EVT studies have recently been performed on extreme solar flares (Elvidge & Angling, 2018) and in the solar wind (Moloney and Davidsen, 2010, 2011), but this work did not differentiate solar cycle dependence as we do here; we will obtain GPD fits for solar maximum intervals of solar cycles 20–24.

However, as well as the important problem of the most extreme events, the natural hazard communities in space weather (Boteler, 1991) and elsewhere, notably hydrology (Rodríguez-Iturbe et al., 1972), have always needed to consider the chance of exceeding more typical, and thus more frequently exceeded, thresholds (Vanmarcke, 2010). But it is still not widely realized that the GPD itself can be applied at thresholds other than just very high ones. Indeed, Hosking and Wallis (1987) noted that the GPD's "applications include use in the analysis of extreme events, in the modeling of large insurance claims, as a failure-time distribution in reliability studies, and in any situation in which the exponential distribution might be used but in which some robustness is required against heavier tailed or lighter tailed alternatives." Here we employ the GPD simply as a flexible fitting distribution. We find that the transition quantile identified by QQ plots identifies an appropriate threshold. Above that threshold we find a distinct GPD master distribution for observations of each of the  $F10.7$  index, the solar wind dynamical pressure, and effective electric field  $E_y$  and the  $D_{\text{ST}}$  and  $AE$  indices. These distributions can be long tailed but are not in the heavy-tailed class (Embrechts et al., 1997), so that a relatively small data sample is required to quantify the mean and variance compared to that needed to resolve the distribution long tail. An estimate of the mean and variance could then be obtained over a relatively short time interval at the beginning of an interval of solar maximum activity, and this would then provide a prediction of the likelihood of occurrence, and hence the return times, that could be expected for the more extreme values that may occur during that solar maximum interval.

## 2. Rescaling and Curve Collapse of Distribution Long Tails

### 2.1. Data Sets and Sampling of the Solar Maximum Intervals

We analyze time intervals (defined below) spanning each of the last five solar maxima using observations on an hourly time base; all of the data analyzed here is from the extended OMNI2 data set extracted from NASA/GSFC's OMNI data set (King & Papitashvili, 2005) through OMNIWeb (the OMNI data were obtained from the GSFC/SPDF OMNIWeb interface at <https://omniweb.gsfc.nasa.gov>). Solar wind variables are observed upstream of Earth's magnetosphere in situ by satellites that at different times are at different locations and so has been time shifted to a common location in the OMNI2 data set.

We present studies of the  $F10.7$  index, the solar wind dynamic pressure  $P_{\text{Dyn}}$ , an estimate of the solar wind convection electric field  $E_y$ , and geomagnetic indices  $AE$  and  $D_{\text{ST}}$ . The  $F10.7$  radio emissions originate high in the chromosphere and low in the corona of the solar atmosphere. The  $F10.7$  index (Tapping, 2013) tracks solar ultraviolet forcing of the upper atmosphere. Many operational space weather models use the  $F10.7$  index as their prime solar input (Codrescu et al., 2012; Dudok de Wit & Bruinsma, 2011). The dynamic pressure (the solar wind momentum flux density [Lopez et al., 1986; Schwenn, 2006])  $P_{\text{Dyn}} = (1 + 4n_{\alpha}/n_p)n_p v_{\text{sw}}^2$  of the solar wind compresses the dayside magnetosphere (Wing & Sibeck, 1997), enhancing the ring current (where  $n_p$  and  $n_{\alpha}$  are the proton and alpha particle densities and  $v_{\text{sw}}$  is the solar wind speed). We also consider the OMNI2 estimate of the solar wind convection electric field  $E_y = -v_{\text{sw}} B_z$  (magnetic field  $B_z$  in GSM), which is a key driver of geomagnetic activity. Geomagnetic indices are derived from ground-based magnetometer observations (Mayaud, 1980) and are widely used to indicate the intensity of space weather events.

**Table 1**  
Number of OMNI2 Hourly Records in Each Solar Maximum Interval Sample Selected by Criterion (i)

Cycle	$F_{10.7}$	$P_{\text{Dyn}}$	$E_y$	$AE$	$-D_{ST}$
20	30,637	16,893	17,608	30,661	30,661
21	30,661	28,101	27,982	30,661	30,661
22	30,661	12,904	12,630	24,061	30,661
23	30,661	30,203	30,569	30,661	30,661
24	30,637	30,222	30,655	30,661	30,661
$q_E$	0.75	0.97	0.9	0.75	0.95

Note. The bottom line gives the threshold quantile  $q_E$  used to distinguish the long tail of the distribution so that, for example, for the  $AE$  index  $q_E = 0.75$ , and 25% of these records are used in the statistical analysis of the long tail of the  $AE$  index.

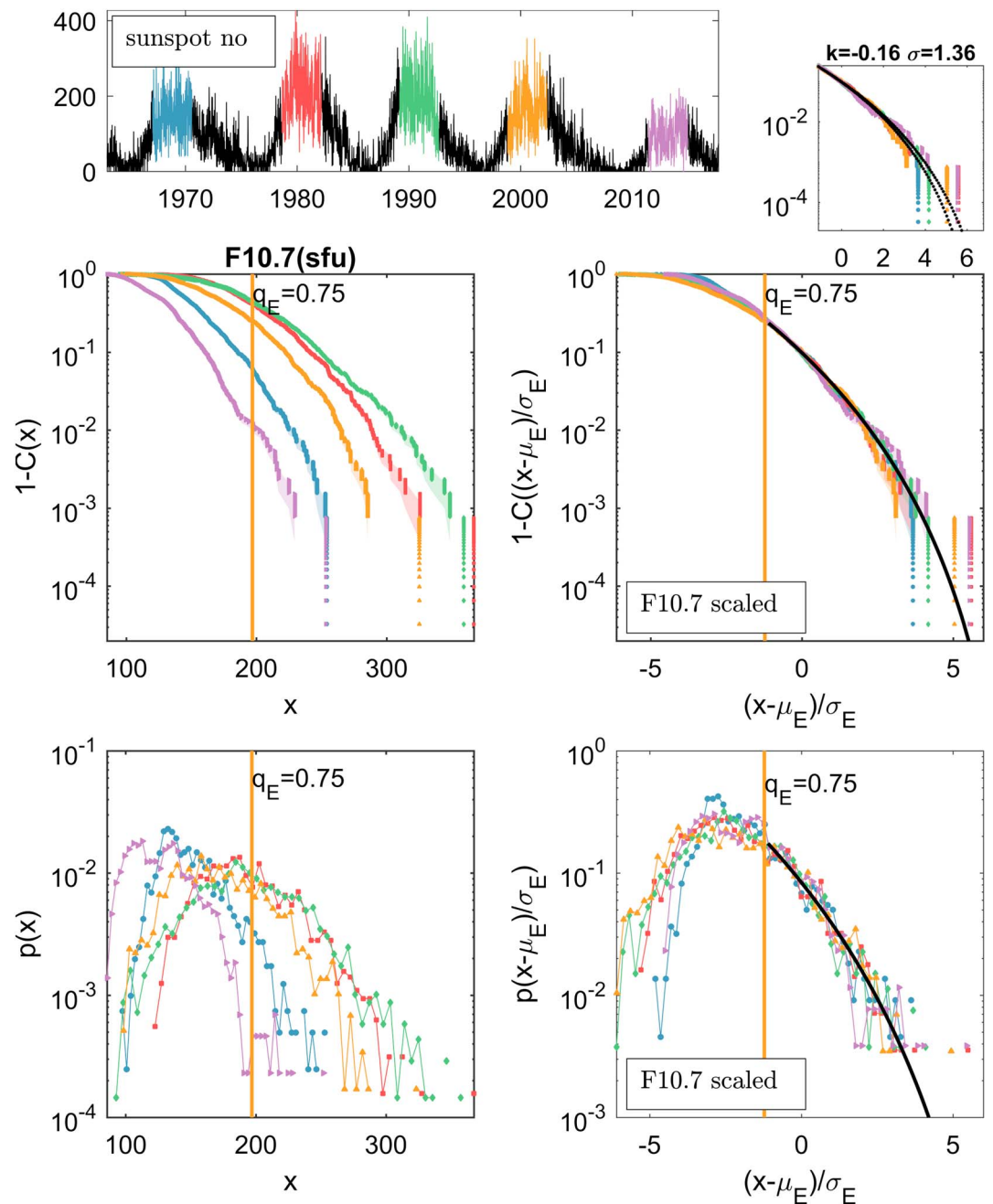
The  $D_{ST}$  index measures low-latitude global variations in the horizontal component of the geomagnetic field, thus representing the strength of the equatorial ring current. The  $AE$  index is comprised of the difference between the most positive and most negative deviations seen at high latitudes.

We will quantify how the statistical distribution of each of these quantities changes from one solar maximum interval to another. In order to form a distribution we need to select time intervals that correspond to each solar maximum interval that are both long enough to capture the more extreme events with statistical significance and short enough that within the sample, the time series can be treated as quasi time stationary. Both the duration and the time between one solar maximum interval and the next are variable. In order to test the robustness of our results we have selected samples of the solar maxima using two different criteria: (i) 3.5-year intervals starting 2.5 years following the previous minimum and (ii) 2-year intervals centered on the monthly sunspot number maximum. We use the following dates (first of each month) for the minima: August 1964, March 1976, September 1986, May 1996, and December 2008; and the maxima: November 1968, December 1979, July 1989, March 2000, and April 2014. We obtained similar results using these two different criteria and we present in detail the results using samples selected with criterion (i). Table 1 gives the number of OMNI2 records in each of these samples. Looking across the past five solar cycles, we will use QQ plots to identify the threshold at which there is a transition in the data, from the bulk of the distribution to the long tail, for each of the key physical variables. These threshold quantiles  $q_E$  are also given in Table 1; the fraction  $1 - q_E$  of these OMNI2 records exceed the threshold and form the distribution long tail.

## 2.2. Curve Collapse of the Distribution Tails

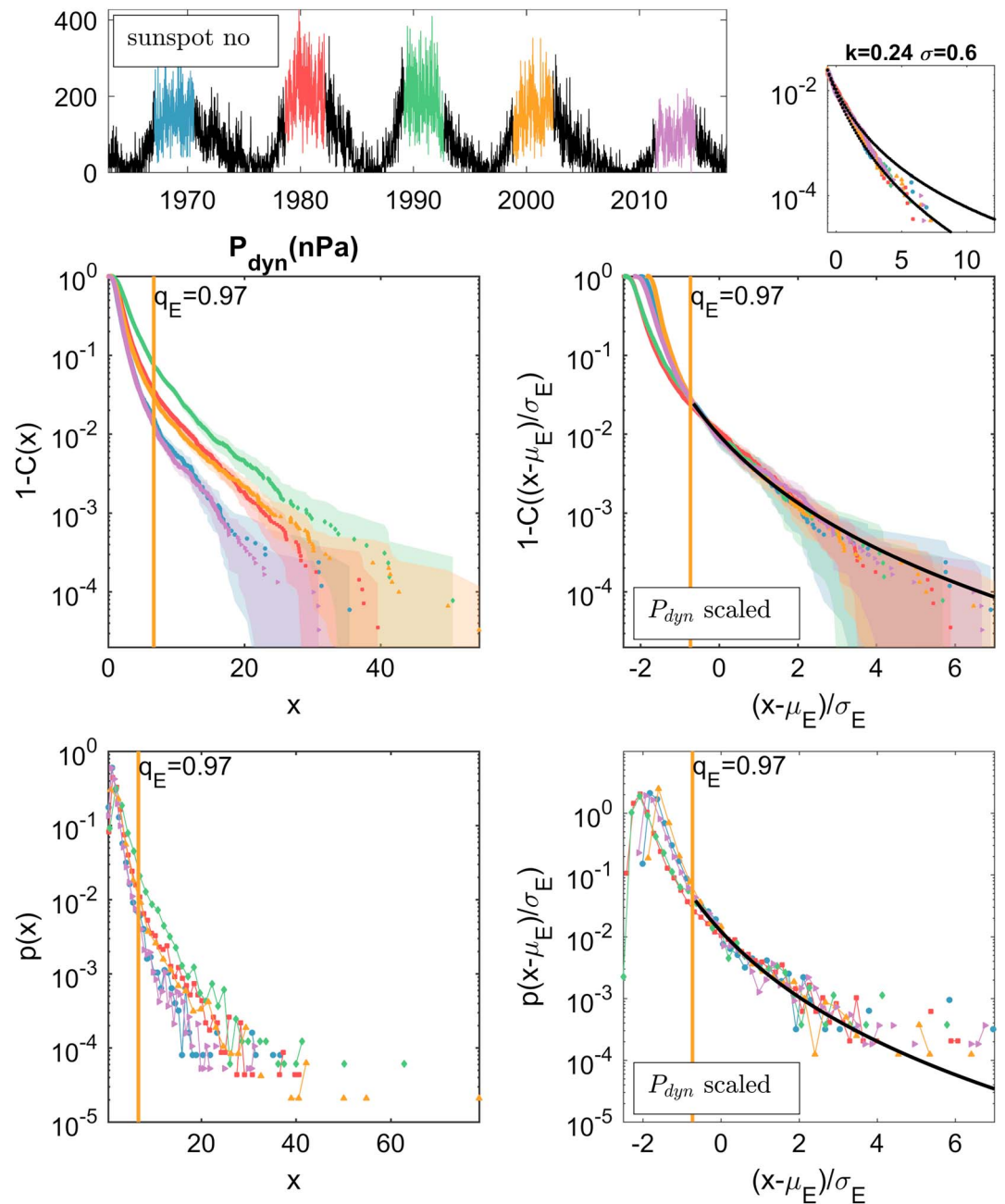
We use the samples of the last five solar maxima to form distributions, which we can plot as the probability density  $P(x)$ , or in terms of the cumulative density function  $C(x)$ , where  $x$  is the  $F_{10.7}$  index (Figure 1), the solar wind dynamic pressure  $P_{\text{Dyn}}$  (Figure 2), the estimated solar wind convection electric field  $E_y$  (Figure 3) or geomagnetic indices  $AE$  (Figure 4) and  $D_{ST}$  (Figure 5). It is convenient to plot the survival distribution  $S(x) = 1 - C(x)$  as this emphasizes the distribution tail of large-to-extreme values, and it gives the return time,  $R(x) = \Delta t / S(x)$ , where  $\Delta t$  is the OMNI2 record cadence of 1 hr as we discuss in more detail in Section 4. The figures plot both the survival functions and the probability densities for comparison. We plot the empirical survival function where for observations  $x_k$ ,  $k = 1 \dots N$ , the  $x$  axis plots the observations in ascending rank order, such that  $k = N$  is the largest value, and the  $y$  axis plots  $1 - k/N$ . Each filled marker on such a plot is a single hourly observation. Uncertainties in the survival functions are indicated by the shaded regions and are estimated using Greenwood's formula (Greenwood, 1926; Kalbfleisch & Prentice, 1980). On the probability density plots the bin-by-bin scatter provides an estimate of the sampling uncertainties; samples are shown down to the one count per bin level.

We focus on the long tails of these distributions, that is, observations that exceed a threshold. The threshold is chosen at a fixed  $q_E$  quantile for each physical variable, using QQ plots as described in detail in the next section. The  $q_E$  quantile identifies the threshold  $x = u(q_E)$  below which the fraction  $q_E$  of observations are found, so that for  $q_E = 0.9$ , 90% of observations take values  $x < u(q_E)$ . The threshold  $u(q_E)$  for fixed  $q_E$  thus varies between one solar maximum interval sample and the next; it is indicated for the maximum of solar cycle 23 on the figures. Other quantiles may be directly read off the survival distribution plots as  $1 - C(x)$  is the fraction of the observations that are found at values greater than  $x$ .



**Figure 1.** Single functional form for the distribution long tail of the F10.7 index. The top panel plots the sunspot number over the last five solar cycles and colored regions indicate the intervals over which data are used to form the plotted distributions for solar cycles 20 (blue), 21 (red), 22 (green), 23 (orange), and 24 (purple). Main panels: Survival distributions (upper panels) and probability densities (lower panels) for the F10.7 index at these maximum intervals of the last five solar cycles. The survival distributions are estimated from the empirical cumulative density function with uncertainties indicated by the shaded regions estimated using Greenwood's formula. The bin-by-bin scatter is indicative of the sampling uncertainty of the probability densities. The left-hand panels plot the distributions of the observed index. Return times  $R(x) = (1\text{hr})/(1 - C(x))$  may be directly read from the survival function plots. The right hand panels plot the distributions of the index rescaled to the mean and standard deviation of the exceedences, that is, observations that exceed the threshold quantile  $u(q_E)$  (indicated for the maximum of solar cycle 23 by the orange line). Overplotted is the generalized Pareto distribution fit for the observations of exceedences aggregated over all five solar maxima. The inset located top-right overplots the 95% confidence bounds of the generalized Pareto distribution fit on the empirical survival distribution.

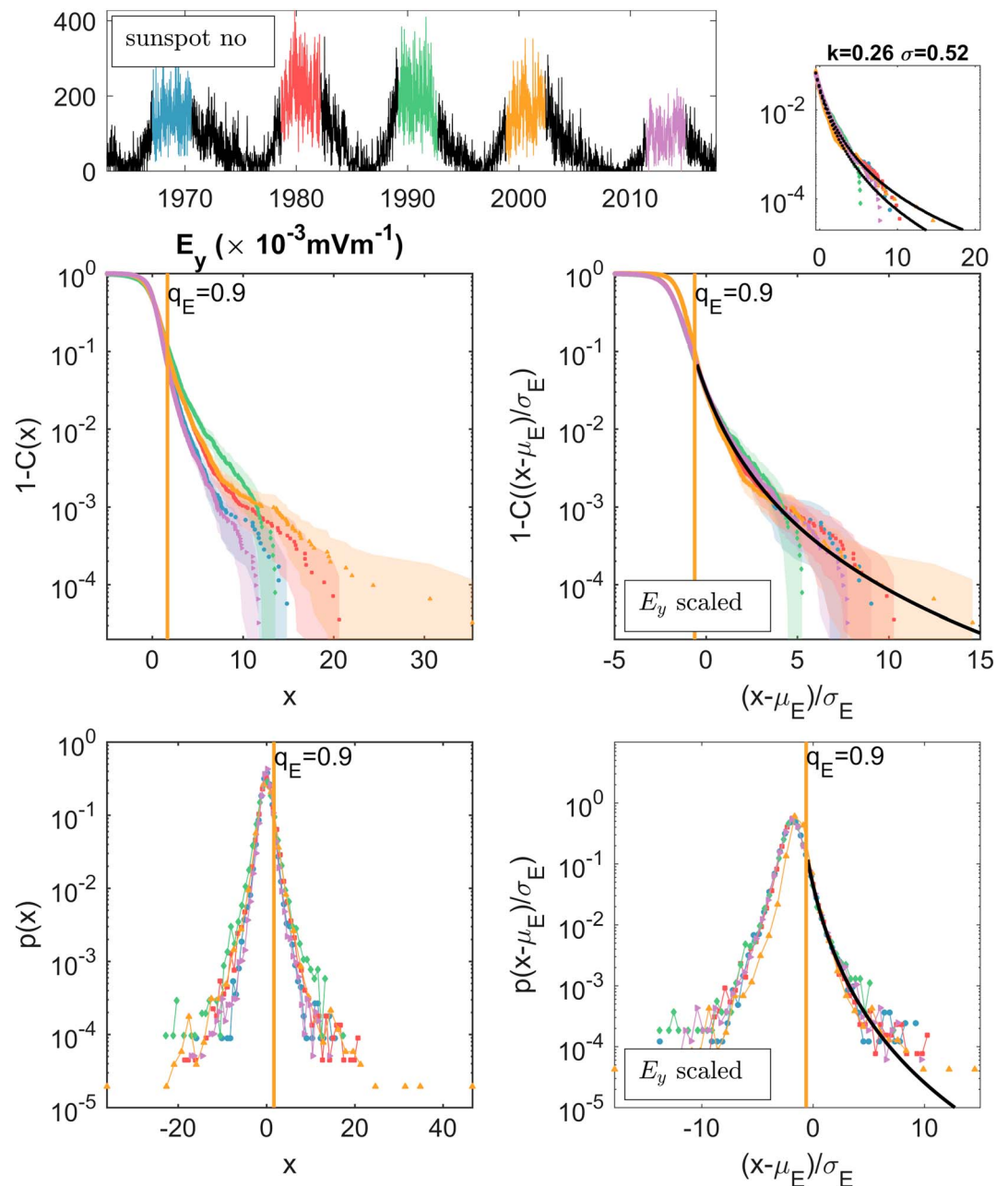




**Figure 2.** Single functional form for the distribution long tail of solar wind dynamic pressure  $P_{\text{dyn}}$  over the last five solar cycles. The format is as in Figure 1.

For each of the samples of the last five solar maxima, the mean  $\mu_E$  and standard deviation  $\sigma_E$  of observations that exceed  $u(q_E)$  are used to rescale the data to  $(x_k - \mu_E)/\sigma_E$ . The right-hand panels of Figures 1–5 then plot the survival functions and the probability densities of the rescaled data. Rescaling the data in this manner collapses the empirical distributions found at different solar maxima onto each other (within uncertainties) for values that exceed  $u(q_E)$ . The possible exception to this is the  $\sim 10$  highest hourly OMNI2 records seen in each 3.5-year-long solar maximum sample of  $-D_{ST}$ . Even though the distributions for the rescaled observations overlap within uncertainties, inspection of the survival distribution suggests a systematic deviation or distribution roll-off, which for  $-D_{ST}$  is most pronounced for the most recent, unusually quiet, solar cycle 24.

This then is our essential result: That for each of the F10.7 index, solar wind  $P_{\text{dyn}}$ , estimated convection  $E_y$ , and the geomagnetic indices AE and  $(-D_{ST})$ , the long tail of the distribution at any of the last five solar maxima



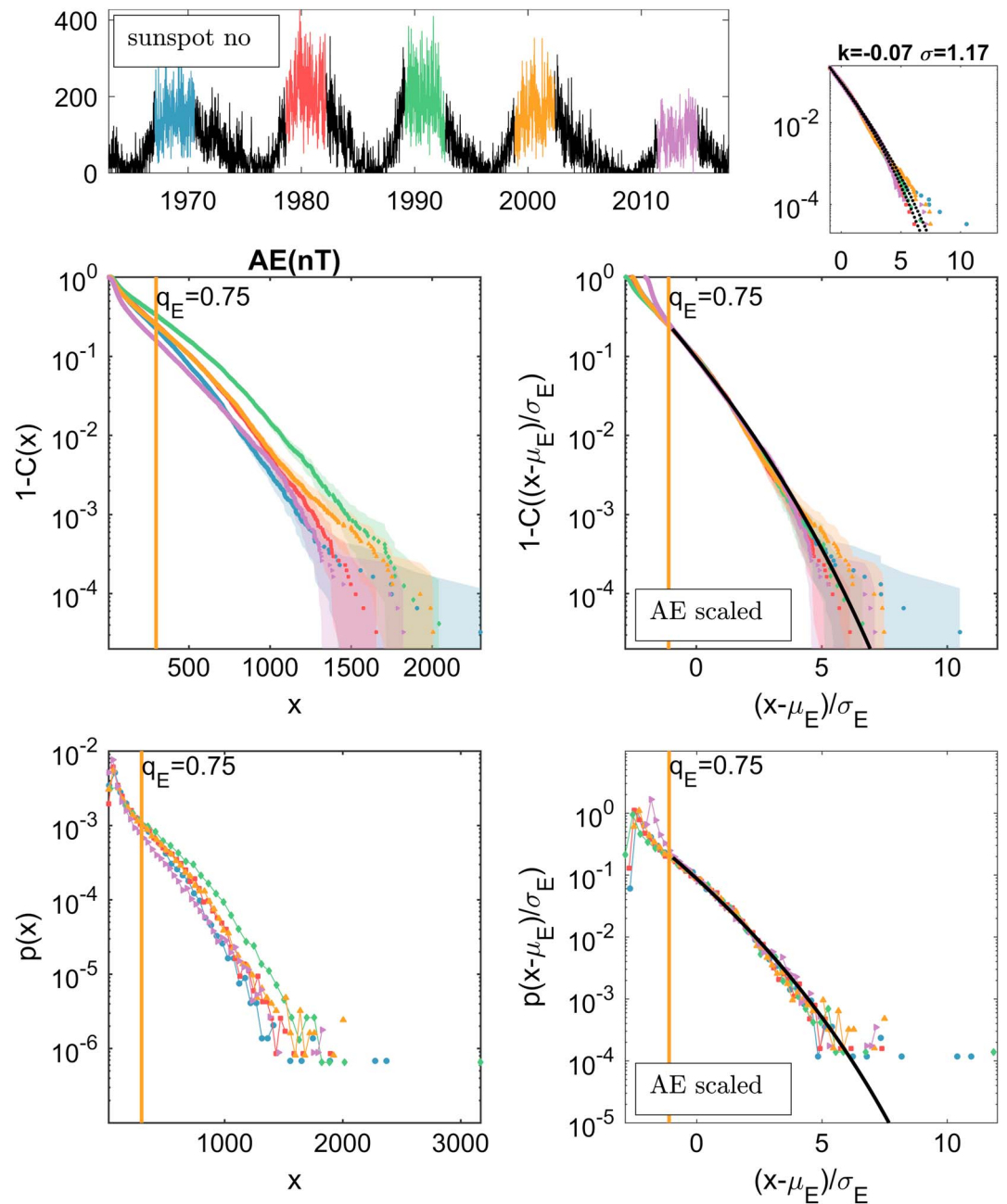
**Figure 3.** Single functional form for the distribution long tail of estimated solar wind convection electric field  $E_y$  over the last five solar cycles. The format is as in Figure 1. Curve collapse is tested on the positive tail, which corresponds to  $-B_z$  IMF.

can be approximately described by a single functional form specific to that physical variable (each has a single master distribution) along with the mean  $\mu_{\epsilon}$  and standard deviation  $\sigma_{\epsilon}$  of that variable's exceedences, specific to each unique solar maximum interval.

### 2.3. QQ Plot Determination of the Thresholds $u(q_F)$

We have taken each of the variables under consideration to be distributed with two components: a core of smaller amplitude, more frequently occurring values, and a long tail of larger, less frequently occurring values that are in excess of a threshold  $q_E$ . We have found that normalizing to the mean and variance of the long tail does not, in general, give a curve collapse for the entire distribution, but it does for the long tail. Therefore, each variable has, for physical reasons, a distinct response in its core and in its long tail to different levels of activity around solar maximum. The value for the threshold  $q_E$  can be determined from the data by using QQ



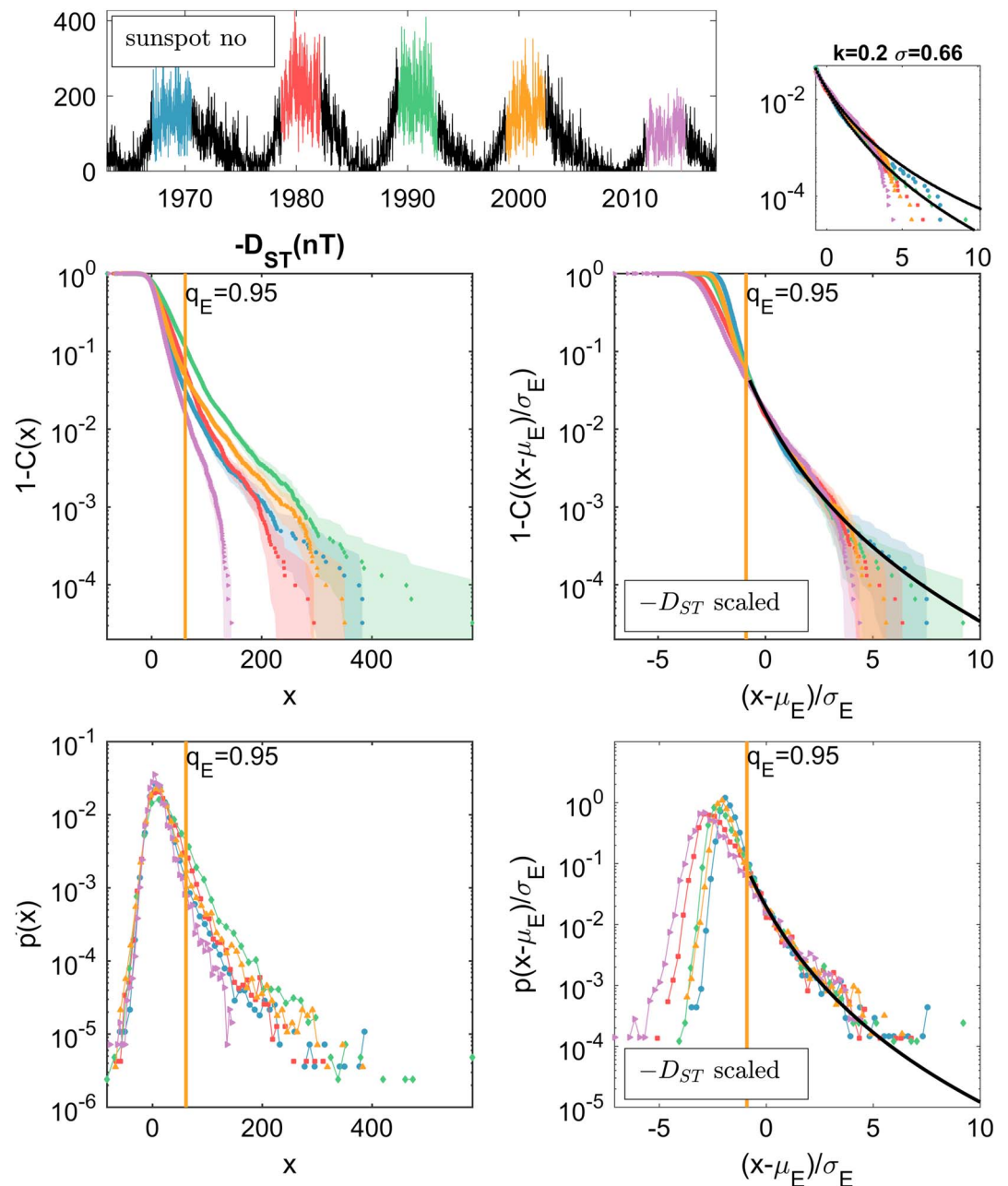


**Figure 4.** Single functional form for the distribution long tail of the AE index over the last five solar cycles. The format is as in Figure 1.

plots. Importantly, the QQ plot can be used to identify this threshold directly from the data without the need to fit specific functions to the observed distribution (Gilchrist, 2000; Tindale & Chapman, 2016, 2017).

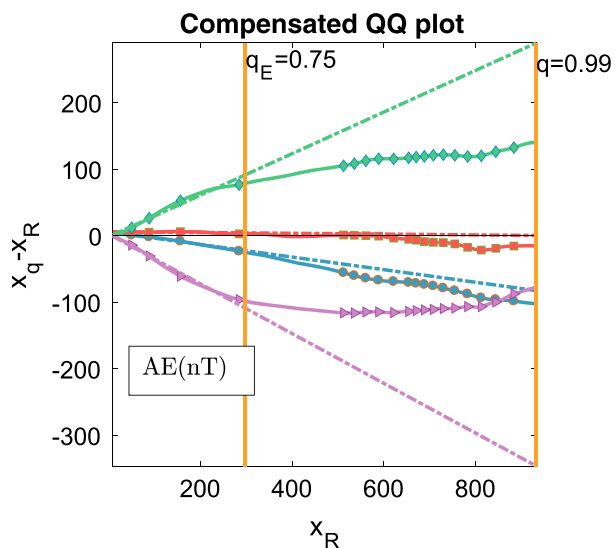
QQ plots of one data sample against another can be used to test whether the samples originate from the same underlying distribution functional form. To compare a sample of observations  $x_q$  with a reference sample  $x_R$  the QQ plot has as its coordinates the quantiles  $(x_R(q), x_q(q))$  where the likelihood  $q$  (value of the cumulative density function) acts as a parametric coordinate. Since the two sets of observations are unlikely to coincide on precisely the same quantiles, it is necessary to resample the data, here using kernel density estimation (using a constant width Gaussian kernel; Coles, 1991; Gilchrist, 2000).

We show an example of one of these plots, for the AE index, in Figure 6. The maximum of solar cycle 23 is used as the reference distribution  $x_R$  as it is recent, and thus less subject to data gaps, but explores a wider



**Figure 5.** Single functional form for the distribution long tail of the  $(-)\ D_{ST}$  index over the last five solar cycles. The format is as in Figure 1.

range of values than the unusually quiet solar cycle 24 maximum. In order to emphasize changes in the distribution, this plot is compensated, that is, we plot  $x_q - x_R$  versus  $x_R$ . If the  $x_q$  and  $x_R$  are drawn from the same distribution, that is, if the distribution for a given solar maximum interval is the same as that for the solar cycle 23 maximum interval, then they have the same quantiles and such a plot gives a straight line  $x_q - x_R = 0$ . If a solar maximum interval has the same underlying distribution functional form  $f$  as that for solar cycle 23 maximum, but has different mean and standard deviation so that  $f(x_q) \sim f(\sigma x_R + \mu)$ , then on the compensated plot  $x_q - x_R = (\sigma - 1)x_R + \mu$ , which will be a single straight line offset from, and at an angle to, the line  $x_q - x_R = 0$ . Straight line fits are overplotted on the  $x_q < u(q_E)$  data. If an entire sample is changing in distribution in the same manner from one solar cycle to another then these straight lines should fit the entire data set. Instead, we can see clear breaks in behavior at  $u(q_E)$ . From the perspective of how the sampled distribution of the physical variable changes from one solar maximum interval to another, this method identifies two components



**Figure 6.** Compensated quantile-quantile (QQ) plots for the solar maximum intervals shown in Figure 4 plotted for the AE index. Quantiles  $x_q$  of samples at the maxima of solar cycles 20 (blue circles), 21 (red squares), 22 (green diamonds), and 24 (purple triangles) are compared to quantiles  $x_R$  of a reference sample, at the maximum of solar cycle 23. To emphasize changes in the distributions w.r.t. the reference we plot compensated quantiles  $x_q - x_R$  versus  $x_R$ . Dash-dot lines are fitted to all data below the quantile corresponding to  $q_E$ . This quantile (shown for the cycle 23 maximum interval, orange line) is approximately at the transition between two regimes in how the distribution changes from one maximum to another.

in the distribution. We then use these plots to approximately identify the quantiles  $u(q_E)$  used as the thresholds for the distribution long tails in Figures 1–5. For simplicity we use a single value of  $q_E$  for each physical variable, identified from QQ plots.

Our finding, that there is more than one physical component to the observed data, which can be identified from their different response to solar cycle changes, is not unexpected; the time series are comprised of relatively quiet intervals punctuated by large bursts, that is, coherent structures in the solar wind seen in  $P_{\text{dyn}}$  and substorms and storms seen in  $AE$  and  $D_{ST}$ . The QQ plots provide a model-independent method to approximately identify the threshold between these two components; the former form the bulk of the distribution, and the latter, the long tail.

### 3. GPD Fits to the Distribution Tails

Within the EVT approach, the GPD can be motivated by the Pickands-Balkema-de Haan theorem (Coles, 1991; Embrechts et al., 1997) and fitted to the observed distribution tail that is then defined as the set of exceedences of observations above a high threshold. This however relies on finding the value of the threshold for which the GPD will hold. In the absence of other information about the system from which the observations derive, this is usually done by finding the range of threshold values for which the mean excess, or the scale and shape parameters of the GPD, do not change (Embrechts et al., 1997; Ghosh & Resnick, 2010). That the GPD should provide a good description does not automatically follow because solar wind variables are not strictly independent, some, for example, have autocorrelation timescales of several tens of hours (Elliott et al., 2013).

The GPD can also be used as a convenient fitting function (Hosking & Wallis, 1987) to characterize the solar cycle independent master distribution for each physical variable. Results of this procedure should then be understood as a model for the observed values rather than a prediction for more extreme events that go beyond the observed range. This is the approach taken here. The Pickands-Balkema-de Haan theorem states that a sample of independent, identically distributed variables  $X$  that exceed a sufficiently high threshold  $\theta$  will have exceedences  $z = (X - \theta)$  that follow the GPD. The GPD's survival distribution is

$$S(z) = 1 - C(z) = \left(1 + k \left(\frac{z}{\sigma}\right)\right)^{-\frac{1}{k}}. \quad (1)$$

We perform a GPD fit to the scaled exceedences of the distribution long tails for each physical variable. For each variable, the samples for the five solar maxima are first independently rescaled to  $(x_k - \mu_E)/\sigma_E$  using each mean  $\mu_E$  and standard deviation  $\sigma_E$  of observations that exceed the threshold  $u(q_E)$  for that solar maximum interval. These five rescaled samples are then aggregated to form a single sample, from which we obtain the exceedences  $z = (x - u(q_E))$  using the quantile  $u(q_E)$  of the aggregated sample. We then perform a maximum likelihood GPD fit to the exceedences of this aggregate over the five solar maxima. An input to the fitting procedure is the threshold  $\theta$  for which we again use the quantile  $u(q_E)$  for  $q_E$  identified from the QQ plots above and for which the distributions collapse onto a single functional form. This is a novel method as it uses two physical insights about the physical variables to determine  $\theta$  (i) that they are expected to be multicomponent and ii) that their multicomponent nature is revealed by how they change in distribution from one solar maximum interval to another. Without this additional physical input, the choice of threshold would rely solely upon standard tests for the GPD, such as a linear dependence of the mean excess on the threshold, or fitted GPD parameters if these can be seen not to vary as the threshold varies.

The GPD fits are overplotted on the right hand panels of Figures 1–5 and can be seen to lie within the data uncertainties. For the survival distributions we also overplot (inset, top right) the 95% confidence intervals of the GPD fit on the data. For  $D_{\text{CT}}$  the single GPD fit does not capture the roll-off seen in the largest few

**Table 2**

Generalized Pareto Distribution Parameters for Fits to Exceedences for the Aggregated Samples of All Five Solar Maxima; the Maxima Are Specified by Two Different Sampling Criteria (i) and (ii)

	$q_E$	$k(i)$	$k(ii)$	$\sigma(i)$	$\sigma(ii)$
$F10.7$	0.75	$-0.157 \pm 0.007$	$-0.088 \pm 0.011$	$1.359 \pm 0.015$	$1.180 \pm 0.019$
$P_{\text{Dyn}}$	0.97	$0.243 \pm 0.041$	$0.275 \pm 0.055$	$0.600 \pm 0.030$	$0.562 \pm 0.038$
$E_y$	0.9	$0.262 \pm 0.021$	$0.234 \pm 0.027$	$0.521 \pm 0.014$	$0.563 \pm 0.020$
$AE$	0.75	$-0.071 \pm 0.006$	$-0.076 \pm 0.009$	$1.169 \pm 0.013$	$1.179 \pm 0.018$
$-D_{ST}$	0.95	$0.198 \pm 0.026$	$0.226 \pm 0.035$	$0.664 \pm 0.022$	$0.625 \pm 0.027$

Note. The exceedence thresholds are at the  $q_E$  quantiles.

$\sim 10$  observations in the interval of solar maximum of cycle 24 and to a lesser extent, that of solar cycle 23. There is also a systematic deviation in  $P_{\text{Dyn}}$  from the GPD although they do coincide within observational statistical uncertainties. These deviations from the GPD fit are for the largest 0.1% of the observations, that is, the largest few tens of hourly OMNI2 records in a 3.5-year-long sample where the uncertainties become large. This suggests a large uncertainty on any extrapolation of the GPD fits to estimate the probability of occurrence of larger, as yet unobserved events. However, we note that other estimates based on fitting specific distributions, and on EVT, also carry large uncertainties (Riley & Love, 2017).

A check on sensitivity to the selected time intervals used to sample each solar maximum is provided by Table 2, which gives the GPD parameters for the two different sampling criteria. These are in agreement within the 95% confidence levels, the exception being the  $F10.7$  index. This can be seen from Figure 1 to have discretization (Tapping, 2013) that is significant in the extreme values, which will in turn affect GPD fitting. An appropriate resampling of the underlying  $F10.7$  data may address this.

The GPD encompasses the three Fisher-Tippett subclasses of EVT (Embrechts et al., 1997). We find that  $F10.7$  is in the Weibull class ( $k < 0$ ), consistent with Vedder and Tabor (1992).  $AE$  ( $k = -0.071$ ,  $\sigma = 1.169$ ) is close to the Gumbel ( $k = 0$ ) case and is almost exponential ( $k = 0$ ,  $\sigma = 1$ ). By contrast  $P_{\text{Dyn}}$ ,  $E_y$ , and  $D_{ST}$  are in the Fréchet class ( $k > 0$ ). The  $F10.7$  index distribution's tail falls off faster than an exponential, while the  $AE$  index decays only just faster than exponentially. This can also be seen in the concave shape of the survival functions when plotted on semilog axes as here. On the other hand the solar wind dynamic pressure and the  $D_{ST}$  index have long tails that decay more slowly than exponential (*subexponential* [Embrechts et al., 1997]), with a correspondingly increased likelihood of extreme events. Importantly for all the exceedence distribution long tails,  $k < 1/2$  so that both the mean and variance are finite (Coles, 1991; Embrechts et al., 1997).

Since  $D_{ST}$  is the ground magnetic field's response to all current systems it may, in addition to capturing changes in the ring current, also be directly influenced by magnetopause currents, which in turn are directly driven by solar wind dynamic pressure (e.g., Zhang et al., 2004, and references therein). Following Gonzalez et al. (1994), we corrected for this and obtained a GPD with fit parameters  $k = 0.199$  and  $\sigma = 0.675$  so that the corrected  $D_{ST}$  long tail has essentially the same distribution as that before correction, consistent with (O'Brien & McPherron, 2000).

The mean excess dependence on threshold for  $D_{ST}$  was previously used by Tsubouchi and Omura (2007) to identify the threshold to obtain a GPD fit for a preselected subset of the observations that identified the most intense events. These most intense events fall into the region of the distribution (see Figure 5 for  $-D_{ST}$ ) where there is a solar cycle dependent roll-off at the largest values.

#### 4. Survival Functions and Return Times

The survival function is directly related to the return time. Consider a return level  $x_m$  that is exceeded once every  $m$  observations. Then the value of the survival function for  $x_m$  is

$$S(x_m) = (1 - C(x_m)) = \frac{1}{m}. \quad (2)$$

If in time interval  $\tau$  we have  $N$  observations, then the average time we will wait to see one observation  $x > x_m$  is the return time  $R(x_m) = m\tau/N$ , which is

$$R(x_m) = \frac{\tau}{N(1 - C(x_m))} = \frac{\Delta t}{(1 - C(x_m))} = \frac{\Delta t}{S(x_m)}, \quad (3)$$

where we make observations every  $\Delta t = 1$  hr here. The return time is an averaged quantity; it indicates the average of the waiting times between the occurrence of events  $x > x_m$ . Importantly, it does not imply that a single observation of  $x > x_m$  will occur once every  $R(x_m)$ . For each solar cycle maximum the return times can thus also be described by their own single master distribution for each variable along with that variable's exceedence mean and variance for that unique solar maximum interval; the master distribution is just that determined for the survival function here.

The GPD is fitted to the distribution exceedences and can be related to the probability of the full set of observations as follows (see Coles, 1991). The GPD is

$$S(X) = \Pr\{X > x \mid X > u\} = \left(1 + k \left(\frac{x - u}{\sigma}\right)\right)^{-\frac{1}{k}}. \quad (4)$$

The probability that an observation exceeds the threshold  $u$  to become part of the population of exceedences, that is,  $X > u$ , is  $S(u) = \Pr\{X > u\}$  so that the probability of exceeding a return level  $x_m$  in the full set of observations is

$$\Pr\{X > x_m\} = S(x_m) = S(u) \left(1 + k \left(\frac{x_m - u}{\sigma}\right)\right)^{-\frac{1}{k}}, \quad (5)$$

which gives the GPD return time from (3). Using (2), the return level estimated from the GPD is

$$x_m = u + \frac{\sigma}{k} \left((mS(u))^k - 1\right). \quad (6)$$

## 5. Discussion and Conclusions

We have quantified the variation between the last five solar maxima of the statistical distributions, that is, survival or return time distributions, of large-to-extreme observations of key physical variables that characterize solar and solar wind driving of space weather and the magnetospheric response: the F10.7 index (Tapping, 2013) of solar coronal radio emissions, the dynamic pressure  $P_{\text{dyn}}$ , and estimated convection electric field  $E_y$  in the solar wind observed in situ upstream of Earth, the enhancement of the ring current  $D_{\text{ST}}$ , and auroral activity at high latitudes, the AE index.

These statistical distributions can be long tailed, and the tail of the distribution that corresponds to the most space weather-effective events changes in a different manner to the bulk of the distribution as we compare one solar maximum interval to another. These components of the distribution may in turn map onto distinct physical processes. QQ plots can be used (Tindale & Chapman, 2016, 2017) to track how an observed distribution has changed from one interval of time to another. Looking across the last five solar maxima, we used QQ plots to identify the threshold between the bulk of the distribution and the tail, which change in a different manner from one solar maximum interval to the next.

Each of the physical variables F10.7,  $E_y$ ,  $P_{\text{dyn}}$ , AE, and  $D_{\text{ST}}$  has a distinct distribution long tail that is comprised of observations above a threshold specific to that variable. We then find that for a given physical variable, the only change in the distribution tail of return times between one solar cycle maximum to the next is in its mean and variance. For each physical variable, the threshold exceedences, that is, the distribution long tail, from a given solar maximum interval rescales to any other when normalized to the exceedence mean and standard deviation of that solar maximum interval. The distinct distribution long tail for each physical variable (F10.7,  $P_{\text{dyn}}$ ,  $E_y$ , AE, and  $D_{\text{ST}}$ ) at each solar cycle maximum interval can be approximately described by a single master distribution for each variable along with the variable's exceedence mean and variance for that unique solar maximum.

If this behavior holds for future solar maxima, then it offers a predictable aspect of space weather climatology. For the class of distributions (GPD with  $k < 1/2$ ) found here, a relatively small data sample is required to quantify the exceedence mean and variance compared to that needed to resolve the distribution long tail. An estimate of the exceedence mean and variance could then be obtained over a relatively short time interval



at the beginning of an interval of solar maximum activity. A prediction for the (large-to-extreme observations) mean and standard deviation could, using the results from our study, be translated into a prediction for the full return time distribution of the large-to-extreme observations. In the case where a variable's distribution of large-to-extreme observations is close to exponential, as found here for  $AE$ , then to a reasonable approximation only a prediction of the mean is required as the variance is then simply related to the mean.

Space weather events seen in these variables are events or *bursts*, that is, intervals in the time series where a given physical variable's value is continually above a crossing threshold. Our results pertain to the occurrence probability of observed values; however, for a stationary time series this will also constrain burst properties from level crossing theory (Boteler, 1991; Cramér & Leadbetter, 2004; Vanmarcke, 2010). If we consider some value for the crossing threshold  $u$ , then the survival function  $1 - C(u)$  at that value of  $u$  constrains the averages  $\langle \dots \rangle$  of the duration of the burst (the time between the upcrossing and the downcrossing)  $\tau_d(u)$  and the time from the start of one burst to the start of the next (the upcrossing time interval)  $\tau_c(u)$  (or alternatively, the burst occurrence frequency  $f_c(u)$ ) through

$$\frac{\langle \tau_d(u) \rangle}{\langle \tau_c(u) \rangle} = \langle \tau_d(u) \rangle \langle f_c(u) \rangle = 1 - C(u). \quad (7)$$

Therefore, our plots of survival functions can be read across directly as plots of  $\langle \tau_d(u) \rangle \langle f_c(u) \rangle$  provided that  $u$  is in the distribution tail. This could in turn be used to quantify the occurrence frequency of different classes of events (corotating interaction regions and coronal mass ejections) provided that another quantity could be used to distinguish the class to which a given burst corresponds.

For each solar cycle maximum interval the the product of average burst duration and average burst frequency can thus also be described by its own single master distribution for each variable along with that variable's exceedence mean and variance for that unique solar maximum; the master distribution is just that determined for the survival function here.

For each physical variable we obtained the corresponding single GPD that characterizes the functional form of the distribution long tail for all of the past five solar cycles. The observations, and their uncertainties, establish the goodness of the GPD fit over the range of values explored by the data. For the  $F10.7$  and  $AE$  indices and  $E_y$  our GPD fit is acceptable from a quite moderate threshold quantile up to the most extreme values that are observed. For  $D_{ST}$  and  $P_{D_{dyn}}$  our GPD fit may not hold for the most extreme (largest 0.1% or few tens of hourly records in a 3.5-year sample) observations;  $D_{ST}$  has an additional solar cycle dependent roll-off in the distribution at very high values for the most recent solar maxima, and  $P_{D_{dyn}}$  has a systematic departure from the GPD.

There is a rich phenomenology in how structures in the solar wind interact with the near-Earth field and plasma environment. The likelihood of a space weather event can in fact depend on a combination of different physical variables; the geoeffectiveness of a large-scale solar wind structure at Earth can be conditioned by the direction of the interplanetary magnetic field (Gonzalez et al., 2008), solar wind Mach number (Borovsky & Denton, 2006; Lavraud & Borovsky, 2008), or by topological details (Lugaz et al., 2016). A natural extension of this work would be to construct joint distribution functions of more than one variable, provided that there is sufficient data to establish statistical significance.

### Acknowledgments

We acknowledge use of NASA/GSFC's Space Physics Data Facility's OMNIWeb service, and OMNI data (the OMNI data were obtained from the GSFC/SPDF OMNIWeb interface at <https://omniweb.gsfc.nasa.gov>). S. C. C acknowledges a Fulbright-Lloyd's of London Scholarship and AFOSR grant FA9550-17-1-0054. E. T. is supported by STFC grant ST/N504506/1. We thank the Santa Fe Institute for hosting a visit during which we worked on this research.

### References

- Baker, D. N., & Lanzerotti, L. J. (2016). Resource letter SW1: Space weather. *American Journal of Physics*, 84, 166–180. <https://doi.org/10.1119/1.4938403>
- Beggs, C., Beamish, D., Richards, A., Kelly, G., & Thomson, A. W. P. (2013). Prediction of extreme geomagnetically induced currents in the UK high-voltage network. *Space Weather*, 11, 407–419. <https://doi.org/10.1002/swe.20065>
- Borovsky, J. E., & Denton, M. H. (2006). The differences between CME-driven storms and CIR-driven storms. *Journal of Geophysical Research*, 111, A07S08. <https://doi.org/10.1029/2005JA011447>
- Boteler, D. H. (1991). Predicting geomagnetic disturbances on power systems. *Eos, Transactions of the American Geophysical Union*, 72(14), 159–160. <https://doi.org/10.1029/EO072i014p00159>
- Campbell, W. H. (1979). Occurrence of  $AE$  and  $d_{ST}$  geomagnetic index levels and the selection of the quietest days in the year. *Journal of Geophysical Research*, 84, 875–881. <https://doi.org/10.1029/JA084iA03p00875>
- Codrescu, M. V., Negrea, C., Fedrizzi, M., Fuller-Rowell, T. J., Dobin, A., Jakowsky, N., et al. (2012). A real-time run of the coupled thermosphere ionosphere plasmasphere electrodynamics (CTIPE) model. *Space Weather*, 10, S02001. <https://doi.org/10.1029/2011SW000736>
- Coles, S. (1991). *An introduction to statistical modelling of extreme values*. New York: Springer.
- Cramér, H., & Leadbetter, M. R. (2004). *Stationary and related stochastic processes: Sample function properties and their applications*. Mineola, New York: Dover.

- Cremades, H., & St. Cyr, O. C. (2006). Coronal mass ejections: Solar cycle aspects. *Advances in Space Research*, 40(7), 1042–1048. <https://doi.org/10.1016/j.asr.2007.01.088>
- Dudok de Wit, T., & Bruinsma, S. (2011). Determination of the most pertinent EUV proxy for use in thermosphere modeling. *Geophysical Research Letters*, 38, L19102. <https://doi.org/10.1029/2011GL049028>.
- Elliott, H. A., Jahn, J.-M., & McComas, D. J. (2013). The Kp index and solar wind speed relationship: Insights for improving space weather forecasts. *Space Weather*, 11, 339–349. <https://doi.org/10.1002/swe.20053>
- Elvidge, S., & Angling, M. J. (2018). Using extreme value theory for determining the probability of Carrington-like solar flares. *Space Weather*, 16, 417–421. <https://doi.org/10.1002/2017SW001727>
- Embrechts, P., Klüppelberg, C., & Mikosch, T. (1997). *Modelling extremal events for insurance and finance*. New York: Springer.
- Feynman, J. (1982). Geomagnetic and solar wind cycles, 1900–1975. *Journal of Geophysical Research*, 87(A8), 6153–6162. <https://doi.org/10.1029/JA087iA08p06153>
- Ghosh, S., & Resnick, S. (2010). A discussion on mean excess plots. *Stochastic Processes and their Applications*, 120, 1492–1517. <https://doi.org/10.1016/j.spa.2010.04.002>
- Gilchrist, W. (2000). *Statistical modelling with quantile functions*. Boca Raton London New York Washington, D.C.: Chapman and Hall.
- Gonzalez, W. D., Joselyn, J. A., Kamide, Y., Kroehl, H. W., Tsurutani, G., Rostoker, B. T., & Vasyliunas, V. M. (1994). What is a geomagnetic storm? *Journal of Geophysical Research*, 99, 5771–5792. <https://doi.org/10.1029/93JA02867>
- Gonzalez, W. D., Tsurutani, B. T., & Gonzalez, A. L. C. (2008). Interplanetary conditions causing intense geomagnetic storms ( $Dst \leq -100$  nT) during solar cycle 23 (1996–2006). *Journal of Geophysical Research*, 113, A05221. <https://doi.org/10.1029/2007JA012744>
- Greenwood, M. (1926). *The natural duration of cancer, in reports of public health and related subjects*. London: HMSO. Vol. 33.
- Hathaway, D. H. (2015). The solar cycle. *Living Reviews in Solar Physics*, 12(4), 1–87. <https://doi.org/10.1007/lrsp-2015-4>
- Hnat, B., Chapman, S. C., Gogoberidze, G., & Wicks, R. T. (2011). Scale-free texture of the fast solar wind. *Physical Review E*, 84(R), 065401. <https://doi.org/10.1103/PhysRevE.84.065401>
- Hnat, B., Chapman, S. C., Kiyani, K., Rowlands, G., & Watkins, N. W. (2007). On the fractal nature of the magnetic field energy density in the solar wind. *Geophysical Research Letters*, 34, L15108. <https://doi.org/10.1029/2007GL029531>
- Hnat, B., Chapman, S. C., & Rowlands, G. (2005). Scaling and a Fokker-Planck model for fluctuations in geomagnetic indices and comparison with solar wind epsilon as seen by WIND and ACE. *Journal of Geophysical Research*, 110, A08206. <https://doi.org/10.1029/2004JA010824>
- Hnat, B., Chapman, S. C., Rowlands, G., Watkins, N. W., & Freeman, M. P. (2002). Scaling in solar wind epsilon and the AE, AL, and AU indices as seen by WIND. *Geophysical Research Letters*, 29(22), 16054. <https://doi.org/10.1029/2002GL016054>
- Hnat, B., Chapman, S. C., Rowlands, G., Watkins, N. W., & Freeman, M. P. (2003). Scaling in long term datasets of geomagnetic indices and solar wind epsilon as seen by WIND. *Geophysical Research Letters*, 30(22), 2174. <https://doi.org/10.1029/2003GL018209>
- Hosking, J. R., & Wallis, J. R. (1987). Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics*, 29(3), 339–349. <https://doi.org/10.2307/1269343>
- Hush, P., Chapman, S. C., Dunlop, M. W., & Watkins, N. W. (2015). Robust statistical properties of the size of large burst events in AE. *Geophysical Research Letters*, 42, 9197–9202. <https://doi.org/10.1002/2015GL066277>
- Kalbfleisch, J. D., & Prentice, R. L. (1980). *The statistical analysis of failure time data*. New York: Wiley and Sons.
- King, J. H., & Papitashvili, N. E. (2005). Solar wind spatial scales in and comparisons of hourly Wind and ACE plasma and magnetic field data. *Journal of Geophysical Research*, 110, A02104. <https://doi.org/10.1029/2004JA010649>
- Kiyani, K., Chapman, S. C., Hnat, B., & Nicol, R. M. (2007). Self- similar signature of the active solar corona within the inertial range of solar wind turbulence. *Physical Review Letters*, 98, 211101. <https://doi.org/10.1103/PhysRevLett.98.211101>
- Koons, H. C. (2001). Statistical analysis of extreme values in space science. *Journal of Geophysical Research*, 106, 10,915–10,921. <https://doi.org/10.1029/2000JA000234>
- Lavraud, B., & Borovsky, J. E. (2008). Altered solar wind-magnetosphere interaction at low Mach numbers: Coronal mass ejections. *Journal of Geophysical Research*, 113, A00B08. <https://doi.org/10.1029/2008JA013192>
- Lockwood, M., Owens, M. J., Barnard, L. A., Scott, C. J., & Watt, C. E. (2017). Space climate and space weather over the past 400 years: 1. The power input to the magnetosphere. *Journal of Space Weather and Space Climate*, 7, A25. <https://doi.org/10.1051/swsc/2017019>
- Lockwood, M., Owens, M. J., Barnard, L. A., Scott, C. J., Watt, C. E., & Bentley, S. (2018). Space climate and space weather over the past 400 years: 2. Proxy indicators of geomagnetic storm and substorm occurrence. *Journal of Space Weather and Space Climate*, 8, A12. <https://doi.org/10.1051/swsc/>
- Lopez, R. E., Freeman, J. W., & Roelof, E. C. (1986). The relationship between proton temperature and momentum flux density in the solar wind. *Geophysical Research Letters*, 13, 640–643. <https://doi.org/10.1029/GL013i007p00640>
- Lugaz, N., Farrugia, C. J., Winslow, R. M., Al-Haddad, N., Kilpua, E. K. J., & Riley, P. (2016). Factors affecting the geoeffectiveness of shocks and sheaths at 1 AU. *Journal of Geophysical Research: Space Physics*, 121, 10,861–10,879. <https://doi.org/10.1002/2016JA023100>
- Luhmann, J. G., Li, Y., Arge, C. N., Gazis, P. R., & Ulrich, R. (2002). Solar cycle changes in coronal holes and space weather cycles. *Journal of Geophysical Research*, 107(A8), 1154. <https://doi.org/10.1029/2001JA007550>
- Mayaud, P. N. (1980). *Derivation, meaning, and use of geomagnetic indices*. *Geophys. 617 Monogr. Ser.*, vol. 22. Washington, DC: American Geophysical Union. <https://doi.org/10.1029/GM022>
- Meredith, N. P., Horne, R. B., Isles, J. D., & Rodriguez, J. V. (2015). Extreme relativistic electron fluxes at geosynchronous orbit: Analysis of GOES E > 2 MeV electrons. *Space Weather*, 13, 170–184. <https://doi.org/10.1002/2014SW001143>
- Moloney, N. R., & Davidsen, J. (2010). Extreme value statistics in the solar wind: An application to correlated Lévy processes. *Journal of Geophysical Research*, 115, A10114. <https://doi.org/10.1029/2009JA015114>
- Moloney, N. R., & Davidsen, J. (2011). Extreme bursts in the solar wind. *Geophysical Research Letters*, 38, L14111. <https://doi.org/10.1029/2011GL048245>
- Newell, P. T., Sotirelis, T., Liou, K., Meng, C.-I., & Rich, F. J. (2007). A nearly universal solar wind-magnetosphere coupling function inferred from 10 magnetospheric state variables. *Journal of Geophysical Research*, 112, A01206. <https://doi.org/10.1029/2006JA012015>
- O'Brien, T. P., Fennell, J. F., Roeder, J. L., & Reeves, G. D. (2007). Extreme electron fluxes in the outer zone Space Weather 5. <https://doi.org/10.1029/2006SW000240>
- O'Brien, T., & McPherron, R. (2000). An empirical phase space analysis of ring current dynamics: Solar wind control of injection and decay. *Journal of Geophysical Research*, 105, 7707–7719. <https://doi.org/10.1029/1998JA000437>
- Richardson, I. G., & Cane, H. V. (2012). Solar wind drivers of geomagnetic storms during more than four solar cycles. *Journal of Space Weather and Space Climate*, 2, A01. <https://doi.org/10.1051/swsc/2012001>
- Riley, P. (2012). On the probability of occurrence of extreme space weather events. *Space Weather*, 10, S02012. <https://doi.org/10.1029/2011SW000734>

- Riley, P., & Love, J. J. (2017). Extreme geomagnetic storms: Probabilistic forecasts and their uncertainties. *Space Weather*, 15, 53–64. <https://doi.org/10.1002/2016SW001470>
- Rodríguez-Iturbe, I., Mejía, J. M., & Dawdy, D. R. (1972). Streamflow simulation: 1. A new look at Markovian models, fractional Gaussian noise, and crossing theory. *Water Resources Research*, 8(4), 921–930. <https://doi.org/10.1029/WR008i004p00921>
- Schwenn, R. (2006). Solar wind sources and their variations over the solar cycle. *Space Science Reviews*, 124, 51–76. <https://doi.org/10.1007/s11214-006-9099-5>
- Shen, C., Chi, Y., Wang, Y., Xu, M., & Wang, S. (2014). Statistical comparison of the ICME's geoeffectiveness of different types and different solar phases from 1995 to. *Journal of Geophysical Research: Space Physics*, 122, 5931–5948. <https://doi.org/10.1002/2016JA023768>
- Silbergleit, V. M. (1996). On the occurrence of geomagnetic storms with sudden commencements. *Journal of Geomagnetism and Geoelectricity*, 48, 1011–1016.
- Silbergleit, V. M. (1999). Forecast of the most geomagnetically disturbed days. *Earth Planets Space*, 51, 19. <https://doi.org/10.1186/BF03352205>
- Siscoe, G. L. (1976). On the statistics of the largest geomagnetic storms per solar cycle. *Journal of Geophysical Research*, 81, 4782. <https://doi.org/10.1029/JA081i025p04782>
- Tanskanen, E. I., Hynönen, R., & Mursula, K. (2017). Seasonal variation of high latitude geomagnetic activity in individual years. *Journal of Geophysical Research: Space Physics*, 122, 10,058–10,071. <https://doi.org/10.1002/2017JA024276>
- Tapping, K. F. (2013). The 10.7 cm solar radio flux (F10.7). *Space Weather*, 11, 394–406. <https://doi.org/10.1002/swe.20064>
- Thomson, A. W., Dawson, E. B., & Reay, S. J. (2011). Quantifying extreme behaviour in geomagnetic activity. *Space Weather*, 9, S10001. <https://doi.org/10.1029/2011SW000696>
- Thomson, A. W. P., Gaunt, C. T., Cilliers, P., Wild, J. A., Opperman, B., McKinnell, L.-A., et al. (2010). Present day challenges in understanding the geomagnetic hazard to national power grids. *Advances in Space Research*, 45, 1182–1190. <https://doi.org/10.1016/j.asr.2009.11.023>
- Tindale, E., & Chapman, S. C. (2016). Solar cycle variation of the statistical distribution of the solar wind  $\epsilon$  parameter and its constituent variables. *Geophysical Research Letters*, 43, 5563–5570. <https://doi.org/10.1002/2016GL068920>
- Tindale, E., & Chapman, S. C. (2017). Solar wind plasma parameter variability across solar cycles 23 and 24: From turbulence to extremes. *Journal of Geophysical Research: Space Physics*, 122, 9824–9840. <https://doi.org/10.1002/2017JA024412>
- Tsubouchi, K., & Omura, Y. (2007). Long-term occurrence probabilities of intense geomagnetic storm events. *Space Weather*, 5, S12003. <https://doi.org/10.1029/2007SW000329>
- Vanmarcke, E. (2010). *Random fields: Analysis and synthesis* (2nd ed). Singapore: World Scientific.
- Vedder, J. D., & Tabor, J. L. (1992). Solar F10.7 Radiation: A short-term statistical model. *Journal of Spacecraft and Rockets*, 29, 76–80. <https://doi.org/10.2514/3.26316>
- Veselovsky, I. S., Dmitriev, A. V., & Suvorova, A. V. (2010). Algebra and statistics of the solar wind. *Cosmic Research*, 48(2), 113–128. <https://doi.org/10.1134/S0010952510020012>
- Wheatland, M. S. (2000). The origin of the solar are waiting-time distribution. *Astronomy Journal*, 136(2), L109–L112. <https://doi.org/10.1086/312739>
- Wing, S., & Sibeck, D. G. (1997). Effects of interplanetary magnetic field Z component and the solar wind dynamic pressure on the geosynchronous magnetic field. *Journal of Geophysical Research*, 102, 7207–7216. <https://doi.org/10.1029/97JA00150>
- Zhang, J., Liemohn, M. W., Kozyra, J. U., Lynch, B. J., & Zurbuchen, T. H. (2004). A statistical study of the geoeffectiveness of magnetic clouds during high solar activity years. *Journal of Geophysical Research*, 109, A09101. <https://doi.org/10.1029/2004JA010410>