Space Weather®

RESEARCH ARTICLE
10.1029/2022SW003304

Extreme Event Statistics in Dst, SYM-H, and SMR Geomagnetic Indices

A. Bergin[1], S. C. Chapman[1,2], N. W. Watkins[1,3,4], N. R. Moloney[5,6], and J. W. Gjerloev[7,8]

[1]Centre for Fusion, Space and Astrophysics, Physics Department, University of Warwick, Coventry, UK, [2]Department of Mathematics and Statistics, University of Tromso, Tromso, Norway, [3]London School of Economics and Political Science, Grantham Research Institute on Climate Change and the Environment, London, UK, [4]Faculty of Science, Technology, Engineering, and Mathematics, School of Engineering and Innovation, The Open University, Milton Keynes, UK, [5]Institute of Mathematics and Computer Sciences, São Paulo University, São Carlos, Brazil, [6]Department of Physics, Imperial College London, London, UK, [7]Johns Hopkins University Applied Physics Laboratory, Laurel, MD, USA, [8]Department of Physics and Technology, University of Bergen, Bergen, Norway

Key Points:

- These indices are not totally interchangeable, consideration should be given to index choice in model validation or cross-study comparison
- Hourly averaged SMR and SYM-H return levels track Dst for return periods below 10 years. Above that they exceed Dst; at 100 years by >10%
- One minute cadence SMR and SYM-H 5, 10, 50, and 100 year return levels exceed that of Dst by about 10%, 12%, 20%, and 25% respectively

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

A. Bergin,
aisling.bergin@warwick.ac.uk

Citation:

Bergin, A., Chapman, S. C., Watkins, N. W., Moloney, N. R., & Gjerloev, J. W. (2023). Extreme event statistics in Dst, SYM-H, and SMR geomagnetic indices. *Space Weather*, *21*, e2022SW003304. https://doi.org/10.1029/2022SW003304

Received 28 SEP 2022
Accepted 6 MAR 2023

Abstract Extreme space weather events are rare, and quantifying their likelihood is challenging, often relying on geomagnetic indices obtained from ground-based magnetometer observations that span multiple solar cycles. The Dst index ring-current monitor, derived from an hourly average over four low-latitude stations, is a benchmark for extreme space weather events, and has been extensively studied statistically. We apply extreme value theory (EVT) to two geomagnetic ring current indices: SYM-H (derived from 6 stations) and SMR (derived from up to 120 stations). EVT analysis reveals a divergence between the return level found for Dst, and those for SYM-H and SMR, that increases non-linearly with return period. For return periods below 10 years, hourly averaged SYM-H and SMR have return levels similar to Dst, but at return periods of 50 and 100 years, they respectively exceed that of Dst by about 10% and 15% (SYM-H) and about 7% and 12% (SMR). One minute resolution SYM-H and SMR return levels progressively exceed that of Dst; their 5, 10, 50, and 100 year return levels exceed that of Dst by about 10%, 12%, 20% and 25% respectively. Our results indicate that consideration should be given to the differences between the indices if selecting one to use as a bench mark in model validation or resilience planning for the wide range of space weather sensitive systems that underpin our society.

Plain Language Summary Extreme space weather events, which disturb Earth's near-Earth plasma and magnetic environment, have the potential to cause significant disruption to our infrastructure. Extreme events are rare, and quantifying their likelihood relies upon long-term continuous observations. High-quality ground-based magnetometer observations underpin geomagnetic indices that monitor space weather. The Dst index has become the de-facto measure for space weather storms. Space weather storms cause magnetic perturbation that can be localized in space and time. We now have the SYM-H and SuperMAG SMR indices, constructed to be 1-min versions of Dst, with SMR using a larger set of magnetometers. We perform the first extreme value theory analysis of SYM-H and SMR. We find that space weather storms are more intense when measured by the 1-min SYM-H or SMR indices but the relationship between extreme values identified in the Dst timeseries and those in SYM-H or SMR is not linear. Our results indicate that consideration should be given to the differences between the indices if selecting one to use as a benchmark in model validation or resilience planning for the wide range of space weather sensitive systems that underpin our society.

1. Introduction

Severe geomagnetic storms can give rise to widespread societal, economic, and technological disruption (Hapgood et al., 2021). Rare events, such as the 1859 Carrington storm (Cliver & Dietrich, 2013), and the 1921 storm (Hapgood, 2019) have been estimated to potentially result in nationwide disruption of power distribution (Oughton et al., 2017) or significant satellite loss (Horne et al., 2013). Consequently, long-term forecasting of magnetospheric space weather is of significant operational interest (Morley, 2020). A key element of an extreme space weather event is a dramatic and rapid enhancement of the large scale currents carried by the ionosphere and magnetosphere. This perturbs the magnetic field at the surface of the earth. Ground-based magnetometer observations, spanning multiple solar cycles, form the basis of geomagnetic indices which are routinely used

to quantify magnetospheric activity (Mayaud, 1980). Geomagnetic storms drive enhancements of earth's ring current and this can be monitored by low- and mid-latitude magnetometer stations.

The hourly Dst index is based upon an average of observations from four such low-latitude stations (Sugiura, 1964), and is commonly used in the characterization of geomagnetic storms (Gonzalez et al., 1994). Whilst it is understood that the detailed spatial and temporal evolution of a storm, which map directly onto its detailed impact, are not consistently captured by ring current indices (e.g., Sandhu et al. (2021) and references therein), the Dst index combines fidelity with multi-solar cycle coverage. Indices built on ground based magnetometers will inevitably sample space weather events in an inhomogeneous manner. The location of the ground based stations relative to the excited storm-time current systems will vary from one event to another, as will the number of stations and their duty cycles. Our results will therefore provide a lower bound on the event return level, which we anticipate will increase with improved spatial and temporal sampling in SMR as compared to Dst, particularly if these indices respond to current systems other than the symmetric ring current.

Storm occurrence rates and intensities vary with the approximately 11 years cycle of solar activity (Bergin et al., 2022; Chapman, McIntosh, et al., 2020) and each cycle is unique in amplitude and duration. Statistical studies to quantify space weather risk thus require observations spanning multiple solar cycles. Extreme events are by definition rare events, so that, whilst it is well known that their occurrence has a solar cycle modulation (Chapman, McIntosh, et al., 2020; Love, 2021), "benchmark" statistical estimates are based on observations across multiple cycles (Love et al., 2015; Riley, 2012). Provided the asymptotic tail of the underlying distribution (aggregated over a number of solar cycles) is sufficiently sampled, an extreme value analysis via peaks over threshold is still valid. As with previous work (Acero et al., 2018; Love et al., 2015; Riley, 2012; Tsubouchi & Omura, 2007), the study presented here should be understood as a solar cycle aggregate.

The Dst index is used as the de-facto benchmark in studies of space weather hazards (Cliver et al., 2022) and remains valuable as an indicator of geomagnetic field disturbance activity. The observed distribution of geomagnetic storm amplitudes, as seen in the maximum excursion of indices such as Dst, is routinely modeled to infer space weather risk. Such modeling provides an estimate of the amplitude of rare, extreme, "one in a hundred year" events. From physical constraints, there will be an upper limit to the size of such events, estimated at Dst $\sim$ −2,000 nT to −2,500 nT (Liu et al., 2020; Vasyliūnas, 2011). Events that have occurred in the past 100 years fall within this limit. For example, the 1921 storm which today could cause continent-wide disruption (Hapgood, 2019; Oughton et al., 2017) had a Dst estimated at $\sim$ −900 nT, comparable with the Carrington event (Love et al., 2019); whereas the 1989 storm which caused local disruption including a 9 hr power blackout in Quebec reached Dst of −589 nT.

One approach to space weather risk inference is to fit different specific functions to the observed distributions of storm sizes (Love, 2020; Love et al., 2015; Riley, 2012). In the distribution tail this leads to a wide range of estimates. For example, for a power-law tail the probability of another extreme geomagnetic event comparable to the Carrington event occurring within the next 10 years is 10.3% 95% confidence interval (CI) [0.9,18.7] but is only 3.0% 95% CI[0.6,9.0] for a lognormal distribution (Riley & Love, 2017). Confidence intervals are generally large (Love et al., 2015; Riley, 2012). To increase the number of events in the sample and to cover a wider range of solar activity, lower fidelity indices that exist for longer time intervals have been re-engineered to estimate Dst, giving estimates for the 100-year return level of −663 nT, with a bootstrap 68% CI[−694, −497] nT (Love, 2021) based on the constrained Weibull model, and the 150-year Dst return level to be −809 95% CI[−955, −663] nT (Chapman, Horne, & Watkins, 2020) based on an exponential model. Estimated probabilities are sensitive to the assumed source distribution (Love, 2021; Riley & Love, 2017).

An alternative approach is extreme value theory (EVT) (see e.g., Coles, 2001; Embrechts et al., 1997; Leadbetter et al., 1983). EVT offers a systematic method for estimating the functional form of the source distribution, suitable for classes of distribution tail that encompass power law, log-normal and exponential. Previous studies have applied EVT to long-time record but relatively low fidelity indices (Elvidge, 2020; Koons, 2001; Silbergleit, 1999; Siscoe, 1976). The underlying observations are highly coarse grained compared to Dst (Chapman, Horne, & Watkins, 2020, and refs therein.). EVT was first applied to the Dst index for the years 1957–2001 (Tsubouchi & Omura, 2007), giving a 1 in 100 years storm level of −645.3 95% CI[−754.5, −536.1] nT, latterly for Dst over 1957–2014 this becomes −576.9 95% CI[−680.44, −473.37] nT as it incorporates more recent, quieter solar cycles.

In this work we will apply EVT to the 1-min resolution SMR and SYM-H indices, which have been constructed to mimic Dst but at considerably improved resolution. Ground based magnetometers typically have 1 min time resolution, and over three hundred are now operating globally. This data set is now being used through the SuperMAG collaboration (Gjerloev, 2012) to construct the SMR index which incorporates data from up to 120 mid and low latitude stations, and is produced at 1-min cadence (Newell & Gjerloev, 2012). With a greater station density and temporal resolution, SMR is more likely to resolve the space and time localized variations in storm time magnetospheric-ionospheric disturbances that are not consistently resolved by the Dst index (Newell & Gjerloev, 2012; Wanliss & Showalter, 2006). The SYM-H index is also produced at 1-min cadence, based on data from 6 stations from a network of 11 (Iyemori et al., 2010). The presence of more information content in SYM-H, compared to hourly Dst, has been reported (Bej et al., 2020). Although low-latitude indices such as Dst were designed to monitor the ring current, for more extreme events, other current systems can come into play, and can contribute to localized and intense disturbances. Regardless of their source, these ground disturbances and their temporal variations are potential space weather hazards.

We for the first time apply EVT to 1-min SMR and SYM-H indices. We isolate individual extreme events using the standard peak-over-threshold (POT) method and compare Dst events with those identified in 1-min cadence SYM-H and SMR, and their hourly averages. When the return level and period are empirically estimated for individual POT events, SYM-H and SMR track each other. For return periods less than about 10 years, Dst POT events track those of hourly averages SMR and SYM-H, and are systematically offset from POT events in 1 min SYM-H and SMR. The EVT analysis of the full set of POT events reveals a divergence between the return level found for Dst, and those for SMR and SYM-H, that increases non-linearly with return period.

2. Methods

Three geomagnetic indices are of relevance for the present study; Dst, SYM-H, and SMR. The Dst index is compiled by the WDC for Geomagnetism Kyoto (Nose et al., 2015) and is available at 1-hr cadence. Four magnetometer stations located at geomagnetic latitudes between −35° and +30° record horizontal component disturbances in the geomagnetic field. Dst is then the baseline-corrected, magnetic latitude-adjusted, average of the four hourly mean disturbances. Dst for years 1957–2016 is termed "final" Dst and is "provisional" for years 2017–2021. The SYM-H index is also compiled by WDC for Geomagnetism Kyoto (Iyemori et al., 2010), it is available at 1-min cadence. Each month, six stations from the network of 11 geomagnetic observatories with geomagnetic latitudes between −50° and +50° are selected for the derivation of SYM-H, depending on data availability and quality. SYM-H is the baseline-corrected, magnetic latitude-adjusted, average of the disturbance component at each minute for the 6 stations. The SMR index (Newell & Gjerloev, 2012) is a SuperMAG index, based on all ground magnetometer stations contributing to the SuperMAG network at geomagnetic latitudes between −50° and +50°. The number of available stations has increased greatly over time, from the first available SMR observation based on less than 5 stations, to a present maximum count of 120 stations. The 1-min averages of baseline-corrected, magnetic latitude-adjusted, horizontal component disturbances are averaged within four equally sized local time sectors to produce four local time SMR indices, SMR is then the average of these four local time indices.

The following procedure is used in this study to select distinct extreme events: (a) exceedances of threshold, u are identified in the timeseries; (b) exceedances of u which are separated by a duration less than cluster interval, r, are labeled part of the same event, but when the time interval between successive exceedances is greater than r they constitute distinct events; (c) the peak (maximum) level recorded during each event is identified; (d) each event is represented by the timestamp and magnitude of the event peak. Following Acero et al. (2018) and Tsubouchi and Omura (2007), a cluster interval $r = 48$ hr is used in the runs declustering of the geomagnetic indices throughout.

3. Results

We apply EVT to six geomagnetic index timeseries: the 1-min resolution SMR, available from 1973 (Newell & Gjerloev, 2012), 1-hr resolution Dst from the same interval, the full 1-hr resolution Dst, available from 1957 (Sugiura, 1964), an hourly average of SMR, which has the same temporal resolution as Dst, but is still constructed from a larger number of stations, the 1-min resolution SYM-H, available from 1981, and an hourly average of the SYM-H record. For convenience we will use minus the value of the index, for example, (−)Dst, throughout

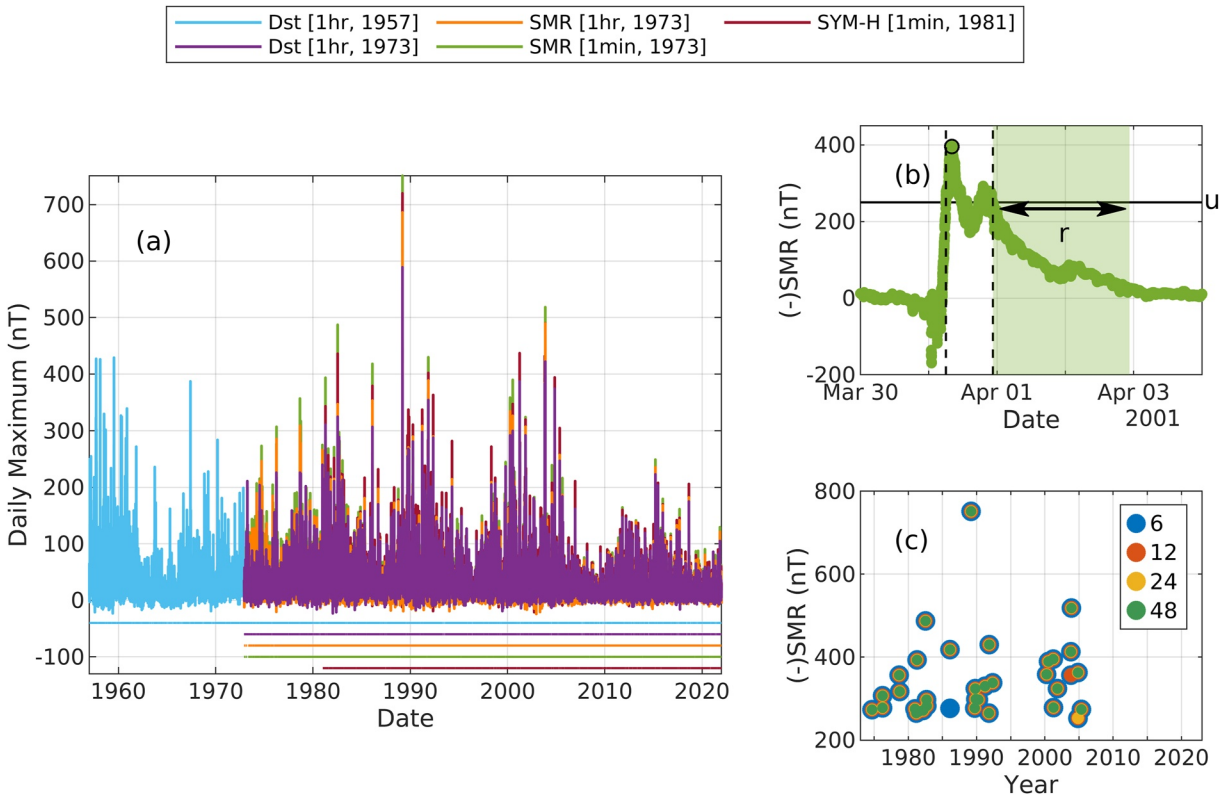

Figure 1. Data sets and declustering. (a) Daily maxima of indices are plotted with data availability indicated below the time series (horizontal line) for (−)Dst [1957–2021] (blue), (−)Dst [1973–2021] (purple), hourly average (−)SMR (orange), 1-min (−)SMR (green), and (−)SYM-H (red) indices. (b) Observations are plotted for (−)SMR [1-min, 1973] time series (green circles) between 30 March and 4 April 2001 showing how declustering is performed; threshold $u = 250$ nT (black solid line), event start and end points (black dashed lines), cluster interval $r = 48$ hr (green shading), and event peak-over-threshold (POT) (black open circle). (c) Event peak-over-thresholds for the (−)SMR [1-min, 1973] time series are plotted in time where distinct events are identified using cluster intervals $r = 6$ hr (blue), 12 hr (orange), 24 hr (yellow), and 48 hr (green).

(neglecting the—sign, geomagnetic storm perturbations are negative). The timeseries are overplotted in Figure 1a. Activity can be seen to vary with the approximately 11 years solar cycle, and with cycle strength, it is relatively weak over the last 11 years (the weak cycle 24). Working with the SMR index from 1973 then misses the relatively active period (1957–1973), however the plot shows that the interval from 1973 does sample a wide range of activity levels. We present results for both the full Dst from 1957, and the Dst from 1973, contemporaneous with SMR. In Figure 1a we can also see that for the most extreme events, hourly averaged SMR often exceeds hourly Dst, and hourly indices are systematically exceeded by 1-min SMR and 1-min SYM-H. Improved spatial resolution (in SMR), and temporal resolution (in both SYM-H and SMR) may translate into increased estimates of return levels for extreme space weather events (Newell & Gjerloev, 2012; Wanliss & Showalter, 2006).

We need to identify individual extreme events in these timeseries. We will use the standard POT method (Davison & Smith, 1990) such that events are identified where observations systematically exceed some high threshold, u. EVT requires that the extreme events to be modeled are independent but, in geomagnetic index time series, temporal clustering of threshold exceedances is expected. Declustering is the procedure most commonly used to obtain a sample of independent events (Coles, 2001; Embrechts et al., 1997; Fawcett & Walshaw, 2012). Here, runs declustering (Smith, 1989) is applied to the timeseries, the result of which is shown in Figure 1b. We specify a minimum cluster interval r such that any sequence of $\geq r$ consecutive observations below the threshold u discriminates two independent events. Following Acero et al. (2018) and Tsubouchi and Omura (2007), we use a cluster interval $r = 48$ hr Figure 1c shows that for $r > 6$ hr the results are insensitive to this choice.

EVT provides a framework to model the probability of extreme events that are rarely observed in a timeseries and to extrapolate these models to unobserved levels. The basis for the POT method (Davison & Smith, 1990) is that, for a stationary process, peak exceedances over a suitably high threshold, separated in time by a suitable cluster

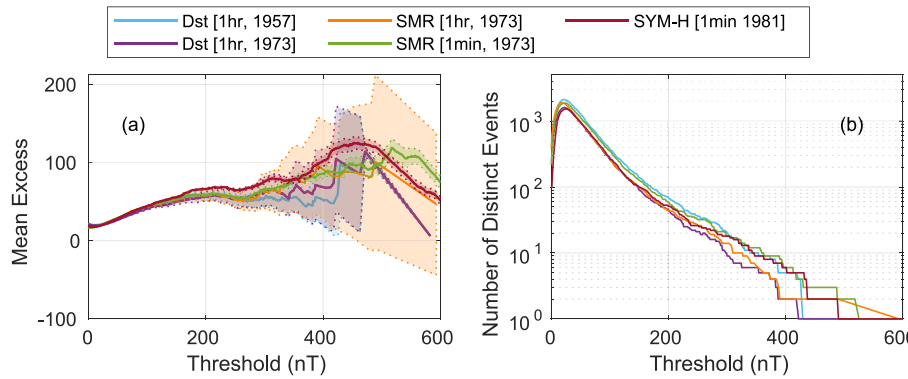

Figure 2. Mean Excess and event histograms. (−)Dst [1957–2021] (blue), (−)Dst [1973–2021] (purple), hourly average (−)SMR (orange), 1-min (−)SMR (green), and (−)SYM-H (red) indices. (a) Mean excess plot (solid) with approximate 95% CI (shaded). (b) Histogram of events.

interval, have a limiting distribution which asymptotically approaches the generalized Pareto distribution (GPD) with survival function:

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

defined on $\{y : y > 0 \text{ and } (1 + (\xi y)/\sigma) > 0\}$, where the GPD limiting distribution has two parameters: scale (σ), and shape (ξ). We obtain the GPD for each geomagnetic index timeseries. The specification of an appropriately high threshold, here $u = 250$ nT, for the GPD to be valid is addressed in the Supporting Information S1. A first indication of the GPD behavior is provided by a plot of the mean excess:

$$E(X - u|X > u) = \frac{1}{n_u}\sum_{i=1}^{n_u}(x_i - u) \tag{2}$$

where $x_1, \ldots, x_{n_u}$ are the n_u observations that exceed u. If a GPD is a reasonable model for excesses of a threshold u_0, for $u > u_0$, $E(X - u|X > u)$ is a linear function of u, with a gradient that determines the shape parameter. This is plotted in Figure 2a where we see that below 300 nT the mean excess is indistinguishable for Dst and SMR, but it diverges at higher thresholds. The SYM-H index mean excess systematically exceeds that of Dst or SMR samples over thresholds in the range 100–500 nT, but the shape of the SYM-H mean excess plot approximately tracks that of 1-min SMR. For a given time interval, the indices record approximately the same number of events, but 1-min indices have 60 times more data points than the 1-hr indices, so that the mean excess uncertainties are much smaller for 1-min SMR and SYM-H than 1-hr Dst or SMR timeseries. From Figure 2a, one can then discern for example, that the 1 min SMR is systematically more long-tailed than Dst; it has a larger positive gradient, hence larger positive shape (ξ) parameter on the mean excess plot, suggesting more long-tailed behavior. Figure 2b plots histograms of the number of events identified in each timeseries as a function of declustering threshold. The longer interval, full Dst record contains more geomagnetic storms than the other timeseries, so can more frequently populate the far-tail. For data taken over the same time interval (from 1973), 1-hr average SMR matches, and can exceed, Dst. Both Dst and 1-hr SMR from 1973 are systematically exceeded by the sizes of events in 1-min SMR and SYM-H. For thresholds greater than 375 nT, the number of POT events in 1-min SYM-H (from 1981) exceeds that of Dst (from 1957), due to the higher values exhibited by the 1-min SYM-H index. For thresholds beyond 300 nT, the number of events in 1-min SMR (from 1973), exceeds that of Dst (from 1957).

We obtain GPD fits to these samples of events. The GPD model fit to the peaks-over-thresholds are shown in Figures 3 and 4 and the parameter estimates with their 95% confidence intervals are given in Table 1. The estimated GPD shape and scale parameters agree with each other; estimates for ξ are close to zero but the confidence intervals span the three families of the GPD; $\xi < 0$, $\xi = 0$, and $\xi > 0$. Estimates of σ have large, overlapping confidence intervals. The more extreme events are, with the exception of the single largest event, reasonably

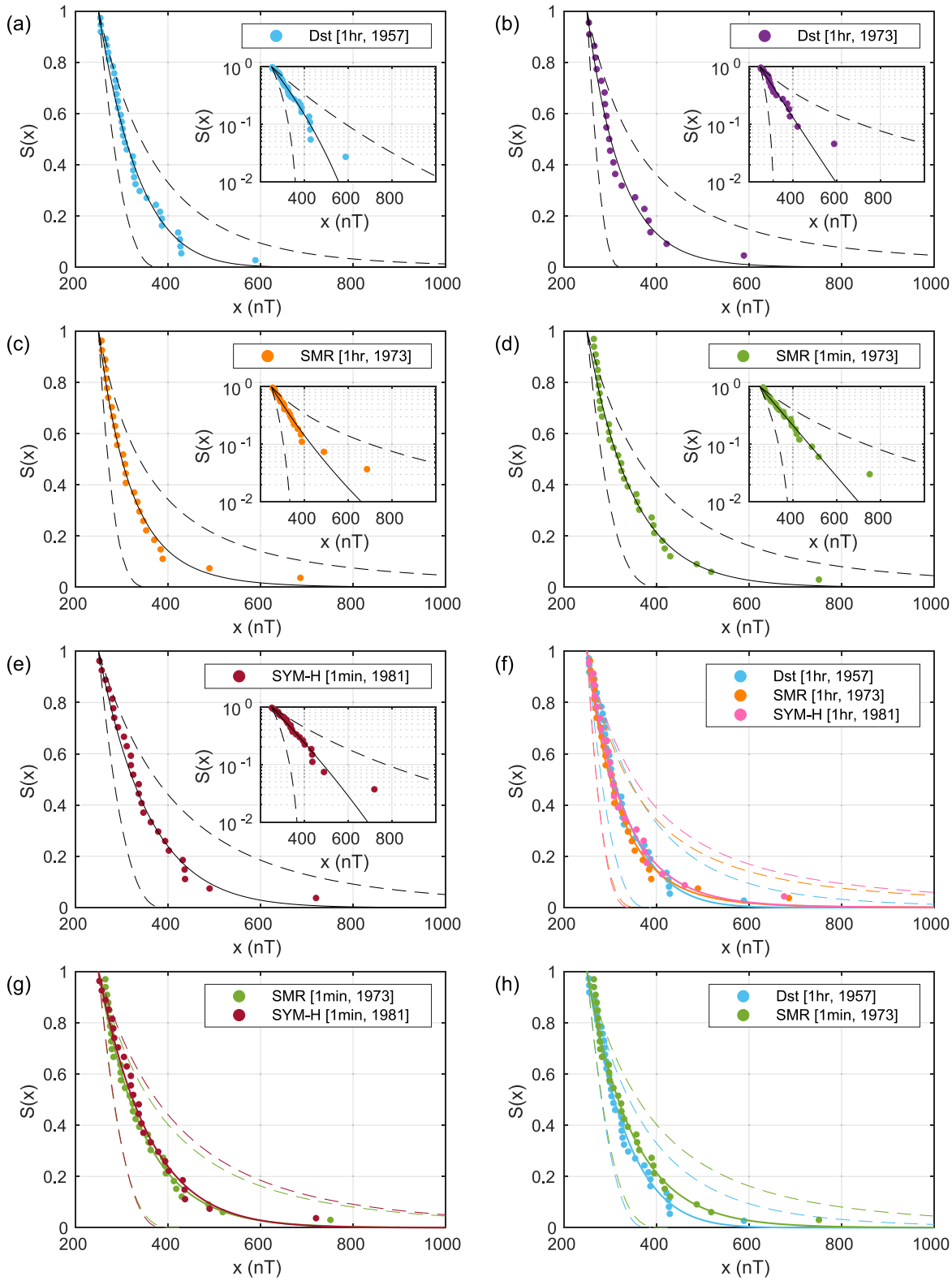

Figure 3. Generalized Pareto Distribution (GPD) models. Event peak-over-threshold values (filled circles) and corresponding GPD fit (black solid line) with 95% CI (black dashed line) for indices (a) (−)Dst [1957–2021] (blue), (b) (−)Dst [1973–2021] (purple), (c) hourly average (−)SMR (orange), (d) 1-min (−)SMR (green), and (e) 1-min (−)SYM-H (red). Insets repeat main plots with semi-log scale. (f) Overplot GPD fits and confidence intervals from (a) and (c) with GPD fit and 95% CI for hourly average (−)SYM-H index (pink). (g) Overplot GPD fits and confidence intervals from (d) and (e). (h) Overplot GPD fits and confidence intervals from (a) and (d).

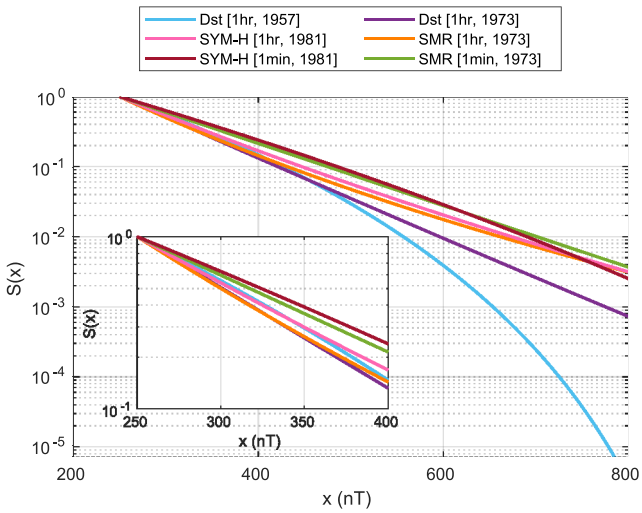

Figure 4. Compare Generalized Pareto Distribution (GPD) models. GPD fits to event peak-over-threshold values for indices (−)Dst [1957–2021] (blue), (−)Dst [1973–2021] (purple), hourly average (−)SMR (orange), 1-min (−)SMR (green), 1-min (−)SYM-H (red), and hourly average (−)SYM-H index (pink) from Figure 3 on semi-log axis. Inset shows zoom for index values between 250 and 400 nT.

tightly clustered around the GPD fit (see Figures 3a–3e), they fall well within the 95% CI. The details of each GPD fit are different, as can be seen in Figures 3f–3h and 4. Comparing their appearance on linear axes, the fits to 1 hr indices are similar (Figure 3f). Despite scatter between the underlying empirical distribution, the GPD fits track each other for Dst [1-hr, 1957], hourly mean SYM-H, and hourly mean SMR indices. Similarly, the 1-min SYM-H and SMR indices (Figure 3g) show some scatter in the underlying empirical distribution but the GPD fits appear similar on a linear axis. Where the 1-hr Dst (from 1957) and 1-min SMR empirical distribution and best GPD fits are compared (Figure 3h), the distributions do not track each other. When the best fit GPD curves are displayed with a semi-log axis (Figure 4), the differences between the curves based on shape and scale parameters tabulated in Table 1 can be more clearly seen. Up to around 350 nT there is reasonable agreement between the 1-hr SYM-H, 1-hr SMR and Dst indices, beyond this point the GPD curves diverge and the 1-hr SYM-H and 1-hr SMR curves exhibit longer tails than either Dst sample. 1-minute SYM-H and SMR distributions track each other reasonably closely and show longer tailed behavior than any of the 1-hr indices up to around 650 nT. We note that Dst (from 1957) gives a bounded GPD distribution $\xi < 0$ consistent with previous authors (Acero et al., 2018), as does 1-min SYM-H, whereas the other samples are unbounded $\xi > 0$. Whilst a bounded GPD distribution would have significant implications for extrapolating to the most extreme, Carrington-class events, we emphasize that the uncertainties contain both bounded and unbounded GPD fits. Our EVT based estimates should not be extrapolated beyond the aforementioned estimated upper limit to the size of events, Dst ∼ 2,000–2,500 nT (Liu et al., 2020; Vasyliūnas, 2011).

Risk is typically framed in terms of return level, for example, the 1 in 100 year return level is the amplitude that the index will exceed, on average, once in 100 years. If we rank-order the events in decreasing size, so that rank 1 is the largest observed event, then the rank is simply the number of times an event of that size has been seen to occur in the time interval of observation. The rank, divided by total number of events, is the survival function, so that our GPD fits translate directly into return level plots. Rather than invert the GPD fits, we use an expression for the return level that directly arises from the EVT (Coles, 2001, p. 103), as detailed in the SI. Estimates for the 5, 10, 50 and 100 year return levels for each index are tabulated in Table 2. Table 3 presents the percentage enhancement of return level estimates, relative to equivalent return period estimate and 95% CI for Dst (from 1957). Large confidence intervals for return levels translate to a large range in the confidence interval for percentage enhancement. Figures 5a–5e provide return level plots for the index timeseries. Return level estimates and

Table 1
Parameters of the Generalized Pareto Distribution Fit to Declustered Peaks-Over-Threshold Are Tabulated for Indices (−)Dst [1-hr, 1957], (−)Dst [1-hr, 1973], Hourly Average (−)SYM-H [1-hr, 1981], Hourly Average (−)SMR [1-hr, 1973], (−)SMR [1-min, 1973], and (−)SYM-H [1-min, 1981]

Index	Year	u	n	ξ	σ
Dst [1-hr, 1957]	1957	250	36	−0.13 [−0.39, 0.14]	88 [51, 125]
Dst [1-hr, 1973]	1972	250	21	0.01 [−0.41, 0.43]	73 [29, 117]
SYM-H [1-hr, 1981]	1981	250	22	0.06 [−0.35, 0.47]	80 [33, 126]
SMR [1-hr, 1973]	1972	250	26	0.10 [−0.28, 0.48]	70 [32, 107]
SYM-H [1-min, 1981]	1981	250	26	−0.06 [−0.39, 0.27]	109 [54, 164]
SMR [1-min, 1973]	1972	250	32	0.00 [−0.31, 0.32]	97 [52, 142]

Note. Year refers to the first year of the time series and n is the number of distinct events in the timeseries that is, declustered peaks over threshold. The GPD fit parameters consist of u the threshold, ξ the shape parameter and σ the scale parameter. Estimates of the 100-year return level for the indices based on the GPD model are presented. Estimates of the lower and upper bounds of 95% confidence intervals are in brackets.

Table 2
Estimates for 5, 10, 50, and 100 Year Return Levels (RL) Based on the Generalized Pareto Distribution Model Are Presented for Each Index

Index	5-year RL	10-year RL	50-year RL	100-year RL
Dst [1-hr, 1957]	337 [304, 371]	384 [342, 427]	485 [409, 560]	524 [424, 624]
Dst [1-hr, 1973]	315 [276, 353]	362 [310, 414]	476 [365, 587]	527 [369, 685]
SYM-H [1-hr, 1981]	331 [287, 375]	385 [320, 450]	537 [351, 723]	617 [328, 906]
SMR [1-hr, 1973]	324 [287, 361]	373 [318, 429]	516 [358, 673]	592 [348, 835]
SYM-H [1-min, 1981]	371 [319, 424]	433 [363, 504]	583 [426, 739]	650 [429, 870]
SMR [1-min, 1973]	359 [316, 403]	420 [352, 489]	603 [381, 824]	703 [352, 1054]

Note. Estimates of the lower and upper bounds of 95% confidence intervals are in brackets.

confidence bounds are overplotted with the rank-order estimates for the return levels of extreme events within the timeseries (the same events that form the basis of the GPD fits in Figure 3) and we can again see that the extreme events are, with the exception of the largest event, well within the 95% CI. The upper bound ($\xi < 0$) of Dst (from 1957) is evident in the curved return level in Figure 5a. Comparing the return level estimates based on Dst (from 1957) to that for Dst (from 1973) in Figure 5b, 95% CI bounds are larger for the estimate based on the shorter time period but contain the estimate for the full available timeseries. Though SMR can have up to 120 stations and SYM-H is limited to 6 stations, comparing the 1-min SYM-H and SMR return level estimates (Figure 5f), the similarity between best GPD fits (Figure 3g) equates to return period estimates that are also consistent, despite the scatter between SMR and SYM-H POT empirical distributions. Despite the difference in numbers of stations, the extreme value analysis of the two 1-min timeseries returns very similar results. The hourly average SYM-H and SMR index return levels are compared to 1-hr Dst (Figures 5g and 5h). Comparing the return levels in Table 2 and the percentage enhancement over Dst (from 1957) in Table 3, it is seen that, up to 10 year return periods, the return level estimates are largely similar. At 50-year return periods and at 100-year return periods, the return level estimates are higher for 1-hr SYM-H and SMR than for those obtained for 1-hr Dst. The 1-min SYM-H and SMR index return levels are compared to 1-hr Dst (Figures 5i and 5j). The return level estimates for the 1-min indices are larger than those for Dst; the percentage increases are progressively bigger for longer return periods. Similarity between the 1-min SYM-H and SMR return period plots equates to similar percentage increases above the 1-hr Dst return level plot. From Tables 2 and 3, 1-min SYM-H and SMR return level estimates exceed those for Dst by increasing percentages for 5, 10, 50 and 100 year return periods.

SYM-H and Dst have been compared in the literature before, differences in quiet time and Sq variation subtraction are mentioned as possible sources of differences between the indices (Gannon & Love, 2011; Katus & Liemohn, 2013; Wanliss & Showalter, 2006). Wanliss and Showalter (2006) compared hourly-averaged SYM-H values to corresponding hourly Dst values. They found a linear relationship between Dst and SYM-H, with a break in the fit at −300 nT; they found that increased activity coincided with larger differences between Dst and SYM-H, and that intense storms have deviations that are usually less than 20 nT. Wanliss and Showalter (2006) highlight multiple possible sources of differences between the indices, including spatial non-uniformity in ring current variations, magnetic field-aligned currents, and ionospheric currents at low latitudes; geologic

Table 3
Percentage Enhancement of 5, 10, 50, and 100 Year Return Level Estimates (RL) Relative to the Estimates for the Dst [1-hr, 1957] Index Timeseries

Index	5-year RL	10-year RL	50-year RL	100-year RL
Dst [1-hr, 1973]	−8% [−21%, 4%]	−7% [−22%, 8%]	−2% [−26%, 22%]	1% [−30%, 32%]
SYM-H [1-hr, 1981]	−2% [−17%, 14%]	1% [−18%, 19%]	10% [−23%, 43%]	15% [−29%, 59%]
SMR [1-hr, 1973]	−4% [−17%, 9%]	−2% [−18%, 14%]	7% [−24%, 37%]	12% [−29%, 54%]
SYM-H [1-min, 1981]	11% [−8%, 29%]	14% [−7%, 34%]	21% [−9%, 50%]	23% [−12%, 59%]
SMR [1-min, 1973]	9% [−7%, 26%]	12% [−6%, 31%]	21% [−9%, 51%]	25% [−14%, 64%]

Note. Based on results tabulated in Table 2, estimates of the 95% confidence intervals are in brackets.

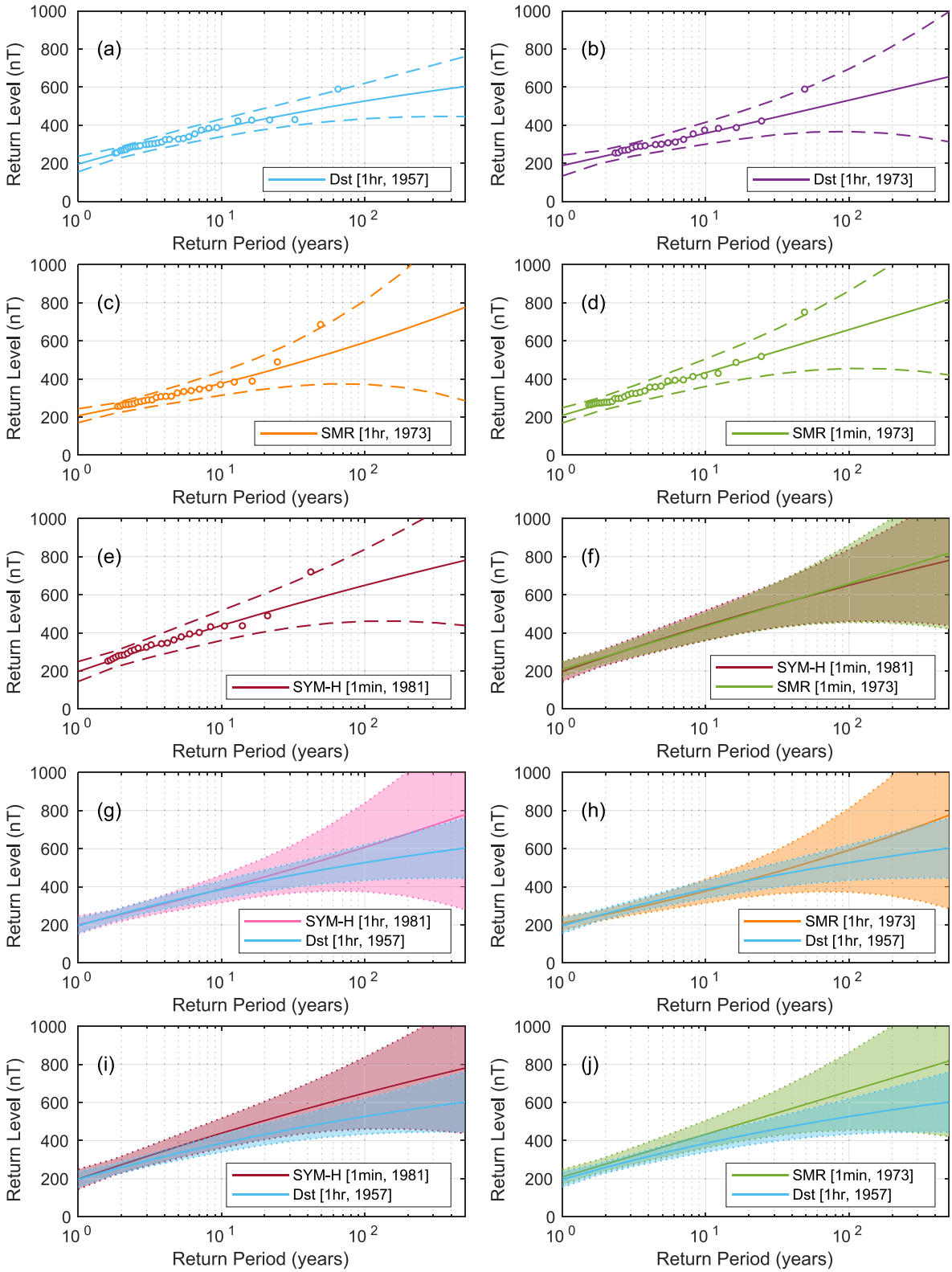

Figure 5. Return level plots. (a) (−)Dst [1957–2021] (blue), (b) (−)Dst [1973–2021] (purple), (c) hourly average (−)SMR (orange), (d) 1-min (−)SMR (green), and (e) (−)SYM-H indices. For each timeseries, extreme value theory (EVT) return level estimates (solid line) with 95% confidence intervals (dashed line) are plotted. Circles indicate empirical estimates of the return levels for observed events (declustered exceedances of $u = 250$ nT). The EVT return level estimates are compared between (f) 1-min (−)SMR and 1-min (−)SYM-H. The EVT return level estimates are compared between 1-hr (−)Dst[1957–2021] and (g) 1-hr (−)SYM-H (pink), (h) 1-hr (−)SMR, (i) 1-min (−)SYM-H, (j) 1-min (−)SMR indices.

conductivity structures beneath the stations used to compute the indices are also variable; furthermore, transients such as those formed by the compression of the magnetopause by coherent structures in the solar wind result in fluctuations that vary on timescales shorter than the cadence of Dst observations. Katus and Liemohn (2013) also show that the differences between SYM-H and Dst are largest during super-storms. They find a difference of approximately 20% between the indices storm peaks. A record of Dst [1-min, 1985–2009] is studied in their analysis, the authors agree with a previous study (Gannon & Love, 2011) linking the differences between Dst [1-min, 1985–2009] and SYM-H indices to greater sensitivity of SYM-H to auroral zone ionospheric currents, as a result of the additional stations at higher magnetic latitudes in the SYM-H network.

In Figure 6, the POT values recorded by different indices are compared for events where both indices register a POT for the same event. Index record lengths vary so identified POT samples also vary from index to index; moreover, for some events one index will exceed 250 nT while another does not cross this threshold, thus a sub-sample of events are established for each pair of indices, where POT values have been identified in both of the index timeseries. Figure 7 gives an indication of which events will be included in the analysis of Figure 6; filled circles represent POT values of the SYM-H or SMR indices for which a corresponding POT has been identified in Dst, open circles indicate an POT value in the SYM-H or SMR timeseries for which no corresponding Dst POT has been identified. POT identification is consistent across the indices for extreme events with estimated return periods greater than 5 years. Comparing the POT values in Figure 6, we find that the SYM-H and SMR indices most closely track each (panels a,d); with percentage enhancement of $\pm$10% for 1-hr SMR over 1-hr SYM-H. 1-min SYM-H and 1-min SMR also show agreement (panels g, j), even for the largest event in the study with SYM-H $\sim$750 nT on 14 March 1989. SYM-H is based on only 6 stations, whereas SMR can include data from up to 120 stations. These results suggest that it is the temporal, rather than spatial, sampling that is the key driver of differences found between 1-min SMR and Dst. We that the hourly average SYM-H and SMR POT values correspond to similar values of Dst (panels b,c,e,f). We find that the most significant difference between POT values arises when 1-hr Dst is compared to the 1-min SYM-H and SMR indices. There is greater deviation to the left of the diagonal (panels h and i), and percentage enhancement of either 1-min index over Dst ranges between $-$10% and +50% (panels k,l), with a weak positive trend seen for greater percentage enhancement for larger values of Dst. In their analysis, Gannon and Love (2011) name the inclusion of higher latitude stations in the SYM-H network as a possible source of difference between 1-min SYM-H and 1-min Dst. This comparison indicates that increased cadence of the SYM-H index relative to Dst has a greater impact than the inclusion of higher latitude stations. Figure 7 shows empirical estimates of the return levels for events identified in the timeseries. Comparing return level estimates for 1-hr indices (Figure 7a), we find that the rank-order empirical estimates for return periods of POT values are comparable, up to periods of 10 years. When the 1-min indices are compared to 1-hr Dst from 1957 (Figure 7b), the empirical return level plots do not track each other; in particular, the relationship between these indices is not linear and for extreme values these geomagnetic indices are not directly interchangeable.

4. Discussion and Conclusions

We performed a comparative extreme value analysis between the recent SuperMAG SMR index, the SYM-H index, and the Dst index, the latter being a benchmark for assessing the risk of geomagnetic storms. Compared to Dst, the SYM-H index offers greater temporal resolution and the SMR index has comparatively higher spatial and temporal resolution. As such, the SYM-H and SMR indices are expected to more closely resolve spatially and temporally localized extremes in the ground magnetic field disturbance (Newell & Gjerloev, 2012; Wanliss & Showalter, 2006). For extreme events, the contributions from the ring current (which these indices are designed to monitor) and contributions from magnetopause current (e.g., sudden commencement), substorm current wedge (both field-aligned currents and change in cross-tail current), or other magnetosphere-ionosphere current systems cannot be discriminated solely from the Dst, SYM-H or SMR indices (see e.g., Ohtani, 2022). Regardless of their source, such extremes are important since these short-time duration spikes in ground magnetic field perturbation are responsible for ground induced currents and hence damage to systems such as power transmission grids.

Whilst the time resolution of the indices does not vary within the record, the number of stations available to construct SMR increases with time (Bergin et al., 2020; Newell & Gjerloev, 2012). The spatial sensitivity of SMR is initially comparable with that of SYM-H but by halfway through the record has increased to over 50 stations. The derivation of the SMR index (see Section 2) involves averaging over stations within four local time sectors;

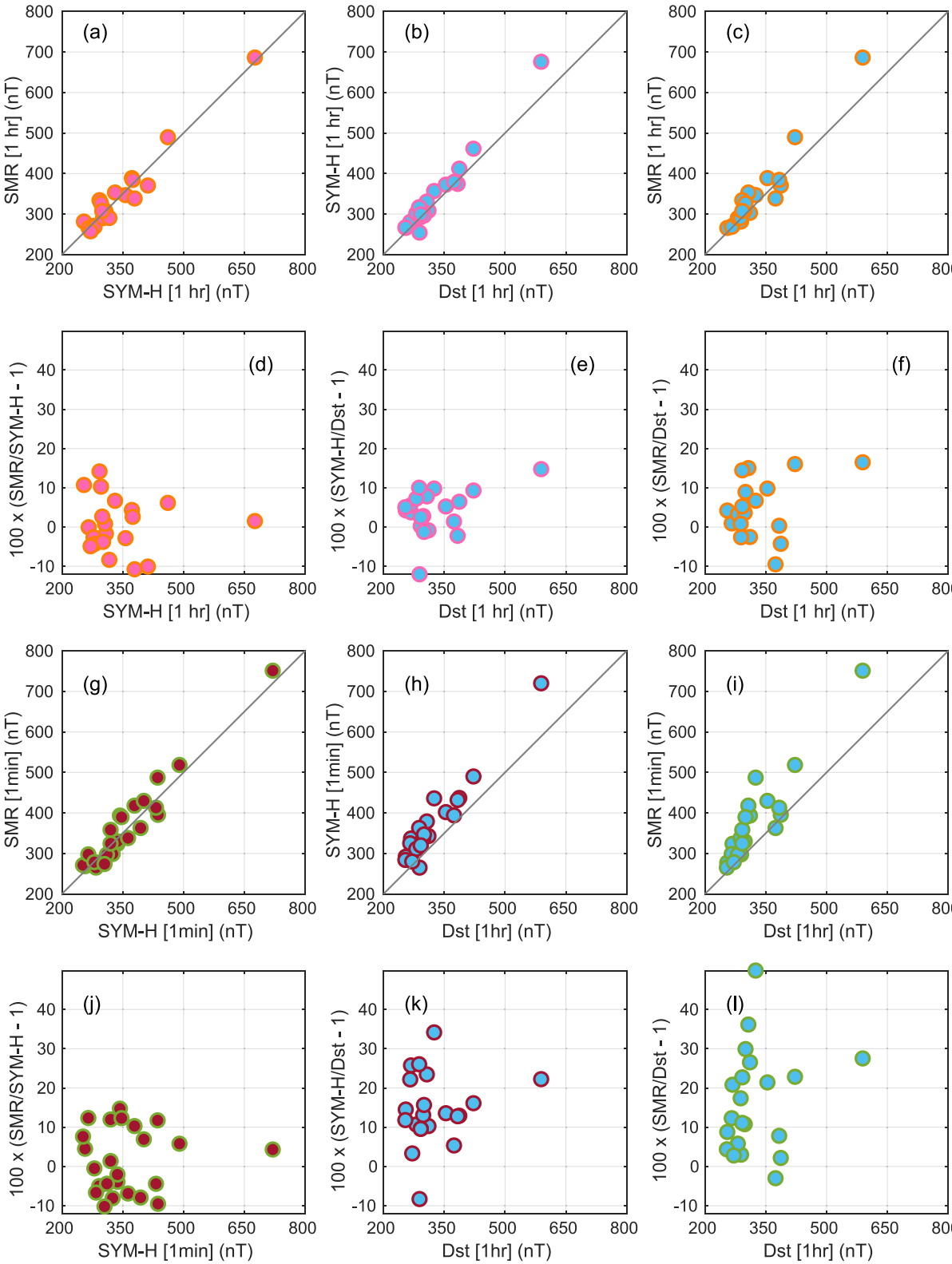

Figure 6. Comparison of peak-over-threshold (POT) events across indices. Corresponding POT values for individual events as parameterized by (a) 1-hr SYM-H and 1-hr SMR, (b) 1-hr Dst and 1-hr SYM-H, (c) 1-hr Dst and 1-hr SMR, (g) 1-min SYM-H and 1-min SMR, (h) 1-hr Dst and 1-min SYM-H, and (i) 1-hr Dst and 1-min SMR. Percentage enhancement of index relative to another for (d) 1-hr SMR over 1-hr SYM-H versus 1-hr SYM-H, (e) 1-hr SYM-H over Dst versus Dst, (f) 1-hr SMR over Dst versus Dst, (j) 1-min SMR over 1-min SYM-H versus 1-min SYM-H, (k) 1-min SYM-H over Dst versus Dst, and (l) 1-min SMR over Dst versus Dst.

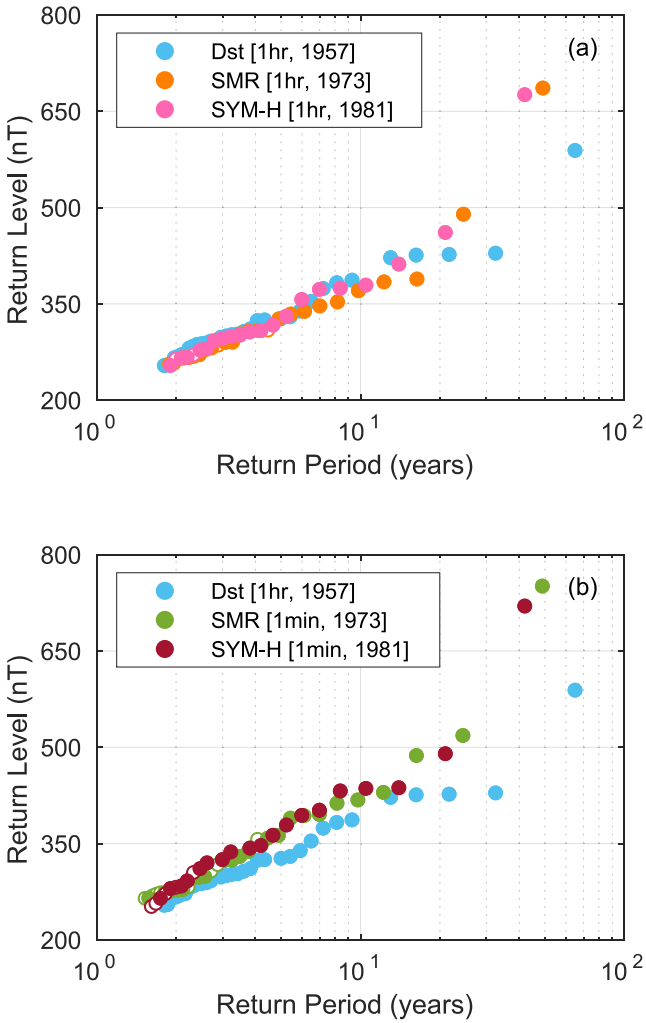

Figure 7. Comparison of event return levels across indices. Empirical estimates of the return levels for peak-over-threshold (POT) values are compared for (−)Dst[1957–2021] (blue), 1-hr (−)SMR (orange), and 1-hr (−)SYM-H indices (pink) (top panel), and (−)Dst[1957–2021] (blue), 1-min (−)SMR (green), and 1-min (−)SYM-H indices (red) (bottom panel). Open circles indicate POT events that were identified in SMR or SYM-H but not Dst that is, the (−)Dst index did not exceed the 250 nT threshold during the event. Filled circles correspond to the events presented in Figure 6.

increasing spatial resolution within each local time sector will act to increase the sensitivity of the index to geomagnetic perturbations. The reasonably close (−10%,+15%) coincidence of 1-min SMR and SYM-H POT values (Figures 6g and 6j) suggests that the significantly different station numbers supporting the indices are not a strong driver of differences in the POT return levels. This suggests that the change in station numbers with time in SMR does not significantly affect our results.

From the GPD distributions fitted to the data, we found that for these more extreme events, SYM-H and SMR predict disturbances that are more intense than previously obtained using Dst, however, the increase does not increase linearly with event severity. Regarding comparison with previous estimates for Dst return levels, we would expect that fitting specific parametric distributions such as power laws or lognormals would also yield significant increases in the estimated return levels. This implies that selection of indices can have significant implications for risk assessment and planning mitigation against space weather hazard, for example, for 1-in-10-year return levels, our estimates based on 1-min SYM-H or SMR exceed those based on Dst by about 12%, while the 1-in-100-year return levels show enhancement of almost 25%.

Risk is commonly parameterized by estimated intensity of the 1 in 100 years event; our results suggest that a more nuanced approach to index selection may be required. Our results suggest that for more extreme events, these geomagnetic indices are not directly interchangeable, they contain different, complementary information and should be used together in any detailed analysis. Such an analysis may improve the correlation between return levels and the impact of severe geomagnetic storms.

In their comparison of Dst and SYM-H, Katus and Liemohn (2013) call attention to the role that global storm indices play in model assessment; the authors highlight that the interchangeable use of indices, without accounting for the differences between the indices, could bias data-model comparison tests. These authors recommended the use of multiple indices in determination of low-to middle-latitude magnetic perturbations and suggested that difference between indices could be considered as an error estimate on these values. Some recent studies such as Brenner et al. (2021) use the Dst, SYM-H, and SMR indices interchangeably. Our results show that the indices are not interchangeable, particularly during extreme events, and further emphasize the consideration that should be given to index choice in model validation and cross-study comparison.

Data Availability Statement

The Dst and SYM-H indices used in this paper were provided by the WDC for Geomagnetism, Kyoto (http://wdc.kugi.kyoto-u.ac.jp/wdc/Sec3.html) and were obtained from the WDC Kyoto interface (Dst: https://wdc.kugi.kyoto-u.ac.jp/dstae/index.html), (SYM-H: https://wdc.kugi.kyoto-u.ac.jp/aeasy/index.html). The SMR index was retrieved from the SuperMAG interface (https://supermag.jhuapl.edu/indices/). The download date was 6 December 2022.

Acknowledgments
We gratefully acknowledge the SuperMAG collaborators (https://supermag.jhuapl.edu/info/?page=acknowledgement). We acknowledge use of the MATLAB EVIM package (Gençay et al., 2001). S.C.C., N.W.W., and A. B. acknowledge AFOSR Grant FA8655-22-7056. S.C.C. acknowledges STFC Grant ST/T000252/1 and support from ISSI via the J. Geiss fellowship. N.R.M. acknowledges support from the Serrapilheira Institute (Grant Serra-1709-16124). We thank the editor and reviewers for suggestions made during the review process.

References

Acero, F. J., Vaquero, J. M., Gallego, M. C., & García, J. A. (2018). A limit for the values of the Dst geomagnetic index. *Geophysical Research Letters*, *45*(18), 9435–9440. https://doi.org/10.1029/2018GL079676

Bej, A., Banerjee, A., Chatterjee, T., & Majumdar, A. (2020). A comparative study between Dst and SYM-H indices based on pattern identification. *Indian Journal of Theoretical Physics*, 13.

Bergin, A., Chapman, S. C., & Gjerloev, J. W. (2020). AE, Dst, and their SuperMAG counterparts: The effect of improved spatial resolution in geomagnetic indices. *Journal of Geophysical Research: Space Physics*, *125*(5), e2020JA027828. https://doi.org/10.1029/2020JA027828

Bergin, A., Chapman, S. C., Moloney, N. R., & Watkins, N. W. (2022). Variation of geomagnetic index empirical distribution and burst statistics across successive solar cycles. *Journal of Geophysical Research: Space Physics*, *127*(1), e2021JA029986. https://doi.org/10.1029/2021JA029986

Brenner, A., Pulkkinen, T. I., Al Shidi, Q., & Toth, G. (2021). Stormtime energetics: Energy transport across the magnetopause in a global MHD simulation. *Frontiers in Astronomy and Space Sciences*, *180*. https://doi.org/10.3389/fspas.2021.756732

Chapman, S. C., Horne, R. B., & Watkins, N. W. (2020). Using the index over the last 14 solar cycles to characterize extreme geomagnetic activity. *Geophysical Research Letters*, *47*(3), e2019GL086524. https://doi.org/10.1029/2019GL086524

Chapman, S. C., McIntosh, S. W., Leamon, R. J., & Watkins, N. W. (2020). Quantifying the solar cycle modulation of extreme space weather. *Geophysical Research Letters*, *47*(11), e2020GL087795. https://doi.org/10.1029/2020GL087795

Cliver, E. W., & Dietrich, W. F. (2013). The 1859 space weather event revisited: Limits of extreme activity. *Journal of Space Weather and Space Climate*, *3*, A31. https://doi.org/10.1051/swsc/2013053

Cliver, E. W., Schrijver, C. J., Shibata, K., & Usoskin, I. G. (2022). Extreme solar events. *Living Reviews in Solar Physics*, *19*(1), 1–143. https://doi.org/10.1007/s41116-022-00033-8

Coles, S. (2001). *An introduction to statistical modeling of extreme values*. Springer London. https://doi.org/10.1007/978-1-4471-3675-0

Davison, A. C., & Smith, R. L. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society: Series B*, *52*(3), 393–425. https://doi.org/10.1111/j.2517-6161.1990.tb01796.x

Elvidge, S. (2020). Estimating the occurrence of geomagnetic activity using the Hilbert-Huang transform and extreme value theory. *Space Weather*, *18*(8), e2020SW002513. https://doi.org/10.1029/2020SW002513

Embrechts, P., Klueppelberg, C., & Mikosch, T. (1997). *Modelling extremal events: For insurance and finance*. Springer. https://doi.org/10.1007/978-3-642-33483-2

Fawcett, L., & Walshaw, D. (2012). Estimating return levels from serially dependent extremes. *Environmetrics*, *23*(3), 272–283. https://doi.org/10.1002/env.2133

Gannon, J. L., & Love, J. J. (2011). USGS 1-min Dst index. *Journal of Atmospheric and Solar-Terrestrial Physics*, *73*(2–3), 323–334. https://doi.org/10.1016/j.jastp.2010.02.013

Gençay, R., Selçuk, F., & Ulugülyagci, A. (2001). EVIM: A software package for extreme value analysis in Matlab. *Studies in Nonlinear Dynamics & Econometrics*, *5*(3). https://doi.org/10.2202/1558-3708.1080

Gjerloev, J. W. (2012). The SuperMAG data processing technique. *Journal of Geophysical Research*, *117*(A9), A09213. https://doi.org/10.1029/2012JA017683

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/10.1029/93JA02867

Hapgood, M. (2019). The great storm of May 1921: An exemplar of a dangerous space weather event. *Space Weather*, *17*(7), 950–975. https://doi.org/10.1029/2019SW002195

Hapgood, M., Angling, M. J., Attrill, G., Bisi, M., Cannon, P. S., Dyer, C., et al. (2021). Development of space weather reasonable worst-case scenarios for the UK National Risk Assessment. *Space Weather*, *19*(4), e2020SW002593. https://doi.org/10.1029/2020SW002593

Horne, R. B., Glauert, S. A., Meredith, N. P., Boscher, D., Maget, V., Heynderickx, D., & Pitchford, D. (2013). Space weather impacts on satellites and forecasting the Earth's electron radiation belts with SPACECAST. *Space Weather*, *11*(4), 169–186. https://doi.org/10.1002/swe.20023

Iyemori, T., Takeda, M., Nose, M., Odagi, Y., & Toh, H. (2010). Mid-latitude geomagnetic indices "ASY" and "SYM" for 2009 (Provisional). In *Data analysis center for geomagnetism and space magnetism*. Graduate School of Science, Kyoto University.

Katus, R. M., & Liemohn, M. W. (2013). Similarities and differences in low-to middle-latitude geomagnetic indices. *Journal of Geophysical Research: Space Physics*, *118*(8), 5149–5156. https://doi.org/10.1002/jgra.50501

Koons, H. (2001). Statistical analysis of extreme values in space science. *Journal of Geophysical Research*, *106*(A6), 10915–10921. https://doi.org/10.1029/2000JA000234

Leadbetter, M. R., Lindgren, G., & Rootzén, H. (1983). *Extremes and related properties of random sequences and processes*. Springer. https://doi.org/10.1007/978-1-4612-5449-2

Liu, Y. D., Chen, C., & Zhao, X. (2020). Characteristics and importance of "ICME-in-sheath" phenomenon and upper limit for geomagnetic storm activity. *The Astrophysical Journal Letters*, *897*(1), L11. https://doi.org/10.3847/2041-8213/ab9d25

Love, J. J. (2020). Some experiments in extreme-value statistical modeling of magnetic superstorm intensities. *Space Weather*, *18*(1), e2019SW002255. https://doi.org/10.1029/2019sw002255

Love, J. J. (2021). Extreme-event magnetic storm probabilities derived from rank statistics of historical Dst intensities for solar cycles 14–24. *Space Weather*, *19*(4), e2020SW002579. https://doi.org/10.1029/2020SW002579

Love, J. J., Hayakawa, H., & Cliver, E. W. (2019). Intensity and impact of the New York railroad superstorm of May 1921. *Space Weather*, *17*(8), 1281–1292. https://doi.org/10.1029/2019SW002250

Love, J. J., Rigler, E. J., Pulkkinen, A., & Riley, P. (2015). On the lognormality of historical magnetic storm intensity statistics: Implications for extreme-event probabilities. *Geophysical Research Letters*, *42*(16), 6544–6553. https://doi.org/10.1002/2015GL064842

Mayaud, P. (1980). *Derivation, meaning, and use of geomagnetic indices* (Vol. 22, p. 607). Washington DC American Geophysical Union Geophysical Monograph Series. https://doi.org/10.1029/GM022

Morley, S. K. (2020). Challenges and opportunities in magnetospheric space weather prediction. *Space Weather*, *18*(3), e2018SW002108. https://doi.org/10.1029/2018SW002108

Newell, P., & Gjerloev, J. (2012). SuperMAG-based partial ring current indices. *Journal of Geophysical Research*, *117*(A5), A05215. https://doi.org/10.1029/2012JA017586

Nose, M., Iyemori, T., Sugiura, M., & Kamei, T. (2015). *Geomagnetic Dst index*. World Data Center for Geomagnetism. https://doi.org/10.17593/14515-74000

Ohtani, S. (2022). New insights from the 2003 Halloween storm into the Colaba 1600 nT magnetic depression during the 1859 Carrington storm. *Journal of Geophysical Research: Space Physics*, *127*(9), e2022JA030596. https://doi.org/10.1029/2022JA030596

Oughton, E. J., Skelton, A., Horne, R. B., Thomson, A. W. P., & Gaunt, C. T. (2017). Quantifying the daily economic impact of extreme space weather due to failure in electricity transmission infrastructure. *Space Weather*, *15*(1), 65–83. https://doi.org/10.1002/2016SW001491

Riley, P. (2012). On the probability of occurrence of extreme space weather events. *Space Weather*, *10*(2). https://doi.org/10.1029/2011SW000734

Riley, P., & Love, J. J. (2017). Extreme geomagnetic storms: Probabilistic forecasts and their uncertainties. *Space Weather*, *15*(1), 53–64. https://doi.org/10.1002/2016SW001470

Sandhu, J., Rae, I., & Walach, M.-T. (2021). Challenging the use of ring current indices during geomagnetic storms. *Journal of Geophysical Research: Space Physics*, *126*(2), e2020JA028423. https://doi.org/10.1029/2020JA028423

Silbergleit, V. M. (1999). Forecast of the most geomagnetically disturbed days. *Earth Planets and Space*, *51*(1), 19–22. 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*(25), 4782–4784. https://doi.org/10.1029/JA081i025p04782

Smith, R. L. (1989). Extreme value analysis of environmental time series: An application to trend detection in ground-level Ozone. *Statistical Science*, *4*(4), 367–377. https://doi.org/10.1214/ss/1177012400

Sugiura, M. (1964). Hourly values of equatorial Dst for the IGY. *Annals of the International Geophysical Year*, *35*(9).

Tsubouchi, K., & Omura, Y. (2007). Long-term occurrence probabilities of intense geomagnetic storm events. *Space Weather*, *5*(12), S12003. https://doi.org/10.1029/2007SW000329

Vasyliūnas, 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

Wanliss, J. A., & Showalter, K. M. (2006). High-resolution global storm index: Dst versus SYM-H. *Journal of Geophysical Research*, *111*(A2). A02202. https://doi.org/10.1029/2005JA011034

References From the Supporting Information

Hall, W. J., & Wellner, J. A. (1981). Mean residual life. In M. Csörgö, D. A. Dawson, J. N. Rao, & A. K. Saleh (Eds.), *Proceedings of the international symposium on statistics and related topics held in Ottawa, Canada, May 1980* (pp. 169–184). North-Holland Publishing Company.

Leadbetter, M. R. (1983). Extremes and local dependence in stationary sequences. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, *65*(2), 291–306. https://doi.org/10.1007/bf00532484

Moloney, N. R., Faranda, D., & Sato, Y. (2019). An overview of the extremal index. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *29*(2), 022101. https://doi.org/10.1063/1.5079656

Pickands, J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics*, 119–131. https://doi.org/10.1214/aos/1176343003

Smith, R. L., & Weissman, I. (1994). Estimating the extremal index. *Journal of the Royal Statistical Society: Series B*, *56*(3), 515–528. https://doi.org/10.1111/j.2517-6161.1994.tb01997.x