Some Experiments in Extreme-Value Statistical Modeling of Magnetic Superstorm Intensities

Jeffrey J. Love

1Geomagnetism Program, U.S. Geological Survey, Denver, CO, USA

Abstract  In support of projects for forecasting and mitigating the deleterious effects of extreme space weather storms, an examination is made of the intensities of magnetic superstorms recorded in the $D_{st}$ index time series (1957–2016). Modified peak-over-threshold and solar cycle, block-maximum samplings of the $D_{st}$ time series are performed to obtain compilations of storm maximum $-D_{stm}$ intensity values. Lognormal, upper limit lognormal, generalized Pareto, and generalized extreme-value model distributions are fitted to the $-D_{stm}$ data using a maximum-likelihood algorithm. All four candidate models provide good representations of the data. Comparisons of the statistical significance and goodness of fits of the various models give no clear indication as to which model is best. The statistical models are used to extrapolate to extreme-value intensities, such as would be expected (on average) to occur once per century. An upper limit lognormal fit to peak-over-threshold $-D_{stm}$ data above a superstorm threshold of 283 nT gives a 100-year extrapolated intensity of 542 nT and a 68% confidence interval (obtained by bootstrap resampling) of [466, 583] nT. A generalized extreme-value distribution fit to solar cycle, block-maximum $-D_{stm}$ data gives a nine-solar cycle (approximately 100-year) extrapolated intensity of 591 nT. The $D_{st}$ data are found to be insufficient for providing usefully accurate estimates of a statistically theoretical upper limit on magnetic storm intensity. Secular change in storm intensities is noted, as is a need for improved estimates of pre-1957 magnetic storm intensities.

Plain Language Summary  Estimating the statistical likelihood of extreme-value geomagnetic storm intensities depends on the selection of data and models. The data used here are discrete samples of a standard time series that measures geomagnetic disturbance. Different methods are examined for identifying data values that can serve as nonredundant representations of storm maximum intensity. Four different statistical models (two formal models and two physically motivated models) are examined as candidate descriptions of the data. Comparisons and contrasts are made of the goodness of model descriptions of the data. Extrapolations are made to estimate once-per-century storm intensities; an upper limit on storm intensity cannot be confidently established; slow secular change is noted in the occurrences of intense storms. General suggestions are made for improving inferences about the future occurrence of intense storms. Results inform projects concerned with forecasting and mitigating the deleterious effects of extreme space weather storms.

1. Introduction

Rare, but extremely intense, space weather storms (e.g., Cliver & Dietrich, 2013; Jiggens et al., 2014; Riley et al., 2018) are hazardous for a variety of modern technologies and their operations (e.g., Baker, 2008; Cannon et al., 2013; Daglis, 2005). The storm of May 1967, for example, brought widespread disruption to radio communication, notably including that of the U.S. military (e.g., Knipp et al., 2016). The great storm of March 1989 induced the failure of the high-voltage Canadian Hydro-Québec electricity power grid, damaged and interfered with the operation of satellites, and disrupted geophysical surveys and over-the-horizon radio communication in many parts of the world (e.g., Allen et al., 1989). The space weather storms of October and November 2003 caused an electricity blackout in Sweden (e.g., Lundstedt, 2006), damaged and interfered with satellites, degraded the accuracy of global positioning systems, disrupted radio communication, and forced commercial airlines to reroute transpolar flights in order to maintain in-flight communication and avoid excessive human exposure to radiation (e.g., Balch et al., 2004).

Since intense space weather storms can be so widely and significantly consequential, it is important to estimate their occurrence frequency with as much accuracy as possible and also to estimate confidence intervals...
Figure 1. The $\text{Dst}$ time series, 1957–2016 (solar cycles 19–24). The seven highest peak-over-threshold $-\text{Dst}_{\text{m}}$ values (7) and the six solar cycle, block-maximum $-\text{Dst}_{\text{BM}}^{\text{m}}$ values (BM) are labeled.

(“error estimates”) on their occurrence frequency. Obtaining such estimates is usually accomplished by fitting statistical models to a relatively small number of historical data recording the most intense storms—this permits extrapolation to even greater storm intensities. A variety of models are used. Some of these models are motivated less by physics than by formal statistical theories (e.g., Albeverio et al., 2006; Coles, 2001; Embrechts et al., 1997), under which extreme data are viewed as conditional samples from the far tail of a hypothetical parent distribution. Extreme-value theory has been applied in many areas of geophysics (e.g., Corral & González, 2019; Ghil et al., 2011; Papalexioú & Koutsoyiannis, 2013), including for the analysis of extreme solar flare intensities, coronal mass ejection speeds, solar proton fluences, solar radio bursts, magnetic storm intensities, and geoelectric field amplitude. Formal statistical models used in analyses of these space weather phenomena include the generalized Pareto distribution (e.g., Elvidge & Angling, 2018; Koons, 2001; Lotz & Danskin, 2017; Thomson et al., 2011; Tsiftsi & De la Luz, 2018; Tsubouchi & Omura, 2007) and the generalized extreme-value distribution (e.g., Nikitina et al., 2016; Schumann et al., 2013; Tsubouchi & Omura, 2007), as well as more specialized extreme-value distributions, such as the exponential (e.g., Echer et al., 2011), Fréchet (e.g., Weigel & Baker, 2003), Gumbel (e.g., Willis et al., 1997), power law (e.g., Aschwanden et al., 2016; Bougeret et al., 1984; Kataoka, 2013; Riley, 2012; Ruzmaïkin et al., 2011; Yermolaev et al., 2013), and Weibull (e.g., Gopalswamy; Moriña et al., 2019; Watari et al., 2001). In contrast, other analyses of extreme space weather events have assumed that storm intensity data are realized (in probability) from a physically plausible stochastic process, and, often, a lognormal statistical model is used (e.g., Love et al., 2015, 2016a; Pulkkinen et al., 2012; Riley & Love, 2017; Verbeeck et al., 2019).

The selection of data and statistical models can significantly affect extrapolated extreme-event probabilities. But one might reasonably ask: How much difference does it all make? Seeking a specific answer to this question, we focus attention on magnetic storm disturbance as measured by $\text{Dst}$ index time series (1957–2016). We sample this time series in different ways to obtain storm maximum intensities and solar cycle, storm maximum intensities. As statistical descriptions of the largest storm maximum intensities, we compare and contrast formal extreme-value models with lognormal models in terms of their theoretical properties, their capacities for fitting the data, and the extrapolated estimates they permit of 100-year storm intensities. Results inform coordinated projects (e.g., Hapgood, 2011), both in the United States (National Science and Technology Council, 2015a, 2015b, 2019) and internationally (e.g., Schrijver et al., 2015), concerned with anticipating and mitigating the deleterious effects of extreme space weather storms.

2. The $\text{Dst}$ Time Series

The $\text{Dst}$ index time series is the longitudinal average of low-latitude, ground-level geomagnetic disturbance (Sugiura, 1964; World Data Center for Geomagnetism et al., 2015). It is useful for identifying the primary evolutionary phases of individual magnetic storms (e.g., Loewe & Prölls, 1997; McPherron) and for long-term statistical analysis on storm occurrence rates. During the main phase of a magnetic storm, intensification of the westward directed magnetospheric ring current causes a proportional decrease in low-latitude, horizontal-component geomagnetic intensity. This causes a storm time decrease of $\text{Dst}$ to negative values...
Figure 2. The $D_{st}$ time series for November 1957. The quiet time level $\beta$ is labeled, as are four storm maximum $-D_{stm}$ values for periods defined by the begin ($\tau_b$) and end ($\tau_e$) times.

(e.g., Daglis, 2001) that is so distinctive it is sometimes taken to be the defining characteristic of a magnetic storm (e.g., Gonzalez et al., 1994).

The Kyoto World Data Center in Japan calculates the standard version of $D_{st}$ from hourly mean, geomagnetic time series collected at four long-running observatories (Sugiura & Kamei, 1991): Hermanus (South African National Space Agency) (e.g., Kotzé, 2018), Kakioka (Japan Meteorological Agency) (e.g., Minamoto, 2013), and Honolulu and San Juan (both operated by the U.S. Geological Survey) (e.g., Love & Finn, 2011). A nonstormy, quiescent baseline is subtracted from each observatory time series; the residual disturbances are averaged, with latitude weighting, to give the 1-hr resolution $D_{st}$ time series. Each hourly $D_{st}$ value is centered on the bottom of the Universal Time hour: 00:30, 01:30, etc. We use $D_{st}$ for years 1957–2016, a 60-year extent of time that encompasses numerous magnetic storms of various intensities that occurred over the course of six solar cycles, from the rising phase of solar cycle 19 through to just past the maximum of cycle 24; see Figure 1. There are 525,960 separate hourly $D_{st}$ data values in the time series considered here; there are no gaps or obvious erroneous values.

3. Sampling for Extreme Intensities

The $D_{st}$ time series is autocorrelated, and, especially from hour to hour, it is serially correlated. Individual magnetic storms typically evolve over the course of a couple of days (or so) and within the depth of a storm’s main phase, the particular hour when $-D_{st}$ attains its maximum value is typically preceded and followed by durations of time that also exhibit negative $D_{st}$ values. Furthermore, the occurrence of one storm is often related to the occurrence in time of another storm. So, for example, a persistent sunspot group might be the source of several coronal mass ejections causing separate magnetic storms (e.g., Gonzalez et al., 2011); similarly, long-lasting coronal holes can give rise to a series of magnetic storms, often separated in time by the duration of one solar rotation (e.g., Richardson & Cane, 2012; Tsurutani et al., 2006). Furthermore, an unusually active solar cycle can give rise to more extremely intense magnetic storms than a relatively subdued solar cycle, and, insofar as the activity level of a given solar cycle represents the state of the quasi-oscillatory solar dynamo, magnetic storms within each cycle are not strictly statistically independent. Recognizing the autocorrelational properties of $D_{st}$ time series while, at the same time, seeking to characterize the occurrence rate of extreme-intensity magnetic storms in terms of statistical models, we sample the $D_{st}$ time series in two different ways that are inspired by standard (conditional) sampling methods used in extreme-event analysis (e.g., Coles, 2001; Mudelsee, 2014, Chapter 6).

First, we use a modified “peak-over-threshold” algorithm similar to that used by others in related analyses (Lotz & Danskin, 2017; Love et al., 2015, 2018). The 1-hr $D_{st}$ values are treated in a manner illustrated in Figure 2 for a series of storms that occurred in September 1957. We begin by ranking all of the $D_{st}$ values by size; the greatest maximum $-D_{stm}$ value is identified from among all the storms, and this value is set aside for statistical analysis. Next, a quiet level $\beta$ is chosen that, itself, is significantly less than the $-D_{stc}$ threshold delimiting the weakest storm maximum value we analyze. For this, we choose $\beta = 30$ nT, but we have found that the statistical results presented here are relatively insensitive to the choice of other similar quiet time levels. Then, a stormy period is defined by the last previous (first subsequent) moment $\tau_b$ ($\tau_e$).
in time representing the beginning (end) of the storm such that all \(-D_{st}\) values within the stormy period are greater than \(\beta\). While keeping the maximum \(-D_{st_m}\) value, the other \(-D_{st}\) values within this stormy period are excluded from further analysis. The remaining \(-D_{st}\) values are ranked, the next largest maximum \(-D_{st_m}\) value is identified and kept, and, again, other previous and subsequent stormy values are excluded, etc. This compilation process is terminated when the highest remaining \(-D_{st_m}\) value falls below \(-D_{st_c}\). As a quality check, we visually inspect the periods identified as stormy, especially those with maxima that are marginally above the cutoff \(-D_{st_c}\) threshold. Our process of trimming each storm’s \(-D_{st}\) time series back to a quiet time level mostly excludes redundant countings that would otherwise occur for storms with complex evolution—those with \(-D_{st}\) time series that momentarily exceed \(-D_{st_c}\), then fall below it, and then exceed it again. If such redundant exceedances were not excluded, then some storms would be counted more than once. For peak-over-threshold sampling, our algorithm yields a compilation similar to those of Tsubouchi & Omura (2007, Section 2); for 1957–2001, we estimate \(N = 336\) (\(N = 26\)) with \(-D_{st_c} > 100\) nT (\(-D_{st_c} > 280\) nT), and they estimate \(N = 322\) (\(N = 26\)).

For the month of September 1957, we note the occurrence of four separate intense magnetic storms. The storm of 13 September 1957 attained a maximum \(-D_{st_m}\) value of 427 nT at 10:30 Universal Time, the third-highest storm maximum value attained in the 1957–2016 extent of the \(D_{st}\) time series. This value is counted once in our statistical analysis. Similarly, the great storm of March 1989, with its maximum \(-D_{st_m}\) value of 589 nT, the highest recorded in the \(D_{st}\) time series, is only counted once. The great storm of July 1959, with a maximum \(-D_{st_m}\) of 429 nT, is also counted only once, etc. For the years 1957–2016, we identify \(N = 390\) separate storms with maximum \(-D_{st_m}\) exceeding 100 nT; \(N = 70\) storms have \(-D_{st_m}\) exceeding 200 nT, \(N = 21\) exceed 300 nT, and \(N = 5\) exceed 400 nT. Although our peak-over-threshold algorithm removes most of the intrastorm serial correlation found in the \(D_{st}\) time series, some interstorm autocorrelation (or inter–solar cycle clustering) might remain. We note that three of the four most intense storms from 1957–2016 occurred in 1957, 1958, and 1959, each within solar cycle 19. To estimate the probability of such a clustering, we recognize that the number of ways \(k\) indistinguishable objects can be arranged into \(n\) different boxes is \(\binom{n+k-1}{k}\). The probability, then, that three out of four storms would be realized in any one of six solar cycles is \(6 \times \binom{3}{1} / \binom{4}{4} = 0.2381\). This is not an especially low probability, and so inter–solar cycle clustering might be regarded as relatively small.

Second, we use “block-maximum” sampling, whereby peak data values are taken from representative windows of time covering a sequence of data. Block-maximum sampling is used, especially, in the analysis of phenomena exhibiting periodic or quasiperiodic variation. So, for example, hydrological analysis is sometimes made of annual–maximum rainfall or annual–maximum river flow (e.g., Jenkinson, 1955; Westra et al., 2013). For this study of magnetic storm occurrence statistics, the most natural modulation to consider is that due to the solar cycle. As shown in Figure 1, for the six solar cycles (19 through 24) covered by the extent of the \(D_{st}\) time series 1957–2016, we identify six \(-D_{st_m}^\text{BM}\) with block-maximum storms occurring in [July 1959, May 1967, July 1982, March 1989, November 2003, March 2015]. Note that block-maximum sampling of solar cycles omits some very intense magnetic storms—the great storm of September 1957, for example, is not in the block-maximum sampling because it is eclipsed in intensity by the storm of July 1959. For comparison, in Figure 1, we show the seven most intense storms as defined by peak-over-threshold sampling and occurring in [September 1957, February 1958, July 1959, May 1967, March 1989, March 2001, November 2003]—the storm of May 1967 had the same maximum intensity, to within the 1-nT resolution of \(D_{st}\), as the storm of March 2001.

4. Theoretical Statistical Models

We assume that the solar-terrestrial conditions that gave rise to extremely intense magnetic storms in the past will be seen again in the future. Under that assumption, we use statistical models, fitted to the \(-D_{st_m}\) data, to estimate storm occurrence rates as a function of storm intensity and to make extrapolated estimates of the occurrence rate of storms that are even more intense than those recorded in the \(-D_{st_m}\) data. For many statistical analyses, the normal distribution is of fundamental importance; for a random variable \(x\), usually assumed to be independent, the probability density function is

\[
f_X(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[ -\frac{(x - \mu)^2}{2\sigma^2} \right], \quad \text{where } x \in (-\infty, +\infty),
\]
where \( \mu \) and \( \sigma^2 \) are the \( x \) population mean (or “location”) and variance (or “scale”), and the corresponding complementary cumulative distribution is

\[
P_{\mathcal{L}}(x|\mu, \sigma) = \int_x^\infty f_{\mathcal{L}}(\theta|\mu, \sigma) \, d\theta = \frac{1}{2} \operatorname{erf} \left( \frac{\ln x - \mu}{\sqrt{2}\sigma} \right).
\]

(2)

The relative ubiquity of the normal distribution is due to the central limit theorem (e.g., Bohm & Zech, 2010, Chapter 13.1.2), according to which the addition of many underlying statistical processes, described in terms of distributions having means and variances, will result in data that are normally distributed.

In contrast, a positive random variable \( x \) is lognormally distributed if \( \ln x \) is normally distributed. With a change of variables that conserves differential probability, the random variable \( x \) has the density function

\[
f_{\mathcal{L},\mathcal{N}}(x|\mu, \sigma) = \frac{1}{x \sqrt{2\pi\sigma^2}} \exp \left[ -\frac{(\ln x - \mu)^2}{2\sigma^2} \right], \quad \text{where } x \in (0, +\infty),
\]

(3)

and where \( \ln x \) denotes the natural logarithm of \( x \), and \( \mu \) and \( \sigma^2 \) are the \( \ln x \) population mean and variance. The corresponding complementary cumulative distribution is

\[
P_{\mathcal{L},\mathcal{N}}(x|\mu, \sigma) = \int_x^\infty f_{\mathcal{L},\mathcal{N}}(\theta|\mu, \sigma) \, d\theta = \frac{1}{2} \operatorname{erf} \left( \frac{\ln x - \mu}{\sqrt{2}\sigma} \right).
\]

(4)

It is important to recognize that a central limit theorem also applies for the lognormal distribution, according to which the multiplication of many independent random variables, each of which is positive and drawn from a distribution having a mean and variance, will be lognormally distributed (e.g., Aitchison & Brown, 1957, Chapter 2.6; Bohm & Zech, 2010, Chapter 3.6.10; Crow & Shimizu, 1988, p. 5).

Following on from previous work (Love et al., 2015, 2018), we hypothesize that storm maximum intensities, as represented by \( -D_{st,m} \), are realized from a giant lognormal stochastic process. In terms of underlying processes that might give rise to a lognormal distribution of magnetic storm intensities, we note, first, that the solar dynamo, with quasiperiodic autoregressive amplification and deamplification of magnetic energy (e.g., Charbonneau, 2010; Conway, 1998; Solanki et al., 2000), is manifest as an 11-year modulation of sunspot number and a corresponding modulation of the rate of coronal mass ejections and geomagnetic storms. Second, the geoeffectiveness of solar wind-magnetosphere coupling is proportional to a multiplicative combination of solar wind velocity, density, and interplanetary magnetic field (e.g., McPherron et al., 1995; Burlaga & Lazarus, 2000; Veselovsky et al., 2010; Yurchyshyn et al., 2005). And, third, within-storm evolutions of geomagnetic disturbance can be viewed as an autoregressive filter of solar wind forcing and past states of the magnetospheric-ionospheric system (e.g., Chandorkar et al., 2017; Vassiliadis et al., 1999). Considering these three processes together, then, we can plausibly invoke a lognormal model as a description of the many solar-terrestrial processes that give storm maximum \( -D_{st,m} \) data. And, furthermore, we suggest that this lognormal model is most appropriate for the maximum intensities of the most intense storms that are both far-separated in time from previous and subsequent intense storms (and so relatively uncorrelated and causally unrelated), and, also, far removed from nonstatistical quiet time space weather conditions (which is highly autocorrelated).

As a further refinement, we can also consider the likely possibility that magnetic storm intensity has an effective upper bound. Vasyliunas (2011) argues that ring current intensity is limited by a balance between plasma pressure within the magnetosphere and the pressure of the magnetic dipole. At the point of such a balance, he calculates that the greatest possible value for \( -D_{st,m} \) is approximately 2,500 nT. In this context, we might hypothesize that magnetic storm intensities are realized from an upper limit lognormal process. Here, the variable \( \ln [x/(v - x)] \) is normally distributed, and \( x \) is positive and upper limit lognormally distributed, with density function

\[
f_{U,\mathcal{L}}(x|\mu, \sigma, v) = \frac{1}{\sqrt{2\pi\sigma^2} x(v-x)} \exp \left[ -\frac{(\ln [x/(v-x)] - \mu)^2}{2\sigma^2} \right], \quad \text{where } x \in (0, v);
\]

(5)

for which the corresponding complementary cumulative distribution is

\[
P_{U,\mathcal{L}}(x|\mu, \sigma, v) = \int_x^v f_{U,\mathcal{L}}(\theta|\mu, \sigma, v) \, d\theta = \frac{1}{2} \operatorname{erf} \left( \frac{\ln [x/(v-x)] - \mu}{\sqrt{2}\sigma} \right).
\]

(6)
(Bezdek & Solomon, 1983; Mugele & Evans, 1951). We note that Vasyl’iūnas’s upper limit on $-DSt_m$ is far higher than that which occurred for the March 1989 storm, far higher than that realized for the storm of May 1921 $-DSt_m = 907$ nT (Love et al., 2019b), and far higher than for the Carrington event of 1859, approximately 850 nT (Siscoe et al., 2006) to 1,050 nT (Gonzalez et al., 2011).

In describing the statistics of extreme-value data, formal asymptotic theories can be invoked. These theories apply for data that are envisioned as having been taken, as conditional samples, from the large-value tail of some hypothetical “parent” distribution, far away from the core part of the distribution where central limit concepts usually hold. In particular, consider sequences of $n$ independent data samples, each drawn from a “parent” distribution $P_0(y)$, which can be any one from a wide class of distributions. Denote one such sequence as $Y_n = \{y_1, y_2, \ldots, y_n\}$ and its (conditional) maximum value as $X_n = \{\max(Y_n)\}$. Under the Fisher-Tippett-Gnedenko theorem (e.g., Albeverio & Piterbarg, 2006; Gomes & Guillou, 2015), after normalizing for location and scale and taking the limit $n \to \infty$, the realization in probability of these “block maxima” is given by the generalized extreme-value distribution, for which the complementary cumulative is

$$P_{GEV}(x|\mu, \sigma, \xi) = \begin{cases} 1 - \exp\left[-\left(1 + \frac{(x-\mu)}{\sigma}\right)^{-\frac{1}{\xi}}\right] & \text{for } \xi \neq 0 \\ 1 - \exp\left[-\exp\left(\frac{x-\mu}{\sigma}\right)\right] & \text{for } \xi = 0. \end{cases}$$

(e.g., Bali, 2003; Walshaw). Depending on the “shape parameter,” $\xi$, the generalized extreme-value distribution reduces to one of three standard and widely used distributions,

- Fréchet: $P_F$ over $x \in (\mu - \sigma / \xi, +\infty)$ for $\xi > 0$.
- Gumbel: $P_G$ over $x \in (-\infty, +\infty)$ for $\xi = 0$.
- Weibull: $P_W$ over $x \in (-\infty, \mu - \sigma / \xi)$ for $\xi < 0$.

A particular parent distribution, $P_0$, with block maxima that are approximately distributed by one of these three extreme-value distributions (Fréchet, Gumbel, and Weibull), is said to be in the “domain of attraction” of that particular extreme-value distribution (e.g., Albeverio & Piterbarg, 2006; Gomes & Guillou, 2015). In this respect, the lognormal distribution is in the domain of attraction of the Gumbel distribution (e.g., Embrechts et al., 1997, Example 3.3.31; Walshaw, 2013).

Alternatively, consider the maxima of sequences of data samples $Y_n$ drawn independently from a parent distribution $P_0(y)$, suitably normalized and conditional on the maximum of each sequence exceeding some prechosen threshold $\gamma$, denoted as $X_n = \{\max(Y_n) > \gamma\}$. Under the Pickands-Balkema-de Haan theorem, after taking the limits $n \to \infty$ and $\gamma \to \infty$, the realization in probability of these “peak-over-threshold maxima” is given by a generalized Pareto distribution (e.g., Arnold, 2008), for which the complementary cumulative is

$$P_{GP}(x|\mu, \sigma, \xi) = \begin{cases} \left[1 + \frac{(x-\mu)}{\sigma}\right]^{-\frac{1}{\xi}} & \text{for } \xi \neq 0 \\ \exp\left[-\frac{x-\mu}{\sigma}\right] & \text{for } \xi = 0. \end{cases}$$

Depending on the parameters, the generalized Pareto distribution reduces to one of three standard and widely used distributions,

- power law: $P_{\text{PL}} = \left(\frac{x}{\mu}\right)^{-\xi}$ over $x \in [\mu, +\infty)$ for $\xi > 0$ and $\sigma = \mu \xi$.
- exponential: $P_{\text{Exp}}$ over $x \in [\mu, +\infty)$ for $\xi = 0$.
- beta: $P_{\text{Beta}}$ over $x \in [\mu, \mu - \sigma / \xi)$ for $\xi < 0$.

where the power law (e.g., Newman, 2005), often the preferred distribution in extreme-value theory, is a special case of $P_{GP}$ for $\xi > 0$ (e.g., Arnold, 2008, p. 122; Cirillo, 2013, p. 5948). In particular, physical systems that can be described in terms of self-organizing criticality often exhibit event occurrences that can be described in terms of power law or “fractal” statistics (e.g., Aschwanden et al., 2016; Turcotte, 1999). A power law distribution can also result from an autoregressive process perturbed by multiplicative noise (e.g., Sornette) and from a lognormal process perturbed by additive noise (e.g., Montebruno et al., 2019). In the context of space weather, the occurrence of individual magnetic substorms, such as caused by abrupt collapses of the magnetotail, has sometimes been described in terms of self-organizing criticality dynamics (Angelopoulos et al.,
It is important to recognize that extreme-value theory rests on the assumption that the extreme-value data on hand are actual (conditional) statistical samples of a parent distribution $F_\circ$. Although the sampling methods assumed under extreme-value theory certainly inspire the designs of our sampling methods described in section 3, our fundamental motivation for performing a sampling of the $Dst$ time series is removal of serial correlation, especially the intrastorm $Dst$ values recording the continuous-time evolution within each storm. And since the $Dst$ time series is not statistical, the $-Dst_m$ values we obtain from the $Dst$ do not, in concept, closely resemble a conditional statistical sample of an underlying parent distribution $F_\circ$, such as assumed under the Pickands-Balkema-de Haan theorem. That we have hypothesized that storm maximum intensities, as measured by $-Dst_m$, might be viewed as the realization of a lognormal process does not amount to an equating of $-Dst_m$ with $F_\circ$. On the other hand, the block-maximum $-DstBM_m$ values, being the statistical maxima of many $-Dst_m$ data values realized across the solar cycles, do closely resemble independent conditional statistical samples from a parent distribution $F_\circ$, as assumed under the Fisher-Tippett-Gnedenko theorem. In this case, we might reasonably view the $-DstBM_m$ as samples from the lognormal process we hypothesize as giving rise to the $-Dst_m$. Having said all of this, we recognize that all statistical models are idealizations. Therefore, we suggest that it is useful to compare and contrast the capacities for the generalized extreme-value and lognormal models for describing the $-Dst_m$ data.

5. Asymptotic Properties of the Models

In examining the basic theoretical properties of the candidate model distributions, we consider, first, those supported on the infinite domain with right endpoint $x \to \infty$. Easily seen is the effect of a horizontal shift of the exponential distribution ($\xi = 0$) by an amount $h$,

$$F_{\xi}(x + h) = \exp \left[ -\frac{h}{\sigma} \right], \quad (9)$$

which results in an exponential scaling of the distribution. In the limit of large $x$, the Gumbel distribution has a similar property,

$$\lim_{x \to \infty} \frac{F_{\xi}(x + h)}{F_{\xi}(x)} = \exp \left[ -\frac{h}{\sigma} \right]. \quad (10)$$

But for the lognormal distribution,

$$\lim_{x \to \infty} \frac{F_{LN}(x + h)}{F_{L}(x)} = 1. \quad (11)$$

This means that with increasing $x$, the tail of a lognormal progresses toward 0 more slowly than an exponential distribution—the lognormal is "long-tailed." Generally speaking, with only pathological (nonsmooth) exceptions, a long-tailed distribution is also "heavy" and "subexponential" (e.g., Foss et al., 2013, Chapters 1 and 2; Zwart, 2009, Chapter 2.3), and, indeed, the lognormal is subexponential (e.g., Foss et al., 2013, Chapter 3.4). Usually, these properties are associated with the Fréchet and power law distributions ($\xi > 0$); we note that

$$\frac{F_{PL}(x + h)}{F_{PL}(x)} = 1 = \lim_{x \to \infty} \frac{F_{PL}(x + h)}{F_{PL}(x)} \quad (12)$$

(e.g., Coles, 2001, Example 4.2). The subexponential property of the lognormal distribution holds even though it is within (outside) the domains of attraction of the exponential (power law) and Gumbel (Fréchet) distributions (Goldie & Resnick, 1988, Section 3).

Consider, next, a rescaling of the power law and Fréchet distributions ($\xi > 0$) by a positive factor $r$,

$$\frac{F_{PL}(rx)}{F_{PL}(x)} = r^{-\frac{1}{\xi}} = \lim_{x \to \infty} \frac{F_{PL}(rx)}{F_{PL}(x)}, \quad \text{for } r \in [0, +\infty). \quad (13)$$
This means that the power law distribution and the extreme-value end of the Fréchet distribution are “regularly varying” (e.g., Beirlant et al., 2004, Chapter 2.9.2; Bingham et al., 1987; Zwart, 2009, Chapter 2.1), “scale-invariant,” or “self-similar” (e.g., Corral & González, 2019; Ghil et al., 2011; Sornette, 2006)—that is, a scaling of the independent variable \( x \) gives a corresponding power scaling of the distribution. In contrast, the lognormal distribution does not have this self-similar property,

\[
\lim_{x \to \infty} \frac{F_{\lognormal}(rx)}{F_{\lognormal}(x)} = \begin{cases} 
0 & r \in (1, +\infty) \\
1 & r = 1 \\
\infty & r \in (-\infty, 1)
\end{cases} \quad (14)
\]

This means that, with increasing \( x \), the tail of the lognormal progresses toward 0 faster than a power law (e.g., Malevergne et al., 2011, equation 3). Regular variation at infinity is a necessary and sufficient condition for a parent distribution to be in the Fréchet domain of attraction (e.g., Albeverio & Piterbarg, 2006; Chapter 3.2.1). The lognormal (14) is not regularly varying at infinity, and, indeed, its tail behavior is like that for the exponential and Gumbel distributions,

\[
\frac{F_{\text{ex}}(rx)}{F_{\text{ex}}(x)} = \begin{cases} 
0 & r \in (1, +\infty) \\
1 & r = 1 \\
\infty & r \in (-\infty, 1)
\end{cases} = \lim_{x \to \infty} \frac{F_{\text{gumbel}}(rx)}{F_{\text{gumbel}}(x)}. \quad (15)
\]

From the preceding simple examination of shifting and rescaling properties, we understand that the lognormal distribution is rather special—although it resides in the exponential-Gumbel domain of attraction, it has properties that straddle boundaries between the exponential-Gumbel and the power law-Fréchet families of distributions.

And, finally, consider the model distributions defined over finite domains (shape parameter \( \xi < 0 \)). As \( x \) increases toward the right endpoint \( x_m = \mu - \sigma/\xi \) of the beta and Weibull distributions,

\[
\lim_{x \to \infty} \frac{F_{\beta}(x_m - r/x)}{F_{\beta}(x_m - 1/x)} = r^{-1} = \lim_{x \to \infty} \frac{F_{\text{weibull}}(x_m - r/x)}{F_{\text{weibull}}(x_m - 1/x)}, \quad \text{for } r \in [0, +\infty) \quad (16)
\]

(e.g., Hashorva et al., 2010, equation 2.6). These are symmetric with the corresponding properties of the power law and Fréchet distributions (13); indeed, equation (16) defines regular variation near the right endpoint (e.g., Albeverio & Piterbarg, 2006, Chapter 3.2.1). In contrast, for the upper limit lognormal, as \( x \) increases toward its right endpoint \( \alpha \),

\[
\lim_{x \to \infty} \frac{F_{\lognormal}(x_m - r/x)}{(\alpha - r/x)F_{\lognormal}(\alpha - 1/x)} = \begin{cases} 
\infty & r \in (1, +\infty) \\
1 & r = 1 \\
0 & r \in (-\infty, 1)
\end{cases} \quad (17)
\]

which is antisymmetric with the corresponding properties of the lognormal (14) and exponential and Gumbel (15) distributions. As the right-hand endpoint is approached, the upper limit lognormal decreases more rapidly than either of the beta or Weibull distributions. Interestingly, even though the upper limit lognormal has, like the beta and Weibull, a finite right endpoint, it is not in their domains of attraction because it is not regularly varying (e.g., Albeverio & Piterbarg, 2006, Chapter 3.2.1).

The preceding observations demonstrate both asymptotic similarities and differences between the lognormal distributions and the formal extreme-value distributions. Perhaps most importantly, the lognormal (upper limit lognormal) distribution is seen to be heavy, even though its far tail converges toward 0 more rapidly than do the far tails of the Fréchet and power law (beta and Weibull) distributions. Practically speaking, this might potentially affect extrapolated estimates of occurrence rates—if the assumed distributional model is heavier (lighter) than that for the actual underlying process, then the far-tail, extreme-event occurrence rate might be overestimated (underestimated). The degree to which occurrence rate estimates are erroneous, and whether or not they are over- or underestimates, depends on how far out in the tail one extrapolates and the degree to which the model distribution approximates the actual process. We suggest that it is useful to compare and contrast extrapolations obtained from generalized extreme-value distributional models with those obtained from the lognormal models.
6. Reconnaissance Inspection of the Data

Before proceeding with our detailed statistical analysis, it is worth making a reconnaissance inspection of the $-Dstm$ data to help identify a $-Dst_c$ threshold value, above which storm maximum intensities might be realized from the tail of a hypothetical statistical distribution. Within the wider extreme-value analysis community, it is standard to inspect plots of the mean (expected value) of the excess as a function of threshold (e.g., Coles, 2001; Scarrott & MacDonald, 2012). For the generalized Pareto distribution, the mean excess is linear in the threshold $\gamma$,

$$E_{GP}(x - \gamma | x > \gamma) = \sigma + \xi \gamma \frac{1}{\xi - \mu}, \quad \text{where } \gamma \in [0, \infty) \quad \text{for } \xi < 1$$

(e.g., Davison & Smith, 1990, Section 2; Coles, 2001, equation 4.9).

Recognizing this, in Figure 3, we plot the empirical mean excess $E$ of the $-Dstm$ data as a function of $-Dst_c$ threshold. Note that for thresholds of $-Dstm \geq -Dst_c = 250$ nT ($N \leq 38$ storms), the excess is approximately linear, possibly indicating that storms of such intensities are realized from the distributional tail where generalized extreme-value approximations might apply (Tsubouchi & Omura, 2007, Figure 3). Conversely, storms with $-Dstm \leq -Dst_c = 250$ nT might not be realized from an extreme-value distributional tail. Here, we note that a $-Dst_c \approx 250$ nT threshold is approximately the level said to define “superstorms” or “superintense” storms that are, almost exclusively, driven by interplanetary coronal mass ejections; less intense storms, those for which $-Dstm \lesssim 250$ nT, can be driven by either interplanetary coronal mass ejections or corotating interactive regions at the front of high-speed streams of solar wind (e.g., Gonzalez et al., 2011; Kilpua et al., 2017; Richardson & Cane, 2012). Recognizing that superstorms are essentially mono-causal, it is, perhaps, not surprising that the excess over the threshold $-Dst_c \geq 250$ nT apparently exhibits some asymptotic linearity.

But we should be careful to avoid overinterpretation of Figure 3. For the lognormal distribution, we note that, for large $\sigma$,

$$E_{LN}(x - \gamma | x > \gamma) \approx \frac{\sigma^2 \gamma}{\ln(\gamma) - \mu}, \quad \text{where } \gamma \in [0, \infty)$$

(Cirillo, 2013, equation 15; Ghosh & Resnick, 2010, Section 5.2.1). Since $\ln(\gamma)$ changes slowly compared to $\gamma$, equation (19) is, over a finite domain, approximately linear in $\gamma$ (e.g., Mitzenmacher, 2004; Perline, 2005). From this, we understand that, although linearity in mean excess might be indicative of data realized from
the far-tail, asymptotic regime of a hypothetical distribution, this is not necessarily a Pareto—the data might be realized from some other distribution—even, possibly, a lognormal.

7. Numerical Methods

For detailed analysis, we fit the candidate statistical models to the \(-D_{stm}\) data by maximizing the likelihood (e.g., Bohm & Zech, 2010, Chapter 6.5)

\[
L(A, \mu, \sigma, \cdots) = \prod_j g(-D_{stm}|A, \mu, \sigma, \cdots),
\]

where the product in \(j\) is taken over all the \(-D_{stm}\) data, where

\[
g(-D_{stm}|A, \mu, \sigma, \cdots) = A \cdot f(-D_{stm}|\mu, \sigma, \cdots),
\]

and where \(A\) is a normalizing factor,

\[
A \int_{-D_{stc}}^{\infty} f(\theta|\mu, \sigma, \cdots)d\theta = \frac{N(-D_{stm} \geq -D_{stc})}{60 \text{ years}}.
\]

We use a simplex algorithm (e.g., Press et al., 1992, Chapter 10.4) to obtain maximization and to estimate the parameters \(\{\mu, \sigma, \cdots\}\). Once this is done, the cumulative probability for a magnetic storm with maximum \(-D_{stm}\) exceeding some level \(x \geq -D_{stc}\) is the cumulative,

\[
F(x|\mu, \sigma, \cdots) = A \cdot \int_{x}^{\infty} f(\theta|\mu, \sigma, \cdots)d\theta.
\]

We use a bootstrap method (e.g., Efron & Tibshirani, 1993), resampling the data (with replacement) and refitting each sampled data set with models, to estimate the statistical errors associated with extrapolated storm occurrence rates (e.g., Love et al., 2015; Riley & Love, 2017).

8. Peak-Over-Threshold Fits

We begin our summary of detailed results with a presentation of fits of the (modified) peak-over-threshold \(-D_{stm}\) data provided by the lognormal model, equation (3), the upper limit lognormal model, equation (5), and the generalized Pareto model, equation (8), where for the latter two cases, we “bound” the maximum allowable \(-D_{stm}\) according to the suggestion of Vasyliunas (2011) by fixing, respectively, \(v = 2, 500 \text{ nT}\) and \(\mu = \sigma/\xi = 2, 500 \text{ nT} (\xi < 0)\). In Figure 4, we show the cumulatives of the \(-D_{stm} \geq -D_{stc}\) data and the cumulatives of the candidate models, each for a variety of different \(-D_{stc}\) thresholds; fits with low (high) \(-D_{stc}\) are to data with more (fewer) data. Qualitatively, the fitted model functions provide good representations of the \(-D_{stm}\) data, especially for (say) the rare occurrences of magnetic superstorms with \(-D_{stc} \gtrsim 283 \text{ nT}\) (\(N \lesssim 28\) storms). On the other hand, for \(-D_{stc} \lesssim 141 \text{ nT} (N \gtrsim 168), many more storms covering a much greater range of \(-D_{stm}\) are fitted, but in these cases, the results are mixed; the lognormal and bounded upper limit lognormal provide reasonably good fits of the data, but the bounded generalized Pareto does not fit so well. This might be an indication that lower intensity \(-D_{stm}\) values are not realized from an extreme-value distribution (as already suggested in section 6).

9. Statistical Significance

A measure of goodness of fit of the statistical models is provided by the Kolmogorov-Smirnov test (e.g., Bohm & Zech, 2010, Chapter 10.3.5), from which we obtain an estimate of the \(p\) value probability for the realization of a discrepancy in probability between the cumulative of a hypothetical model and the data that are as large as or larger than those actually observed (e.g., Bohm & Zech, 2010, Chapter 10.2.4). Good (poor) fits have large (small) probabilities. For a lognormal model fitted to the data for a superstorm threshold of \(-D_{stc} = 283 \text{ nT}, Figure 4f, the p value is 0.97 (a high probability). But this result, like those we have previously reported for similar analyses (e.g., Love et al., 2015), is based on data that were used to estimate the model parameters—it is circular logic (we now recognize) to calculate a \(p\) value using the same data used to estimate a model’s parameters (e.g., Chave, 2017, Chapter 8.2.4; Clauset et al., 2009, Section 3.4; Corral & González, 2019, Section 3.2; Steinskog et al., 2007).
Figure 4. Cumulatives (a–h) for peak-over-threshold $-D_{st_{m}}$ data (black) and fitted lognormal (blue), bounded generalized Pareto (green), and bounded upper limit lognormal (red) models for different $-D_{st_{c}}$ thresholds. Also given are the number $N$ of $-D_{st_{m}}$ data such that $-D_{st_{m}} \geq -D_{st_{c}}$, the natural logarithm of the likelihood $\ln L$, the Kolmogorov-Smirnov $p$ value, and the 100- and 1,000-year extrapolated storm intensities.

We can circumvent this circular bias by calculating multiple $p$ values from multiple bootstrap resamplings of the data. For the lognormal and $-D_{st_{c}} = 283$ nT, the median of the resampled $p$ values is 0.24. This objective measure of fit is still too high to motivate outright rejection of the lognormal model. Median values for resampled $p$ values for each model and various $-D_{st_{c}}$ are given in Figure 4. Generally, the probabilities are not small enough to motivate rejection of any of the lognormal, upper limit lognormal, and generalized...
Pareto models as descriptions of the peak-over-threshold $-D_{st_{m}}$ data for $-D_{st} \geq 283$ nT. We might interpret this as indicating that high intensity $-D_{st_{m}}$ are realized from a distributional tail where generalized extreme-value approximations might apply (also as suggested in section 6).

10. Extreme-Event Extrapolations

Of particular interest in projects for mitigating the deleterious effects of intense space weather events (e.g., Bernabeu, 2013; Cannon et al., 2013; National Science and Technology Council, 2085; North American Electric Reliability Corporation, 2016) is a 100-year magnetic storm benchmark (e.g., Love et al., 2015, 2016b, 2018; Pulkkinen et al., 2012; Riley & Love, 2017). Occasionally, interest is also expressed in a 1,000-year storm benchmark (e.g., Gopalswamy; Oughton et al., 2018). Therefore, in Figure 4, we show model extrapolations to both 100- and 1,000-year storm intensities. For a superstorm threshold of $-D_{st} = 283$ nT ($N = 28$), Figure 4f, we note that extrapolations to 100-year storm intensities range from 520 nT (generalized Pareto) to 545 nT (lognormal), a difference of 25 nT. On the other hand, for a much higher threshold of $-D_{st} = 400$ nT ($N = 5$), Figure 4h, extrapolations to 100-year storm intensities range from 524 nT (lognormal, generalized Pareto) to 527 nT (upper limit lognormal), a difference of just 3 nT. For the lognormal model, the difference between the 1,000-year extrapolation for $-D_{st} = 283$ nT and that for $-D_{st} = 400$ nT is 21 nT.

Interest in 1,000-year space weather events needs to be contained by extreme caution. One might, for example, imagine that 1,000-year storms are qualitatively different from 100-year storms, or, at least, one might imagine that they are different from those that have occurred since the start of $D_{st}$ in 1957. From Figure 4, for a threshold of $-D_{st} = 283$ nT, extrapolations to 1,000-year storm intensities range from 650 nT (generalized Pareto) to 744 nT (lognormal), a difference of 94 nT. For a much higher threshold of $-D_{st} = 400$ nT, extrapolations to 1,000-year storm intensities range from 651 nT (generalized Pareto) to 696 nT (lognormal), a difference of just 45 nT. For the lognormal model, the difference between the 1,000-year extrapolation for $-D_{st} = 283$ nT and that for $-D_{st} = 400$ nT is 48 nT. We conclude that extrapolations to 1,000-year storm intensities are insensitive to model selection. It is noteworthy that all of the extrapolations to 1,000-year extrapolated intensities are far below the upper limit of 2,500 nT suggested by Vasylıûnas (2011).

11. Superstorm Threshold

In Figure 5, we show the 100- and 1,000-year extrapolated storm intensities for the candidate models, together with 68% (one standard deviation) confidence intervals obtained by bootstrap resampling, refitting, and reextrapolation, each as a function of $-D_{st}$ threshold. Consistent with what might have been expected from section 6, the extrapolated intensities for the generalized Pareto model appear to converge for a superstorm threshold of $-D_{st} \geq 283$ nT ($N \leq 28$), and, notably, so do the extrapolated intensities for the lognormal and upper limit lognormal models. The threshold of $-D_{st} = 283$ nT appears to be the (approximate) critical point above which storm intensities are realized from the extreme-value distributional tail that can be well approximated by any of the candidate models. For the upper limit lognormal, and for $-D_{st} = 283$ nT, the 100-year intensity [and 68% confidence interval] are 542 [466, 583] nT; the corresponding 1,000-year intensity and confidence interval are 715 [552, 797] nT. The statistical uncertainty of these estimates is considerable—the 100- and 1,000-year confidence intervals overlap.

We note that the March 1989 storm, with $-D_{st_{m}} = 589$ nT and the most intense storm recorded in the $D_{st}$ time series, was more intense than the upper limit lognormal 100-year extrapolation above the superstorm threshold. In this respect, the March 1989 storm might be considered to be more than a 100-year storm (e.g., Love et al., 2018), possibly one that just happened to have occurred in the past 60-year time span of the $D_{st}$ time series. Indeed, for an upper limit lognormal model with $-D_{st} = 283$ nT, the March 1989 storm had an intensity that is only expected to be exceeded (on average) once per 189 years. So, in that sense, the March 1989 storm was remarkable (see section 14).

Interestingly, the second most intense storm in the $D_{st}$ time series, that of July 1959 with $-D_{st_{m}} = 429$ nT, falls below the 100-year confidence interval. Furthermore, we recognize that the 100-year storm intensity estimated here is lower than the 645 nT value given by Tsubouchi and Omura (2007, Section 5.2) and much lower than the 880 nT given by Love et al. (2015), where in contrast to the method used here, the latter study used a low threshold (far below superstorm, $-D_{st_{m}} = 63$ nT) so as to include as many storms as appeared to be consistent with a lognormal model. We note, however, that our 100-year extrapolated intensity of 542 nT falls within the (very wide) 95% confidence interval given by Love et al. (2015).
Figure 5. Peak-over-threshold extrapolated (a) 100- and (b) 1,000-year storm intensities for the lognormal (blue), bounded generalized Pareto (green), and bounded upper limit lognormal (red) models, each for a range of $-Dst_c$ thresholds. Also shown are 68% (one standard deviation) confidence intervals (dotted lines) obtained by bootstrap resampling, refitting, and reextrapolation.

12. Statistically Elusive Upper Limit

Because it is physical, we have, until now, fixed the $\nu$ parameter in the upper limit lognormal model (5) and the $\xi$ parameter in the generalized Pareto model (8) by imposing the 2,500 nT upper bound on the intensity of a magnetic storm given by Vasyliūnas (2011). If, instead, $\xi$ and $\nu$ are allowed to assume their optimal value as free parameters in maximum-likelihood fits of the $-Dstm$ data, one might hope to infer a theoretical upper limit consistent from the statistics of the $-Dstm$ data. To investigate this, we fitted generalized Pareto and upper limit lognormal models to the $-Dstm$ with both $\xi$ and $\nu$ free and for a range of $-Dst_c$ thresholds. In Figure 6a, we see that, for $-Dst_c \gtrsim 218$ nT, $\xi$ changes from positive to negative depending on the chosen threshold—this means that the data are insufficient, on purely statistical grounds, for determining whether or not the far tail of $-Dstm$ is heavy (and unbounded) or light (and bounded). Relatedly, in Figure 6b, we see that, for $-Dst_c \gtrsim 218$ nT, the $-Dstm$ data are, again, on purely statistical grounds, insufficient for inferring a usefully accurate theoretical upper limit on storm intensity. From this, we understand that fixing an upper limit on storm intensity provides a small amount of stability in maximum-likelihood estimation of model parameters. We doubt the estimate drawn by Acero et al. (2018), based only on qualitative grounds, that magnetic storm intensity has an upper limit of $-Dstm \approx 850$ nT. The data do not support such an estimate.
Figure 6. Shape parameter $\xi$ (a) for the generalized Pareto model and (b) the estimated upper limit $\nu$ for the upper limit lognormal (red) and the estimated upper limit $\mu - \sigma/\xi$ for the generalized Pareto (green) models, each for peak-over-threshold fits of $-\text{Dst}_m$ data for a range of $-\text{Dst}_c$ thresholds. For estimated $\xi \geq 0$, the generalized Pareto does not have a finite upper limit, and its tail is said to be “heavy.”

If a most extreme storm intensity estimate is to be made, it needs to be quoted with confidence intervals, but from what we have seen here, those intervals are likely to be very wide.

13. Model Selection

Likelihood ratio is a standard model selection criterion (e.g., Bohm & Zech, 2010, Chapter 10.3.4). Among candidate models, the “best” one might be taken as that with the highest relative likelihood. For a lognormal model fitted to the data for a superstorm threshold of $-\text{Dst}_c = 283$ nT, Figure 4f, the upper limit lognormal has the highest likelihood, $\ln L = 5.1652$. But for a slightly lower (higher) threshold, such as $-\text{Dst}_c = 238$ (336) nT, Figure 4e (4g), the generalized Pareto has the highest likelihood, $\ln L = 5.4070$ ($\ln L = 5.3081$). In terms of the $p$ values, for $-\text{Dst}_c = 238, 283,$ and 336 nT, the upper limit lognormal has the highest probability, though the $p$ values for the other models are at least comparable to those for the upper limit lognormal. Insofar as all of the candidate models provide good representations of the $-\text{Dst}_m$ data (above superstorm threshold), it is, perhaps, not surprising that comparisons of likelihoods and $p$ values do not provide a clear indication as to which candidate model is best. We prefer the upper limit lognormal model because it makes the most physical sense to us. We find it acceptable because it provides a good representation of the $-\text{Dst}_m$ data, but other models cannot be easily rejected.
Figure 7. Cumulatives and bounded upper limit lognormal fits of peak-over-threshold $-\text{Dst}_c = 283$ nT data, (a) for a leave-one-out stability test, with (red) and without (black) the $-\text{Dst}_m = 589$ nT value of the March 1989 storm, (b) for a bisection stability test, with the first half (brown, 1957–1986) and the second half (black, 1987–2016) of the $\text{Dst}$ time series. Also given are the number $N$ of $-\text{Dst}_m$ data used in each fit, the natural logarithm of the likelihood $\ln \text{L}$, the Kolmogorov-Smirnov $p$ value, and the 100- and 1,000-year extrapolated storm intensities.

14. Stability and Secular Change

Recalling from section 10 that extrapolations to 1,000-year storm intensities were stable under model selection, we now consider stability under data selection. Recognizing that the March 1989 storm was of unusually great intensity, in Figure 7a, we show fits for the upper limit lognormal model with $-\text{Dst}_c = 283$ nT and extrapolations with and without the $-\text{Dst}_m = 589$ nT value of the March 1989 storm. The leave-one-out extrapolations are lower than for those in which the intensity of the March 1989 storm is included. For the upper limit lognormal, the leave-one-out 100-year extrapolation [and 68% confidence interval] are 473 [447, 491] nT; the corresponding 1,000-year intensity and confidence interval are 567 [521, 599] nT. We note that this leave-one-out 100-year confidence interval is not wide enough to accommodate the intensity of the March 1989 storm, but that of the 1,000-year interval is. Although this leave-one-out experiment is artificial, it shows that extrapolations to 100- and 1,000-year storm intensities depend very much on the most extreme storm intensities in the $\text{Dst}$ time series.

Next, we show fits and extrapolations under a bisection test, separately using samplings from the first half (1957–1986) and the second half (1987–2016) of the historical $\text{Dst}$ time series. We show results for the upper limit lognormal model with $-\text{Dst}_c = 283$ nT in Figure 7b, where we also show results for sampling the entire $\text{Dst}$ time series (1957–2016). The upper limit lognormal 100-year extrapolation [and 68% confidence interval], for the first half of the $\text{Dst}$ time series, are 479 [425, 509] nT; the corresponding 1,000-year intensity and confidence interval are 571 [489, 619] nT. The 100-year extrapolation and confidence interval for the second half of the $\text{Dst}$ time series are 600 [467, 640] nT; the corresponding 1,000-year intensity and confidence interval are 864 [557, 876] nT. That the second-half extrapolations are clearly higher than the first-half extrapolations might indicate a slow, secular increase in magnetic storm intensity (related to a secular increase in solar activity) over the past three solar cycles. Noting that the first-half 100- and 1,000-year extrapolations fall within the second-half confidence intervals, but the second-half 100- and 1,000-year extrapolations fall outside of the first-half confidence intervals, an increase in storm intensity might be seen as being statistically significant.

These observations are consistent with the assessment made above that the 1989 storm was of unusually high intensity. It is also consistent with other time series related studies (e.g., Clilverd et al., 1998; Love, 2011) and physics-based studies of secular change in space climatology (e.g., Stamper et al., 1999). The recent increased storm intensity might also represent a reversion to a long-term mean, since, prior to the start of the $\text{Dst}$ time series in 1957, a number of magnetic storms had very high $-\text{Dst}_m$. In particular, the Carrington event of September 1859 had (1-hr average) $-\text{Dst}_m \approx 850$ nT (Siscoe et al., 2006) to approximately 1,050 nT (Gonzalez et al., 2011), the storm of May 1921 had $-\text{Dst}_m = 907$ nT (Love et al., 2019b), and the storm of September 1909 had $-\text{Dst}_m = 595$ nT (Love et al., 2019a). Each of these, along with the storm of March 1989, had intensities that exceed our 100-year intensity estimate of 542 [466, 583] nT for the upper limit lognormal model.
15. Effects of Conditional Sampling

Looking back to sections 3 and 4, we consider the effects of sampling the Dst time series. In Figure 8, we show cumulatives of the $-Dst_c$ data obtained by our modified peak-over-threshold sampling method for $-Dst_c = 119$ nT and $-Dst_c = 283$ nT, each with a quiet time level of $\beta = 30$ nT. We recall that our modified sampling method excludes most multiple exceedances of $-Dst_c$ that occur during some storms. In Figure 8, we also show cumulatives for a compilation of $-Dst_c$ data obtained with a more traditional peak-over-threshold sampling method, such as envisioned under the Pickands-Balkema-de Haan theorem; for these, $-Dst_c = \beta = 119$ nT and $-Dst_c = \beta = 283$ nT (quiet time level the same as the exceedance threshold). The traditional sampling method admits multiple crossings of $-Dst_c$ for some storms. For comparison, in Figure 8, we show the cumulative of all the $-Dst$ data (these are autocorrelated). We see, immediately, that both our peak-over-threshold sampling method and the traditional sampling method result in distributions for $-Dst_c$ data that are very different from that for the autocorrelated $-Dst$ data. Notably, for $-Dst_c = 119$ nT, Figure 8a, our modified (the traditional) peak-over-threshold sampling method reduces number data from $N = 3,193$ to 263 (422), and for $-Dst_c = 283$ nT, Figure 8b, the reduction is from 135 to 28 (34). More interesting, however, are comparisons of the distributions for the different peak-over-threshold sampling methods. Our modified method results in fewer small $-Dst_c$ values, those marginally above $-Dst_c$, than does the traditional sampling method—this is due to redundant counting of some storms under the traditional sampling method. These differences in distribution affect the model fits. The data selected by our modified peak-over-threshold method are better fitted by the bounded upper limit lognormal model, as measured by either $\ln L$ or $p$ value, than are the data selected by the traditional peak-over-threshold method.

16. Block-Maximum Fits

Next, in Figure 9a, we show fits of solar cycle, block-maximum $-Dst_{BM}$ data by the lognormal model, equation (3), the bounded upper limit lognormal model, equation (5), and the bounded generalized extreme-value models, equation (7), where for the latter two cases, and in parallel with the results presented in section 8, the maximum allowable $-Dst_{BM}$ is fixed, $\nu = 2$, 500 nT and $\mu - \sigma/\xi = 2$, 500 nT ($\xi < 0$). The cumulatives of the $-Dst_{BM}$ data and the cumulatives of the candidate statistical models show that all of the model functions provide good and almost identical representations of the solar cycle, block-maximum $-Dst_{BM}$ data.

Regarding statistical significance, bootstrap resampling and refitting gives median Kolmogorov-Smirnov $p$ values ranging from 0.43 for the lognormal model to 0.35 for the upper limit lognormal model. These are not small enough to motivate, on purely statistical grounds, rejection of any of the candidate models as descriptions of the solar cycle, block-maximum $-Dst_{BM}$ data. Regarding goodness of fit, the generalized extreme-value model has the highest likelihood, $\ln L = 36.7157$, but this is only slightly higher than the model with the lowest likelihood, the upper limit lognormal model, $\ln L = 36.4650$. As was the case for modeling of
Figure 9. Cumulatives (black) for solar cycle, block-maximum $-\text{Dst}_{BM}$ data and fitted lognormal (blue), bounded generalized extreme-value (green), and bounded upper limit lognormal (red) models where (a) the $-\text{Dst}_{BM} = 589$ nT value of the March 1989 storm is used, and (b) under a replace-one stability test, the $-\text{Dst}_{BM} = 354$ nT value of the November 1991 storm is used instead of the value for the March 1989 storm.

In Figure 9a, we also show model extrapolations to 9 and 90 solar cycles, corresponding, roughly, to 100- and 1,000-year storm intensities if we assume an 11-year solar cycle period. The nine-solar cycle extrapolations range from 553 nT (upper limit lognormal) to 591 nT (generalized extremevalue), a difference of just 38 nT. The 90-solar cycle extrapolations range from 759 nT (upper limit lognormal) to 899 nT (generalized extremevalue), a difference of 140 nT. As with 1,000-year extrapolations of peak-over-threshold $-\text{Dstm}$ data, section 10, all of the extrapolations to 90-solar cycle intensities are far below the upper limit of 2,500 nT suggested by Vasyliunas (2011). Because only six-solar cycle, block-maximum $-\text{Dst}_{BM}$ data are available, it is not practical to use bootstrap resampling to estimate statistical confidence intervals.

Similar to the leave-one-out test performed in section 14, we perform a replace-one stability test of fits of the solar cycle, block-maximum $-\text{Dst}_{BM}$ data. Noting that the March 1989 storm, with $-\text{Dst}_{m} = 589$ nT, was the most intense storm of solar cycle 22, we replace its value with that of the second most intense storm of the cycle, the November 1991 storm, $-\text{Dst}_{m} = 354$ nT. We show fits in Figure 9b. All candidate models
provide good representations of the data. For corresponding model fits, extrapolations to 9- and 90-solar cycle intensities are lower for fits in which the November 1991 \(-\text{Dst}_m\) value is used than for those in which the March 1989 storm is used. As we concluded in section 14, extrapolations to extreme storm intensities depend, critically, on the most extreme storm intensities that have actually been recorded.

17. Small Irreconcilable Differences

We note that the 9- and 90-solar cycle, block-maximum extrapolations (553 and 759 nT) are slightly higher than the 100- and 1,000-year peak-over-threshold extrapolations, each obtained for an upper limit lognormal, 542 [466, 583] and 715 [552, 796] nT for \(-\text{Dst}_m = 283 \text{ nT}\). This is a systematic difference. With block-maximum sampling of solar cycles 19 through 24, we have the ranked \(-\text{Dst}_{m_{BM}}\) values of [589, 429, 422, 387, 325, 222 nT], but for peak-over-threshold sampling, the seven most intense storms have six ranked \(-\text{Dst}_{m}\) values of [589, 429, 427, 426, 422, 387 nT] (two storms have \(-\text{Dst}_{m} = 387 \text{ nT}\)). The distribution of the block-maximum intensities is flatter than the distribution of the peak-over-threshold intensities. Flatter distributions give more extreme extrapolations, but the differences in the extrapolations are, in this case, small. For the upper limit lognormal 100-year and nine-solar cycle extrapolations, the difference is just 11 nT. This is well within the 68% confidence interval obtained by bootstrap resampling of the peak-over-threshold samples.

18. Conclusions

The existing \(-\text{Dst}_m\) data (1957–2016) can be about as well described by lognormal models as by formal extreme-value models. The models permit extrapolation to 100-year storm intensities. Such extrapolations are useful, but, beyond that, we have found it challenging to infer other integrated qualities of extreme-event storm intensities—the data are about as well-described by models that have infinitely long tails as by those having finite right endpoints, by models with tails that are either regularly varying or not, and by models with tails that are either heavy or light (though, perhaps, not especially heavy and not especially light). That we have not been able to draw more specific inferences is reflective of the flexibility of the distributional models and the limited information content of the \(-\text{Dst}_m\) data.

This study is an advancement on previous work which we either performed (e.g., Love et al., 2015) or with which we have been associated (e.g., Riley & Love, 2017). Whereas those works were focused exclusively on lognormal and power law distributional models of the \(-\text{Dst}_m\) data, in this study, we additionally consider the upper limit lognormal and formal extreme-value models. Whereas those works were focused on peak-over-threshold data for single, specific thresholds intended to encompass as much data as possible, in this study, we explore a range of thresholds \(-\text{Dst}_n\), selecting a smaller data set that shows possible asymptotic consistency with extreme-value distributions, and we also consider block-maximum data selection. This focus on the most extreme superstorm intensities has led to a lower 100-year extrapolated estimate of 542 nT as compared to 880 nT given by Love et al. (2015). These very different estimates illustrate sensitivity to data and model selection.

Further progress in characterizing the statistical properties of the \(-\text{Dst}_m\) data, and especially in reduction in the uncertainty in extrapolated intensities, would benefit from theoretical refinements in specification of model distributions and their parameters. Although we advocate for use of the upper limit lognormal, tighter specification of the theoretical upper limit of storm intensity is needed if tight confidence intervals are required for extrapolations to extreme intensities. We appreciate that a view far across the past is necessary, though not sufficient, for making modest forecasts of the future (e.g., Saffo, 2007). We recognize that estimating pre-1957 storm intensities is a high priority. Noting that models of the post-1957 \(-\text{Dst}_m\) data yield 100-year extrapolated storm intensities that are eclipsed by several pre-1957 storms, we might reasonably infer that our extrapolations are underestimates.

References

Acero, F. J., Vaquero, J. M., & Gallego, M. C. (2018). A limit for the values of the Dst geomagnetic index. Geophysical Research Letters, 45, 9435–9440. https://doi.org/10.1029/2018GL079676
Aitchison, J., & Brown, J. A. C. (1957). The lognormal distribution: With special reference to its uses in economics, (pp. 1–176). Cambridge, UK: Cambridge University Press.
Albeverio, S., & Piterbarg, V. (2006). Mathematical methods and concepts for the analysis of extreme events. In S. Albeverio, & V. Jentsch (Eds.), Extreme Events in Nature and Society, The Frontiers Collection (pp. 47–68). Berlin, Germany: Springer.
Albeverio, S., Jentsch, V., & Kantz, H. (2006). Extreme events in nature and society. In Extreme Events in Nature and Society, The Frontiers Collection (pp. 1–352). Berlin, Germany: Springer.
Goldie, C., & Resnick, S. (1988). Distributions that are both subexponential and in the domain of attraction of an extreme-value distribution. *Advances in Applied Probability, 20*(4), 706–718. https://doi.org/10.2307/1427356

Gomes, I. M., & Guillou, A. (2015). Extreme value theory and statistics of univariate extremes: A review. *International Statistical Review, 83*(2), 263–292. https://doi.org/10.1111/1751-5823.12058

Gonzalez, W. D., Echer, E., Tsurutani, B. T., Chia de Gonzalez, A. L., & Dal Lago, A. (2011). Interplanetary origin of intense, superintensive and extreme geomagnetic storms. *Space Science Reviews, 158*(1), 69–99. https://doi.org/10.1007/s11214-010-9715-2

Gonzalez, W. D., Joselyn, J. A., Kamide, Y., Kroehl, H. W., Rostoker, G., Tsurutani, B. T., & Vasyliunas, V. M. (1994). What is a geomagnetic storm. *Journal of Geophysical Research, 99*(A4), 5771–5792. https://doi.org/10.1029/93JA0267

Gopalaswamy, N. (2018). Extreme solar eruptions and their space weather consequences. In N. Buzulukova (Ed.), *Extreme space weather: Origins, predictability, and consequences* (Vol. 2, pp. 37–63). Amsterdam, The Netherlands: Elsevier.

Hagwood, M. A. (2011). Towards a scientific understanding of the risk from extreme space weather. *Advances in Space Research, 47*(12), 2059–2072. https://doi.org/10.1016/j.asr.2010.02

Hashorva, E., Pakes, A. G., & Tang, Q. (2010). Asymptotics of random contractions. *Insurance: Mathematics and Economics, 47*(3), 405–414. https://doi.org/10.1016/j.jnsma.2010.08.006

Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements, quart. *Quarterly Journal of the Royal Meteorological Society, 81*(348), 158–171. https://doi.org/10.1002/qj.49708134804

Jiggins, F., Chavy-Macdonald, M., Santin, G., Menicucci, A., Evans, H., & Hilgers, A. (2014). The magnitude and effects of extreme solar particle events. *Journal of Space Weather and Space Climate, 4*. A20. https://doi.org/10.1051/swsc/2014017

Kataoka, R. (2013). Probability of occurrence of extreme magnetic storms. *Space Weather, 11*, 214–218. https://doi.org/10.1029/2004sw000417

Kilpua, E., Koskinen, H. E. J., & Pulkkinen, T. I. (2017). Coronal mass ejections and their sheath regions in interplanetary space. *Living Reviews in Solar Physics, 14*(1), 83. https://doi.org/10.1002/srl2.0009

Klimas, A. J., Valdivia, J. A., Vassiliadis, D., Baker, D. N., Hesse, M., & Takalo, J. (2000). Self-organized criticality in the substorm phenomenon and its relation to localized reconnection in the magnetospheric plasma sheet. *Journal of Geophysical Research, 105*(A8), 18,765–18,780. https://doi.org/10.1029/1999JA000319

Knipp, D. J., Ramsay, A. C., Beard, E. D., Boright, A. L., Cade, W. B., Hewins, I. M., et al. (2016). The May 1967 great storm and radio disruption event: Extreme space weather and extraordinary responses. *Space Weather, 14*, 614–633. https://doi.org/10.1002/2016SW001423

Koons, H. C. (2001). Statistical analysis of extreme values in space science. *Journal of Geophysical Research, 106*(A6), 10,915–10,921. https://doi.org/10.1029/2000ja900234

Kotz, E. (2018). Hermanus Magnetic Observatory: A historical perspective of geomagnetism in southern Africa. *History of Geo- and Space Sciences, 9*(2), 125–131. https://doi.org/10.1016/j.hgsps.9-125-2018

Loewe, C. A., & Prölss, G. W. (1997). Classification and mean behavior of magnetic storms. *Journal of Geophysical Research, 102*(A7), 14,209–14,213. https://doi.org/10.1029/96JA04020

Lundstedt, H. (2006). The sun, space weather and GIC effects in Sweden. *Space Weather, 12*, 120–146. https://doi.org/10.1029/2005SW000684

Love, J. J., Coïsson, P., & Pulkkinen, A. (2016a). Global statistical maps of extreme-event magnetic observatory 1 min first differences in horizontal intensity. *Geophysical Research Letters, 43*, 4126–4135. https://doi.org/10.1002/2016GL068664

Love, J. J., & Finn, C. A. (2011). The USGS geomagnetism program and its role in space weather monitoring. *Space Weather, 9*, S07001. https://doi.org/10.1029/2011SW000664

Love, J. J., Hayakawa, H., & Cliver, E. W. (2019a). On the intensity of the magnetic superstorm of September 1909. *Space Weather, 17*, 37–45. https://doi.org/10.1029/2018SW000279

Love, J. J., Hayakawa, H., & Cliver, E. W. (2019b). Intensity and impact of the New York Railroad superstorm of May 1921. *Space Weather, 17*, 1281–1292. https://doi.org/10.1029/2019SW002250

Love, J. J., Cliver, E. W., Kelbert, A., & Bedrosian, P. A. (2018). Geoelectric hazard maps for the mid-Atlantic United States: 100-year extreme values and the 1989 magnetic storm. *Geophysical Research Letters, 45*, 5–15. https://doi.org/10.1029/2017GL076042

Love, J. J., Rigler, E. J., Pulkkinen, A., & Riley, P. (2015). On the normality of high magnetic storm intensity statistics: Implications for extreme-event probabilities. *Geophysical Research Letters, 42*, 6544–6553. https://doi.org/10.1002/2015GL064842

Lundstedt, H. (2006). The sun, space weather and GIC effects in Sweden. *Advances in Space Research, 37*(6), 1182–1191. https://doi.org/10.1016/j.asr.2005.10.023

Maike, J., & Resnick, S. (2011). Testing the Pareto against the lognormal distributions with the uniformly most powerful unbiased test applied to the distribution of cities. *Physical Review E, 83*, 36111. https://doi.org/10.1103/PhysRevE.83.036111

Mitzenmacher, M. (2004). A brief history of generative models for power law and lognormal distributions. *Internet Mathematics, 1*(2), 226–251. https://doi.org/10.1080/15427951.2004.1029088

Montebrero, P., Bennett, R. J., van Lieshout, C., & Smith, H. (2019). A tale of two tails: Do power law and lognormal models fit firm-size distributions in the mid-Victorian era? *Physics A, 523*, 858–875. https://doi.org/10.1016/j.physa.2019.02.054
National Science and Technology Council (2019). *National space weather strategy and action plan*, pp. 1–13. Washington, DC: Executive Office.

Newell, P. T., Sotirelis, T., Liou, K., Meng, C. I., & Rich, F. I. (2007). A nearly universal solar wind–magnetosphere coupling function inferred from 10 magnetospheric state variables. *Journal of Geophysical Research, 112*, A01206. https://doi.org/10.1029/2006JA011205

Newman, M. E. I. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics, 46*(5), 323–351. https://doi.org/10.1080/0010751050052444

Nikitina, L., Trichtchenko, L., & Boteler, D. (2016). Assessment of extreme values in geomagnetic and geoelectric field variations for Canada. *Space Weather, 14*, 481–494. https://doi.org/10.1002/2015SW001386

North American Electric Reliability Corporation (2016). *Atlanta GA*, pp. 1–26.

Oughton, E. J., Hapgood, M., Richardson, G. S., Beggan, C. D., Thomson, A. W. P., Gibbs, M., et al. (2018). A risk assessment framework for the socio-economic impacts of electricity transmission infrastructure failure due to space weather. *Paper no. 01/2018*. Cambridge, UK: Cambridge Judge Business School.

Papalexou, S. M., & Koustoviyannis, D. (2013). Battle of extreme value distributions: A global survey on extreme daily rainfall. *Water Resources Research, 49*, 187–201. https://doi.org/10.1029/2012WR012557

Perline, R. (2005). Strong, weak and false inverse power laws. *Statistical Science, 20*(1), 68–88. https://doi.org/10.1214/08834230400000215

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1992). *Numerical recipes*, pp. 1–963. Cambridge, UK: Cambridge University Press.

Pulkkinen, A., Bernabeu, E., Eichner, J., Beggan, C., & Thomson, A. W. P. (2012). Generation of 100-year geomagnetically induced current scenarios. *Space Weather, 10*, S04003. https://doi.org/10.1002/2011SW000750

Richardson, I. G., & Cane, H. V. (2012). Solar wind drivers of geomagnetic storms during more than four solar cycles. https://doi.org/10.1013/sws.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., Baker, D., Liu, Y. D., Verronen, P., Singer, H., & Güdel, M. (2018). Extreme space weather events: From cradle to grave. *Space Science Reviews, 214*(1), 21. https://doi.org/10.1007/s11214-017-01456-3

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

Ruzmaikin, A., Feynman, J., & Jun, I. (2011). Distribution of extreme solar energetic proton fluxes. *Journal of Atmospheric and Solar-Terrestrial Physics, 73*(2–3), 300–307. https://doi.org/10.1016/j.jastp.2009.12.016

Saarlo, P. (2007). Six rules for effective forecasting. *Harvard Business Review, 1937–8, 122–131.

Scarfott, C., & MacDonald, A. (2012). A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT-Statistical Journal, 10*(1), 33–60.

Schröder, C. J., et al. (2015). Understanding space weather to shield society: A global road map for 2015–2025 commissioned by COSPAR and ILWS. *Advances in Space Research, 55*(12), 2745–2807. https://doi.org/10.1016/j.asr.2015.03.023

Schumman, A. Y., Moloney, N. R., & Davidsen, J. Extreme value and record statistics in heavy-tailed processes with long-range memory. In A. S. Sharma, A. Bunde, V. P. Dimri, & D. N. Baker (Eds.), *Extreme events and natural hazards: The complexity perspective* (Vol. 196, pp. 315–334). Geophysical Monograph Series. Washington DC: American Geophysical Union. (2013). https://doi.org/10.1029/2011GM001088

Siccoo, G. L., Crocker, N. U., & Clark, C. R. (2006). Dist of the Carrington storm of 1859. *Advances in Space Research, 38*(2), 173–179. https://doi.org/10.1016/j.asr.2005.02.102

Solanki, S. K., Schüssler, M., & Fligge, M. (2000). Evolution of the Sun’s large-scale magnetic field since the Maunder minimum. *Nature, 408*, 445–447.

Sornette, D. (2006). *Critical phenomena in natural sciences: Chaos, fractals, selforganization and disorder: Concepts and tools*. Berlin, Germany: Springer-Verlag.

Stapler, R., Lockwood, M., Wild, M. N., & Clark, T. D. G. (1999). Solar causes of the long-term increase in geomagnetic activity. *Journal of Geophysical Research, 104*(A12), 28,325–28,342. https://doi.org/10.1029/1999JA900311

Steinkogel, D. J., Thonstheim, D. B., & Kvanmto, N. G. (2007). A cautionary note on the use of the Kolmogorov-Smirnov test for normality. *Monthly Weather Review, 135*(3), 1115–1137. https://doi.org/10.1175/MWR3236.1

Sugita, M. (1964). Hourly values of equatorial Dst for the IGY. *Annales Geophysicae, 16*, 191–196. https://doi.org/10.1016/j.jastp.2010.05.012

Sugita, M., & Kamei, T. (1991). Equatorial Dst index 1957–1986, IAGA Bull, vol. 40. Japan: ISGI Publications Office, Saint-Maur-des-Fossés, France.

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

Telles, M. R., & De la Luz, V. (2018). Extreme value analysis of solar flare events. *Space Weather, 16*, 1984–1996. https://doi.org/10.1029/2018SW001938

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

Tzurutani, B., et al. (2006). Corotating solar wind streams and recurrent geomagnetic activity: A review. *Journal of Geophysical Research, 111*, A07S01. https://doi.org/10.1029/2005JA011273

Turcotte, D. L. (1999). Self-organized criticality. *Reports on Progress in Physics, 62*, 1377–1429.

Uritsky, V. M., & Pudovkin, M. I. (1998). Low frequency 1/f-like fluctuations of the AE-index as a possible manifestation of self-organized criticality in the magnetosphere. *Annales Geophysicae, 16*(12), 1580–1588. https://doi.org/10.1007/s00585-998-1580-x

Vassiliadis, D., Klimas, A. J., Valdivia, J. A., & Clauer, C. R. (2006). Dst of the Carrington storm of 1859. *Journal of Geophysical Research, 104*(A11), 24,957–24,976. https://doi.org/10.1029/1999JA900185

Vayuléinas, V. M. (2011). The largest imaginable magnetic storm. *Journal of Atmospheric and Solar-Terrestrial Physics, 73*(11–12), 1444–1446. https://doi.org/10.1016/j.jastp.2010.05.012

Verbeek, C., Kraaijkamp, E., Ryan, D. F., & Podladchikova, O. (2019). Solar flare distributions: Lognormal instead of power law? *The Astrophysical Journal, 884*(1), 1–16. https://doi.org/10.3847/1538-4357/ab3425

Veselovsky, I. S., Dmitriev, A. V., & Susorov, A. V. (2010). Algebra and statistics of the solar wind. *Cosmic Research, 48*(2), 113–128. https://doi.org/10.1134/S0301032X10002012
Walshaw, D. (2013). Generalized extreme value distribution. In A. H. El-Shaarawi, & W. W. Piegorsch (Eds.), Encyclopedia of environmetrics. New York, NY: John Wiley. https://doi.org/10.1002/9780470057339.vae062.pub2

Watari, S., Kunitake, M., & Watanabe, T. (2001). The Bastille Day (14 July 2000) event in historical large Sun-Earth connection events. Solar Physics, 204(1), 425–438. https://doi.org/10.1023/A:1014271227639

Weigel, R. S., & Baker, D. N. (2003). Probability distribution invariance of 1-minute auroral-zone geomagnetic field fluctuations. Geophysical Research Letters, 30(23), 2183. https://doi.org/10.1029/2003GL018470

Westra, S., Alexander, L. V., & Zwiers, F. W. (2013). Global increasing trends in annual maximum daily precipitation. Journal of Climate, 26(11), 3904–3918. https://doi.org/10.1175/JCLI-D-12-00502.1

Willis, D. M., Stevens, P. R., & Crothers, S. R. (1997). Statistics of the largest geomagnetic storms per solar cycle (1844–1993). Annales Geophysicae, 15(6), 719–728. https://doi.org/10.1007/s00585-997-0719-5

World Data Center for Geomagnetism, Kyoto, Nosé, M., Iyemori, T., Sugiura, M., & Kamei, T. (2015). Geomagnetic Dst index. https://doi.org/10.17593/14515-7400

Yermolaev, Y. I., Lodkina, I. G., Nikolaeva, N. S., & Yermolaev, M. Y. (2013). Occurrence rate of extreme magnetic storms. Journal of Geophysical Research: Space Physics, 118, 4760–4765. https://doi.org/10.1002/jgra.50467

Yurchyshyn, V., Yashiro, S., Abramenko, V., Wang, H., & Gopalswamy, N. (2005). Statistical distributions of speeds of coronal mass ejections. The Astrophysical Journal, 639(1), 599–603. https://doi.org/10.1086/426129

Zwart, B. (2009). Lecture notes: Rare events and heavy tails in stochastic systems.