# Variation of geomagnetic index empirical distribution and burst statistics across successive solar cycles.

Aisling Bergin[1], Sandra C Chapman[1], Nicholas R. Moloney[2], and Nicholas Wynn Watkins[3]

[1]University of Warwick
[2]London Mathematical Laboratory
[3]LSE

November 22, 2022

**Abstract**

The overall level of solar activity, and space weather response at earth, varies within and between successive solar cycles and can be characterized by the statistics of bursts, that is, time-series excursions above a threshold. We consider non-overlapping 1 year samples of the auroral electrojet index (AE) and the SuperMAG-based ring current index (SMR), across the last four solar cycles. These indices respectively characterize high latitude and equatorial geomagnetic disturbances. We suggest that average burst duration $\tau$ and burst return period R form an activity parameter, $\tau/R$ which characterizes the fraction of time the magnetosphere spends, on average, in an active state for a given burst threshold. If the burst threshold takes a fixed value, $\tau/R$ for SMR tracks sunspot number, while $\tau/R$ for AE peaks in the solar cycle declining phase. Crossing theory directly relates $\tau/R$ to the observed index value cumulative distribution function (cdf). For burst thresholds at fixed quantiles, we find that the probability density functions of $\tau$ and R each collapse onto single empirical curves for AE at solar cycle minimum, maximum, and declining phase and for (-)SMR at solar maximum. Moreover, underlying empirical cdf tails of observed index values collapse onto common functional forms specific to each index and cycle phase when normalized to their first two moments. Together, these results offer operational support to quantifying space weather risk which requires understanding how return periods of events of a given size vary with solar cycle strength.

# Variation of geomagnetic index empirical distribution and burst statistics across successive solar cycles.

**A. Bergin[1], S. C. Chapman[1,2], N. R. Moloney[3], N. W. Watkins[1,4,5]**

[1]Centre for Fusion, Space and Astrophysics, Physics Department, University of Warwick, Coventry, UK
[2]Department of Mathematics and Statistics, University of Tromso, Norway
[3]Department of Physics, Imperial College London, London, UK
[4]Grantham Research Institute on Climate Change and the Environment, London School of Economics and Political Science, London, UK
[5]School of Engineering & Innovation, Faculty of Science, Technology, Engineering and Mathematics, The Open University, Milton Keynes, UK

## Key Points:

- For fixed value thresholds annual mean burst duration-return period ratio tracks solar cycle for $SMR$ but peaks near cycle decline for $AE$.
- Parameters of bursts in $AE$ and $SMR$ with thresholds at fixed quantile share the same distributions for successive solar cycle maxima.
- Tails of $AE$ and $SMR$ empirical distributions follow a solar cycle invariant functional form for solar cycle maximum, minimum and decline.

Corresponding author: Aisling Bergin, `aisling.bergin@warwick.ac.uk`

**Abstract**

The overall level of solar activity, and space weather response at earth, varies within and between successive solar cycles and can be characterized by the statistics of bursts, that is, time-series excursions above a threshold. We consider non-overlapping 1 year samples of the auroral electrojet index ($AE$) and the SuperMAG-based ring current index ($SMR$), across the last four solar cycles. These indices respectively characterize high latitude and equatorial geomagnetic disturbances. We suggest that average burst duration $\overline{\tau}$ and burst return period $\overline{R}$ form an activity parameter, $\overline{\tau}/\overline{R}$ which characterizes the fraction of time the magnetosphere spends, on average, in an active state for a given burst threshold. If the burst threshold takes a fixed value, $\overline{\tau}/\overline{R}$ for $SMR$ tracks sunspot number, while $\overline{\tau}/\overline{R}$ for $AE$ peaks in the solar cycle declining phase. Crossing theory directly relates $\overline{\tau}/\overline{R}$ to the observed index value cumulative distribution function (cdf). For burst thresholds at fixed quantiles, we find that the probability density functions of $\tau$ and $R$ each collapse onto single empirical curves for $AE$ at solar cycle minimum, maximum, and declining phase and for (-)$SMR$ at solar maximum. Moreover, underlying empirical cdf tails of observed index values collapse onto common functional forms specific to each index and cycle phase when normalized to their first two moments. Together, these results offer operational support to quantifying space weather risk which requires understanding how return periods of events of a given size vary with solar cycle strength.

## Plain Language Summary

Earth's magnetosphere and ionosphere has its own space weather. Space weather storms can cause technological problems including electrical grid damage and satellite system disruption. The overall driving of space weather follows the solar cycle of activity which has a period of approximately 11-years. Geomagnetic indices, based on magnetic field observations at the earth's surface, provide almost continuous monitoring of magnetospheric and ionospheric activity. We analyse two geomagnetic index time series, AE and SMR, which track activity in the auroral region and around the Earth's equator, respectively. We identify bursts or excursions above thresholds in the AE and SMR time series. We find that the ratio of average burst duration to return period provides a useful activity parameter which tracks the solar cycle in a well defined way. No two solar cycles are the same, each solar maximum has a different strength. However the distributions of the bursts, and the observations from which they are constructed, have properties that repeat from one solar cycle to the next. These results provide constraints that could be used in model predictions for the statistics of future space weather on solar cycle scales.

## 1 Introduction

Geomagnetic response at earth to solar driving is a problem of longstanding and topical interest (Baker, 2000; Milan et al., 2017; Pulkkinen, 2007), both in terms of understanding the underlying fundamental non-linear physics of the sun-earth system and improving space weather preparedness. Geomagnetic indices such as the auroral electrojet ($AE$) index (Davis & Sugiura, 1966) and disturbance storm-time ($Dst$) index (Sugiura, 1964) and their higher resolution counterparts (Gjerloev, 2012) are central to characterizing space weather activity and are available over the last five solar cycles.

Statistical studies of long-term geomagnetic indices are aimed both at fundamental understanding of the non-linear magnetospheric response to solar driving and quantification of space weather risk. The properties of burst distributions have been related to those characteristic of a wider class of complex systems (Barabási, 2010; Consolini, 1997; Chapman et al., 1998; Freeman & Watkins, 2002; Takalo, 1993; Watkins et al., 2016).

Space weather risk quantification has inspired statistical studies of the observed values of the geomagnetic indices, and of the return periods of bursts or events identified in these time-series. For example, Riley (2012) and Love et al. (2015) applied probabilistic analysis to $Dst$ to assess the probability of occurrence of extreme space weather events. Chapman et al. (2018) characterized space weather parameter distributions, including $AE$ and $Dst$, for the last five solar maxima and showed that the tail of the distributions follow a single functional form, independent of the strength of each solar cycle. Kakad and Kakad (2020) examined the probability distribution functions associated with $AE$ and $Dst$ indices and noticed significant narrowing in the probability distribution function for solar cycle 24 as compared to cycles 20–23, suggesting a decrease in the strength of associated current systems during this time. Extreme value theory (EVT) may be applied to model the index distributions. The $Dst$ index which is limited to 1-hour cadence and the 1-hour $AE$ timeseries are regularly used in such studies. For example, by applying the peak over threshold method to $Dst$ Tsubouchi and Omura (2007) estimate occurrence probabilities of intense geomagnetic storms and Acero et al. (2018) study an upper bound to the $Dst$ index distribution. Nakamura et al. (2015) utilize EVT to provide statistical evidence for finite upper limits to $AE$ indices and to estimate the annual expected number and probable intensity of extreme substorm events. Alberti et al. (2021) explored the complexity of the 1-minute $AE$ index and the $Dst$-like, 1-minute $SYM{-}H$ index and reported no difference between the last two solar cycles in terms of complexity measures for the two geomagnetic indices.

Geomagnetic activity shows a solar cycle dependence (Borovsky & Denton, 2006; Denton et al., 2006; Gonzalez et al., 1999; Hathaway, 2015; Richardson et al., 2002), driven by changes in the character of the solar-wind during the different phases of the solar cycle (Borovsky, 2020). Interplanetary coronal mass ejections (ICMEs) are associated with intense geomagnetic activity and are more frequent at solar maximum. Corotating interaction regions (CIRs) are most frequent during the declining phase of the solar cycle. CIR-driven storms are normally weaker than CME-driven storms, with less intensive auroral activity, but often have a longer duration of several days. Space-age geomagnetic indices, such as $AE$ and $Dst$, are only available for the last 5 solar cycles. Accessing a larger number of solar cycles relies on the $aa$ or $Ap$ 3-hourly range indices (Mayaud, 1980), however these time series are highly discretized by construction (Chapman, Horne, & Watkins, 2020) and thus cannot be straightforwardly thresholded to construct bursts as in our study here. Nevertheless they reveal solar cycle variation in activity. Lockwood et al. (2018) and (Lockwood et al., 2019) examined the $Ap$, $aa$, $Dst$ and $AE$ indices and found a monotonic relationship between the mean observed annual values and the fraction of a given year during which a 'large event' threshold was exceeded. Lockwood et al. (2018) found that annual distributions of $Ap$, $AE$ and $aa$ values follow a lognormal form which maintains a very constant shape over years of index availability. Haines et al. (2019) found that for geomagnetic storms, as measured by the $aa_H$ index, more intense storms have longer durations. Chapman, McIntosh, et al. (2020) found solar cycle ordering for extreme geomagnetic events in the $aa_H$ index. Elvidge (2020) applied EVT to the $aa$ index to estimate return levels for geomagnetic activity at times of solar maximum and solar minimum. Owens et al. (2021) use the $aa_H$ index to find that storms of all magnitudes occur more frequently during solar maximum than around solar minimum and that extreme events occur more frequently during large solar cycles than small cycles. Chapman, Horne, and Watkins (2020) found a good correspondence between annual minimum $Dst$ values and extreme activity in $aa$ which can be used to translate between the two indices on an annual timescale.

Counting events in order to construct empirical distributions can be problematic as geomagnetic storms can have multiple onsets and substorms are generally not isolated (Kamide et al., 1998). One approach is to define a threshold and investigate bursts, or excursions of the variable, above this threshold (e.g. Consolini, 1997; Freeman et al., 2000; Hush et al., 2015; Moloney & Davidsen, 2011, 2014; Tindale et al., 2018; Uritsky et al., 2001). In order to construct bursts from geomagnetic index time series we will therefore exploit the 1-minute resolution $AE$ index and the recently constructed high time resolution equatorial index, $SMR$, for the years 1975 - 2017. $SMR$ is produced using the SuperMAG collaboration of ground-based magnetometers (Gjerloev, 2012), is available at 1-minute cadence, and has been introduced as a high spatial resolution counterpart to the $Dst$ index. In this paper available $AE$ and $SMR$ geomagnetic index data are divided into 1-year non-overlapping intervals so that properties of bursts may be studied over the evolution of four solar cycles. We consider two definitions of bursts using (i) a fixed value threshold and (ii) a quantile threshold defined by the underlying cumulative distribution function (cdf) of the observations in each year-long sample. These burst definitions are sensitive to different aspects of the overall solar cycle variation of geomagnetic activity. A fixed value threshold directly identifies events exceeding a given amplitude and is sensitive to the overall rise and fall of solar cycle activity. A quantile threshold from the underlying observation cdf will itself rise and fall with the overall level of solar cycle activity, the threshold will resolve the change in behaviour of the underlying functional form of the distribution across different solar cycle phases. We will focus on how burst duration, $\tau$ (time spent above threshold), and return period, $R$ (time between threshold upcrossings), vary across the last four solar cycles. We remark that the return period defined in the threshold crossing problem for discrete time series considered here is different from the waiting time between events from time-dependent point processes, which has also been of continuing interest (e.g. Nurhan et al., 2021; Wheatland, 2000).

The observed *values* of geomagnetic indices track solar cycle activity. Indeed, we show here that quantiles of the $AE$ and $SMR$ index distributions track solar cycle variation over multiple solar cycles. The question is then how the likelihood of *events*, or *bursts*, varies with solar cycle activity. An identity from the theory of level crossings (Lawrance and Kottegoda (1977), see also Chapman et al. (2019)), states that for a given threshold, the dimensionless ratio of mean burst duration to mean burst return time ($\overline{\tau}/\overline{R}$) is directly related to the cdf of the underlying observations from which the bursts were constructed. We find that yearlong $AE$ and $SMR$ samples show clear ordering of $\overline{\tau}/\overline{R}$ with sunspot number (SSN), suggesting that $\overline{\tau}/\overline{R}$ is a useful 'activity parameter': it quantifies the fraction of time a geomagnetic index spends above a fixed value threshold. Taken separately, $\overline{R}$ and $\overline{\tau}$ show more complex behaviour.

Crossing theory gives the average return time for events that exceed a specific quantile for a given average duration. The minute-resolution indices provide a sufficiently large statistical sample that we can directly investigate if, and how, the full distributions of burst return time and burst duration track solar cycle variability, we find the results depend on solar cycle phase. Relating results for quantile thresholds to physical values requires knowledge of the underlying cdf of the index observations used to construct the bursts. Chapman et al. (2018) previously found that the observation cdf 'near-tail' functional form of 1-hour resolution $AE$ and $Dst$ does not vary from one maximum to the next, it is simply scaled by its mean and variance between weaker and stronger solar cycles. We recover this result here with higher time resolution 1-minute $AE$ and $SMR$. This suggests a robust dynamics of space weather 'climate' which does not predict when individual events will occur, but does suggest an overall activity level that floats up and down with the solar cycle level of activity. We suggest how this invariance provides a framework for translating predictions of future solar cycle activity into that of event return times.

The paper is organised as follows. Section 2 introduces the data used throughout this paper and describes how quantiles of the empirical distributions of annual index samples vary with solar cycle activity. Section 3 explains the methodology of burst construction and offers an informal proof of the aforementioned identity from crossing theory. Section 4 presents mean burst parameters calculated for the $AE$ and (-)$SMR$ indices across solar cycles 21-24 for bursts above fixed value and quantile thresholds. In Section 5, the full distribution of burst parameters are compared for fixed value and quantile thresholds at select phases of the solar cycle. Section 6 characterises the functional form for the near-tail region of underlying empirical distributions of $AE$ and (-)$SMR$ observations at maximum, minimum and declining phases of the solar cycle. Section 7 discusses the implications of these results for the study of space weather climatology. Finally, the findings of the paper are summarised in Section 8.

## 2 The $AE$ and $SMR$ indices and their Variation Over the Last Four Solar Cycles

Two key features of magnetospheric response to solar driving, as observed at the Earth's surface, are the enhancement of the ring current and auroral activity at high latitudes. The $AE$ index can be used as a good monitor of global energy deposition rates (Ahn et al., 1983). Indices such as $Dst$, $SYM-H$, and $SMR$, which respond to the horizontal component at the equator, can be interpreted as representing the energy of the suprathermal ions circulating about the Earth in the ring current (Newell & Gjerloev, 2012). As such, the auroral and equatorial indices parameterize the magnetic perturbations on the ground arising from distinct systems of magnetospheric and ionospheric current systems. It has been established that both time series exhibit irregular and bursty behaviour (Alberti et al., 2021). Understanding the underlying temporal changes in these indices can elucidate the behaviour of the overall magnetosphere - ionosphere system. Geomagnetic storms can last several days and substorms have timescales of a few hours but both are characterised by periods of rapid energy release (Sandhu et al., 2019, and references therein). High-time-resolution observations are required to resolve the bursty nature of the system and we require homogeneous, multi-solar cycle observations to investigate statistical variation within and between solar cycles. In this paper we will utilise the $AE$ and $SMR$ indices, as both are available at 1-minute resolution, almost continuously, for solar cycles 21-24.

### 2.1 Data Sets

The $AE$ index was introduced by Davis and Sugiura (1966) and is generated by the World Data Center for Geomagnetism, Kyoto (Nose et al., 2015). $AE$ is produced at 1-minute cadence using data from up to 12 ground based magnetometer stations at latitudes within the band of the auroral oval. Background-subtracted horizontal field (H) components of each station are compared. The upper index, $AU$, is the largest positive H-component disturbance and the lower index, $AL$, is the largest negative H-component disturbance. The $AE$ index is then the difference between upper and lower indices, $AE$ = $AU$ - $AL$. We use the final $AE$ index for the years 1975-1987 and the provisional $AE$ index for the years 1990-2017. There exist two gaps in the $AE$ data availability, 1976-1977 and 1988-1989. The SuperMAG auroral electrojet index, $SME$ (Newell & Gjerloev, 2011), uses data from more magnetometer stations than the official IAGA approved $AE$ indices to capture the electrojet behaviour more effectively. However, the number of magnetometer stations used to construct $SME$ varies over long (multiple solar cycle) time scales so that it is unsuitable for this cross-solar cycle study (Bergin et al., 2020).

The $SMR$ index is the ring current index compiled by the SuperMAG collaboration (Newell & Gjerloev, 2012). It is conceptually the same as the disturbance storm time ($Dst$) index (Sugiura, 1964) and $SYM-H$ index (Iyemori, 1990). $SMR$ is produced at 1-minute cadence. Low and midlatitude magnetometer stations in the SuperMAG collaboration provide H component data. Baseline removal is applied (Gjerloev, 2012), along with a correction factor for magnetic latitude. The stations are separated into four magnetic local time (MLT) zones, with centers at 00, 06, 12 and 18 MLT. The partial ring current index, $SMR_{00}$, is defined as the average corrected H component for all available stations within the 6-hr MLT zone centered at 00 MLT. Likewise for the other partial ring current indices, $SMR_{06}$, $SMR_{12}$ and $SMR_{18}$. $SMR$ is the average of the four partial ring current indices, $SMR = (SMR_{00} + SMR_{06} + SMR_{12} + SMR_{18})/4$. Space weather activity such as enhancement of the ring current results in a decrease in the $SMR$ index. Here, we plot minus the value of the index, (-)$SMR$, throughout for convenience.

We use the 13-month smoothed international SSN published by SILSO World Data Center (1973-2020) to characterise the solar activity cycle and we use the SILSO identification of dates for the minima and maxima of each individual solar cycle. In Sections 5 and 6, 1-year samples of $AE$ and (-)$SMR$ from minima, maxima and declining phases of solar cycles 21-24 are investigated. Samples around solar maximum are the calendar years 1979, 1989, 2001 and 2014 and around solar minimum are the years 1976, 1986, 1996 and 2008. Due to the aforementioned gap in $AE$ data availability for 1976 and 1989, we use $AE$ from the year 1975 for the solar cycle 21 minimum sample and 1990 for the sample of $AE$ at solar maximum of cycle 22. The calendar years 1983, 1993, 2004 and 2016 are selected to represent the declining phase of the solar cycles, following Chapman, Horne, and Watkins (2020).

### 2.2 Quantiles and the Cumulative Distribution Function

In Figure 1 we plot the distributions of $AE$ and (-)$SMR$ index observations for non-overlapping calendar year samples across the last four solar cycles. The cdf for the observed calendar year samples are calculated in two steps. (i) We first obtain the empirical cdf. For a given observation $x_k$, in a set of rank ordered observations $\{x_1 < x_2 < ... < x_k < ... < x_N\}$, taken from a time series, and where $N$ is the number of observations in the set and $k$ is the rank order, the corresponding value of the empirical cdf is $C(x_k) = \frac{k}{N}$. The $q_u$ quantile defines the value $u$ for which the observations $x > u$ exceed that of the proportion $q$ of the rank-ordered data set. In general, $C(u) = q_u$ is the fraction of observations for which $x \leq u$. (ii) We then use the kernel estimator of the cdf (Silverman, 1986) to resample the empirical distributions at 1 nT intervals. We present the cdf and the distribution quantiles of non-overlapping 1-year samples of the $AE$ index (Figure 1bi) and the (-)$SMR$ index (Figure 1bii). We compare the time variation of the distributions to the solar cycle variation exhibited by the SSN and the 13-month smoothed SSN for solar cycles 21-24. Overlaid in the plots are the dates of solar cycle minima/maxima.

We see that quantiles of the indices track solar cycle variation. For each $AE$ index quantile, in Figure 1bi, the peak in activity is seen in the declining phase of the solar cycle, after the solar maximum. The weak nature of cycle 24 is evident, particularly at the deep minimum in quantiles at 2008. We compare the variation of maximum $AE$ value within each calendar year sample to the sunspot cycle in Figure 1ci to demonstrate the extreme values that can be reached by $AE$. The 5000 nT peak associated with the 'great geomagnetic storm' at solar minimum in 1986 (Hamilton et al., 1988) is not reflected in the cdf variation. For each (-)$SMR$ index quantile, in Figure 1bii, the dual-peak solar cycle distribution described by Gonzalez et al. (1990) can be seen, where the dip is centred on solar cycle maximum.

## 3 Burst Statistics and Crossing Theory

An identity from crossing theory directly relates the distribution of the raw observations that constitute any given time series to the average properties of bursts in that time series. We will now explore how the solar cycle variation of the raw observations, discussed above, translates into the solar cycle variation of bursts, that is, events identified by threshold-crossing.

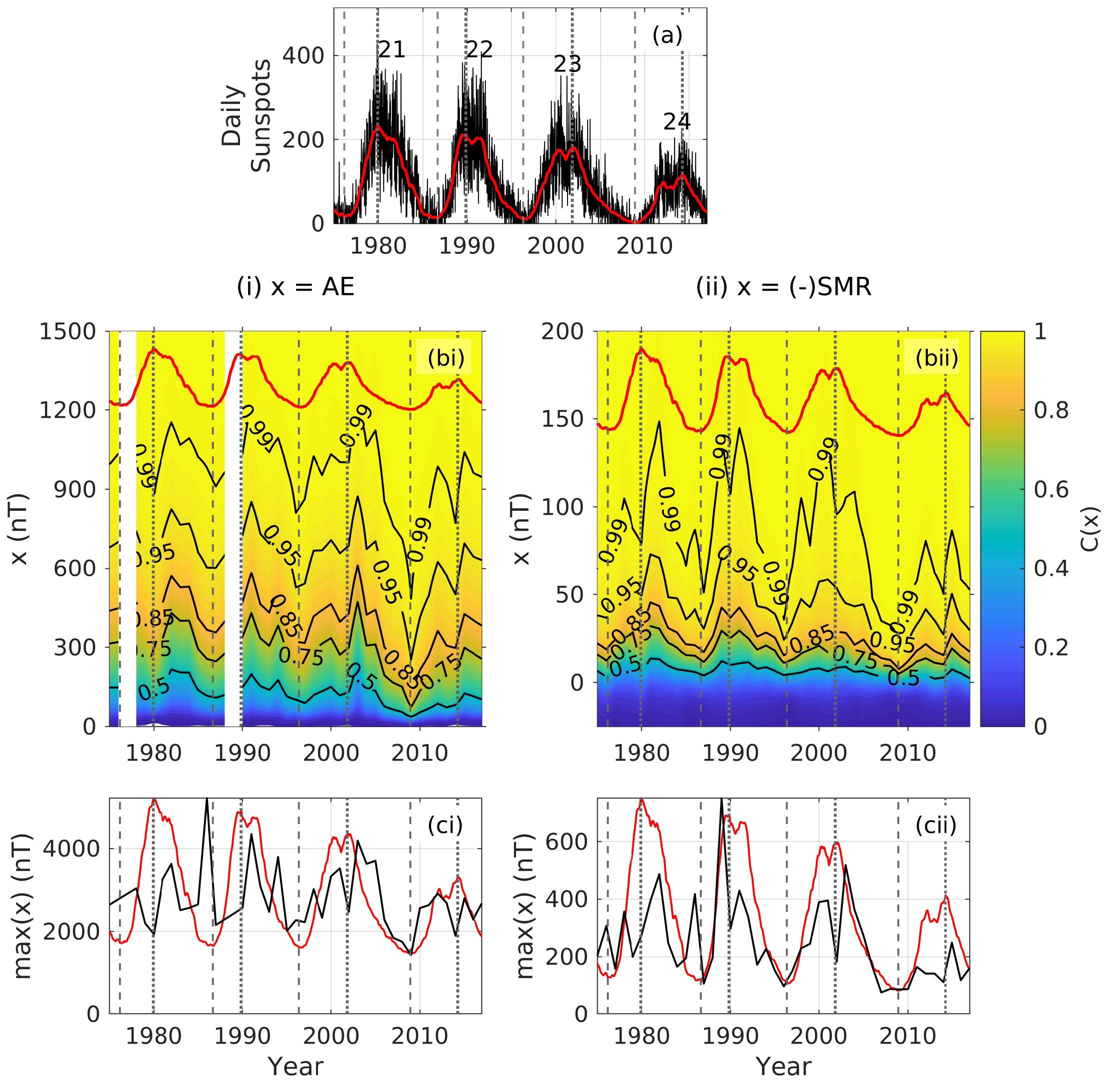


**Figure 1.** (a) Daily sunspot number (SSN) for the years 1975-2017 are plotted (black). Overplotted (red) is the 13-month smoothed monthly SSN. Labelled are solar cycles 21-24. The (bi) available Kyoto 1-minute auroral electrojet ($AE$) index time series and (bii) (-) SuperMAG 1-minute ring current index ($SMR$) from 1975-2017 are shown as a time-variation in distribution. The cumulative distribution function (cdf) values, for each non-overlapping 1 year sample, are indicated by color and are plotted versus index value (y-axis) and time (x-axis). Quantiles are indicated in black solid lines on the cdfs. The maximum (ci) $AE$ and (cii) (-)$SMR$ value in each non-overlapping 1 year sample. In all panels, grey dashed lines indicate solar minimum and dotted lines indicate solar maximum, identified from the monthly smoothed SSN. Overplotted in (b,c) is the transformed 13-month smoothed monthly SSN (red).

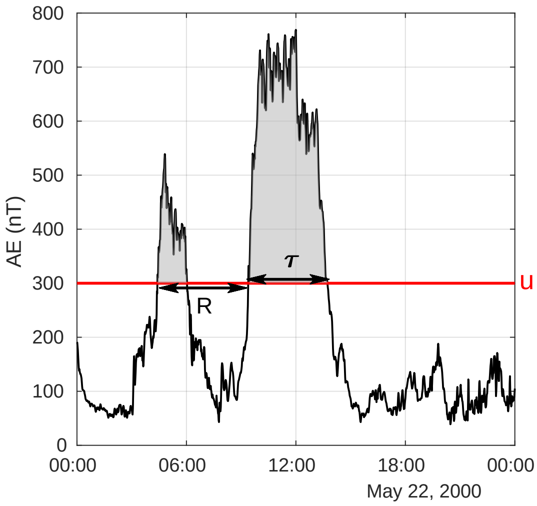


**Figure 2.** Bursts in 24 hours of the $AE$ time series on 22nd May 2000, are plotted (black) with an example threshold of u = 300 nT in red. Definitions of relevant burst parameters are labelled. Burst return period, R, is the time between subsequent threshold upcrossings. Burst duration, $\tau$, is the time between threshold upcrossing and subsequent downcrossing. Burst size is the integrated area between the time series and the threshold and is indicated here by grey shading.

### 3.1 Identifying Bursts in the Time Series

A burst in a timeseries is defined as an excursion above a threshold, $u$. The burst return period, $R$, is the time between consecutive threshold upcrossings, the burst duration, $\tau$, is the time between threshold upcrossing and downcrossing, and size is the integrated area while the timeseries exceeds the threshold, as illustrated in Figure 2. We will consider two different definitions of the burst threshold. Bursts may be identified as the data interval above a fixed value, sample-invariant threshold which is constant across the solar cycles, for example, $u = 500$ nT. Alternatively, bursts may be identified as the data interval above a sample-specific quantile threshold, $q_u$. For any calendar year sample of observations, the time-varying threshold $u(t)$, is defined by the $q_u$th quantile of the sample, for example given $q_u = 0.85$, the burst threshold $u(t)$ varies in time such that it is always at a value that exceeds that of 85% of the rank ordered sample. Figure 1 then plots the observed index values that correspond to a given quantile threshold.

Bursts of duration $\tau$ that approach the time-resolution of the data, that is, $\tau$ shorter than a $T_{R1}$ of 5 minutes, are excluded from this analysis. Likewise, consecutive pairs of bursts separated by R shorter than $T_{R2}$ of 5 minutes are treated as a single burst. We have varied $T_{R1}$ and $T_{R2}$ independently ($T_{R1} \neq T_{R2}$) and find that the number of bursts monotonically decreases with increasing $T_{R1}$ or $T_{R2}$ and their annual means, $\overline{\tau}$ and $\overline{R}$, vary in step but importantly the ratio of $\overline{\tau}$ to $\overline{R}$ does not change within the 95% confidence interval estimate. As a sample, for a threshold of $u = 500$ nT applied to AE where $T_{R1} = 5$ minutes, on average 300 of 1300 fluctuations above $u$ fall into the $\tau < T_{R1}$ category. For a threshold of $u = 40$ nT applied to (-)$SMR$ where $T_{R1} = 5$ minutes, on average 40 of 250 fluctuations fall into this category.

### 3.2 Mean Burst Duration and Return Period from Crossing Theory

The properties of bursts identified in the observed discrete time series are constrained by their empirical cdfs via an identity from crossing theory (Cramér & Leadbetter, 2004;

Lawrance & Kottegoda, 1977; Vanmarcke, 2010). Equations 1 to 4 offer an informal proof of this identity, following the aforementioned references. Let $\{x_k\}_{k=1}^N$ be the time-indexed observations in a time series sample, where $N$ is the number of observations in the sample. For a threshold $u$, the mean burst duration is

$$\overline{\tau}(u) = \frac{\#observations > u}{B(u)} = \frac{\sum_{k=1}^{N} \mathbb{1}(x_k > u)}{B(u)} \tag{1}$$

where $B(u)$ is the number of bursts above the threshold $u$ and $\mathbb{1}$ is an indicator function such that

$$\mathbb{1}(x_k > u) = \begin{cases} 1, & \text{if } x_k > u \\ 0, & \text{if } x_k \leq u. \end{cases} \tag{2}$$

The mean burst return period is

$$\overline{R}(u) = \frac{N}{B(u)}. \tag{3}$$

So the ratio of mean burst duration to mean burst return period is

$$\frac{\overline{\tau}(u)}{\overline{R}(u)} = \frac{\sum_{k=1}^{N} \mathbb{1}(x_k > u)}{N} = \frac{N - \sum_{k=1}^{N} \mathbb{1}(x_k \leq u)}{N} = 1 - \frac{1}{N}\sum_{k=1}^{N} \mathbb{1}(x_k \leq u) = 1 - C(u) \tag{4}$$

where $C(u)$ is the value of the cdf evaluated at the threshold $u$.

Thus mean burst duration and return time are not independent quantities. Together they form a dimensionless ‘activity parameter’ which describes the fraction of time the magnetosphere spends, on average, in an active state. It will be larger when events last longer and/or occur more frequently, signifying enhanced geomagnetic activity. Many short duration bursts will have the same value of the activity parameter as a few, long duration bursts so that we may determine the average event duration for a given average return period, or vice versa. For a fixed threshold value $u$, the activity parameter will track the solar cycle variation in $C(u)$ that we have seen in Figure 1. For a threshold at a fixed quantile, that is, fixed value of $C(u)$, the activity parameter should not vary and we will use this result from crossing theory as a check on the fidelity of our distributions in Section 4.

## 4 Mean Burst Return Periods and Durations in $AE$ and $SMR$

We can directly obtain the distribution of burst parameters from the geomagnetic index timeseries and hence can independently determine the mean burst duration ($\overline{\tau}$) and mean return period ($\overline{R}$) for non-overlapping consecutive calendar year samples of the $AE$ index and the (-)$SMR$ index across solar cycles 21-24. We can then see how these contribute to the ratio $\overline{\tau}/\overline{R}$ across multiple solar cycles. Figure 3 plots $\overline{\tau}$ and $\overline{R}$ for $AE$ bursts identified above a fixed value threshold of 500, 700, and 900 nT in each annual sample. We would expect larger events to have longer duration on average, and indeed the mean burst duration $\overline{\tau}$, in Figure 3a, tends to be smallest around solar minimum and is largest around solar maximum. However, except for the quiet cycle 24, $\overline{\tau}$ peaks after each solar maximum SSN peak. Since events are more frequent during the active phase of the solar cycle we would expect the mean return period $\overline{R}$ to be roughly anti-correlated with the SSN. $\overline{R}$, plotted in Figure 3b and again in Figure 3c with a zoom on the y-axis, does indeed peak at the solar minima, however, except for cycle 24, the average return period is shortest after the SSN peak at maximum, in the declining phase of the solar cycle. In addition to these overall trends, the detailed behaviour varies from one cycle to the next. For example, $\overline{R}$ has an additional peak at 1980 and at 2015 near the solar maxima. The minimum before the notably weak cycle 24 was particularly quiet relative to previous minima, it has a relatively low SSN, and we can see that it has a very long $\overline{R}$. The highest threshold $\overline{\tau}$ does not have a minimum at the cycle 21 minimum.

The ratio of average duration to return period ($\overline{\tau}/\overline{R}$), that is, the activity parameter, is plotted in Figure 3d. Figure 3d shows that $\overline{\tau}/\overline{R}$ exhibits the same pattern of variation through each of the four solar cycles, with minima coinciding with each SSN solar minimum and peaking in the declining phase for all four cycles. This suggests that $\overline{\tau}/\overline{R}$ shows a clearer ordering with the SSN over solar cycle scales than $\overline{\tau}$ or $\overline{R}$ independently. Both $\overline{\tau}$ and $\overline{R}$ contain information about the details of the burst events, but they are not independent of each other as they are constrained by crossing theory (Equation 4), that is they are constrained by the overall activity level of the system which is characterized by $\overline{\tau}/\overline{R}$.

Figure 4 is in the same format Figure 3 except that it now plots $\overline{\tau}$ and $\overline{R}$ for $AE$ bursts identified above quantile thresholds at the 0.85, 0.95 and 0.99 quantile of each annual sample. The observed values of $AE$ to which quantile thresholds correspond are now tracking the overall level of solar cycle activity, as can be seen from Figure 1bi. We now do not see strong solar cycle ordering in the variation of $\overline{\tau}$ in Figure 4a or $\overline{R}$, in Figure 4b and 4c. On these plots, the behaviour at cycle 24 is similar in amplitude to that of the previous cycles, suggesting that once the overall solar cycle activity level is removed, the events in weak cycle 24 are not behaving differently from those in previous cycles. The activity parameter $\overline{\tau}/\overline{R}$ is constrained to be a constant by the crossing theory (Equation 4); Figure 3d confirms that this is indeed the case for our empirically determined $\overline{\tau}/\overline{R}$, confirming the accuracy of our quantile estimation.

Figures 5 and 6 repeat the above analysis for $SMR$. Figure 5 plots $\overline{\tau}$ and $\overline{R}$ for (-)$SMR$ bursts identified above a fixed value threshold of 40, 60 and 100 nT in annual samples. Although $\overline{\tau}$ in Figure 5 is larger at the maxima, the signal shows high variability; it does not show robust ordering with the SSN for all four solar cycles. Peaks in $\overline{R}$ are found at the solar cycle minima and the shortest $\overline{R}$ values are observed either in the declining phase or at the maxima of the solar cycle, most evident for 40 nT and 60 nT thresholds in Figure 5c. The activity parameter $\overline{\tau}/\overline{R}$ in Figure 5d shows a rather clearer signal of variation with the solar cycle of activity including (except for cycle 23) tracking the double peak in the SSN. Figure 6 plots $\overline{\tau}$ and $\overline{R}$ for (-)$SMR$ bursts identified above quantile thresholds at the 0.9, 0.95 and 0.99 quantiles in each annual sample. As with $AE$, we see that using a fixed quantile burst threshold eliminates much of the solar cycle variation.

The $AE$ and $SMR$ indices respond to different magnetospheric and ionospheric current systems and events are characterized by different signatures in the timeseries. Significant disturbances in the auroral electrojets result in rapid sporadic signatures in the $AE$ index whereas the ring current recovers more gradually from large disturbances (Milan et al., 2017). This is reflected in differences in the statistical sampling of the mean return time and duration of bursts constructed with quantile thresholds of the $AE$ and (-)$SMR$ indices. Equation 3 defines $\overline{R}$ for a given threshold as the number of observations in a sample divided by the number of bursts over that threshold. We consider 1-minute resolution, 1-calendar year samples of $AE$ and (-)$SMR$ so that the number of observed values of the indices from which the bursts are constructed is essentially the same across all samples and $T/\overline{R}$, where $T$ is a calendar year, is the time-varying number of bursts in each sample. For example, we applied a quantile threshold of 0.95 to both indices. Where we consider the the $AE$ index in Figure 4, $\overline{R}$ of approximately 10 hours corresponds to approximately 900 bursts in the sample. $\overline{\tau}$ varies between 30 minutes and 1 hour. On the other hand, when the 0.95 quantile threshold is applied to the (-)$SMR$ index, seen in Figure 6, $\overline{R}$ fluctuates around 50 hours, equating to approximately 200 bursts. $\overline{\tau}$ varies between 2 and 4 hours. Bursts in the (-)$SMR$ index are longer in duration so lead to more recorded amplified values per event, thus the quantile threshold picks out fewer events.

To summarize, crossing theory does not specify how $\overline{\tau}$ and $\overline{R}$ vary independently. Whilst $\overline{\tau}$ and $\overline{R}$ show some overall solar cycle trends we find there is considerable variation in how they track each cycle in detail. However, from the crossing theory identity

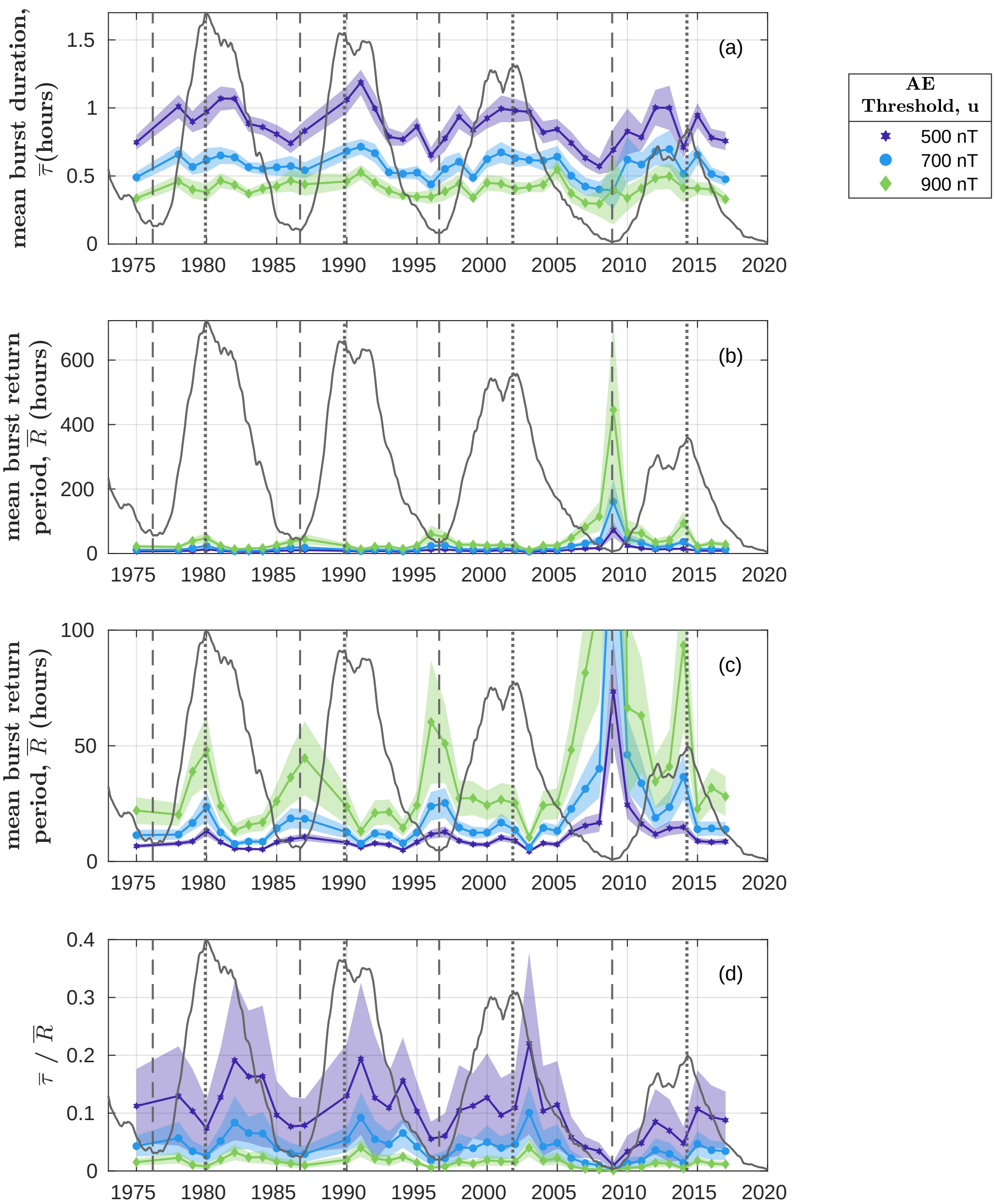


**Figure 3.** *AE* mean burst parameters at fixed value thresholds. Burst analysis of the time series is used to plot the mean (a) burst duration, (b,c) burst return period, and (d) ratio of duration to return period for bursts in the *AE* time series over a threshold of 500 nT (purple), 700 nT (blue) and 900 nT (green). Data is sampled in non-overlapping 1-year periods. Shading indicates 95% confidence intervals for the mean, given by $\bar{x} \pm (1.96 \times \sigma(x)/\sqrt{B})$ where $x$ is duration or return period of bursts in the 1-year sample, $\bar{x}$ is their mean, $\sigma(x)$ is their standard deviation and B is the number of bursts recorded. The 13-month smoothed monthly sunspot number (SSN) (solid grey), rescaled to the y-axes is plotted. Grey dashed lines indicate solar minima and dotted lines indicate solar maxima, identified from the monthly smoothed SSN.

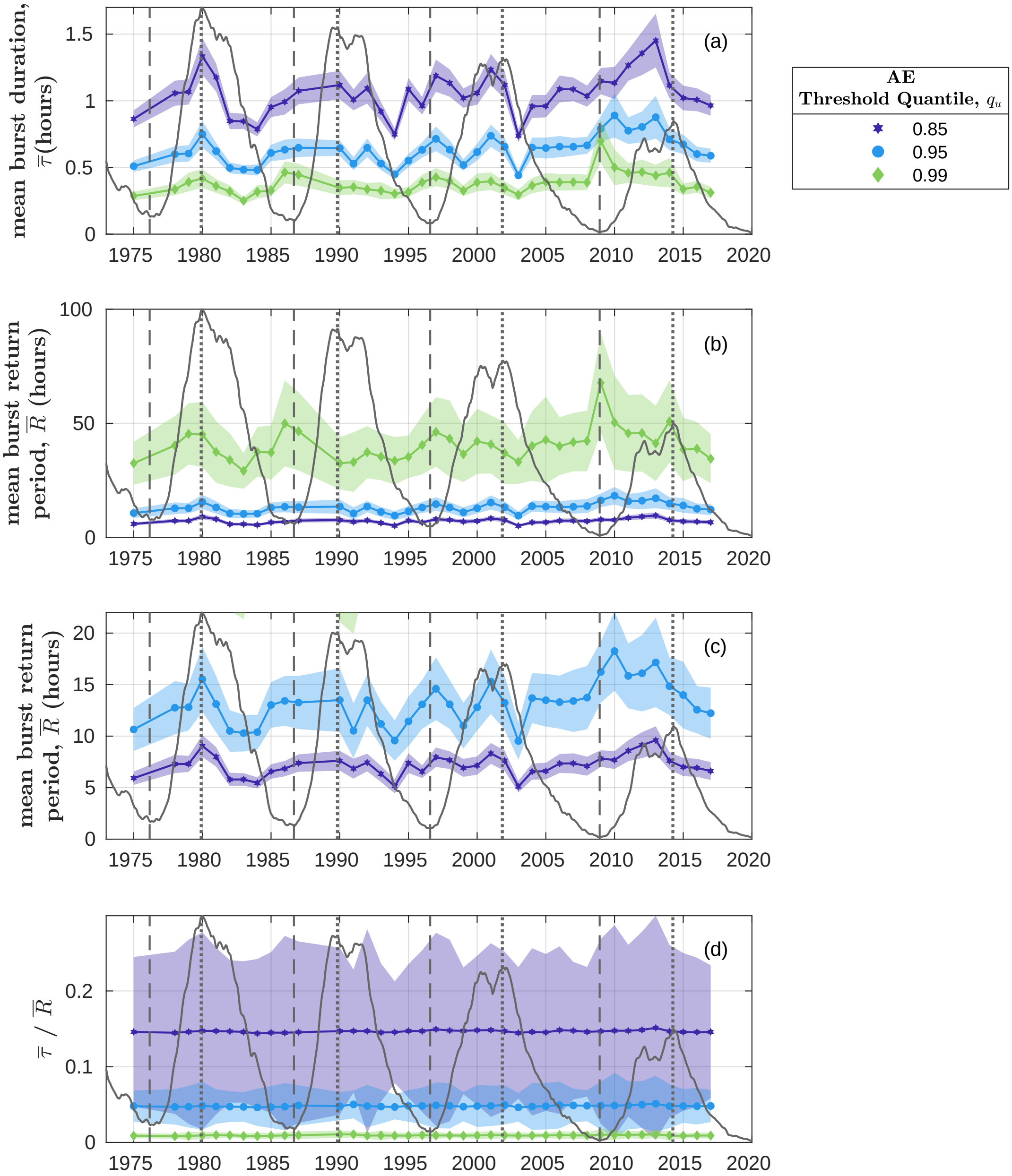


**Figure 4.** *AE* mean burst parameters at quantile thresholds. Burst analysis of the time series is used to plot the mean (a) burst duration, (b,c) burst return period, and (d) ratio of duration to return period for bursts in the *AE* time series over a quantile threshold of 0.85 (purple), 0.95 (blue) and 0.99 (green). Data is sampled in non-overlapping 1-year periods. Format as in Figure 3.

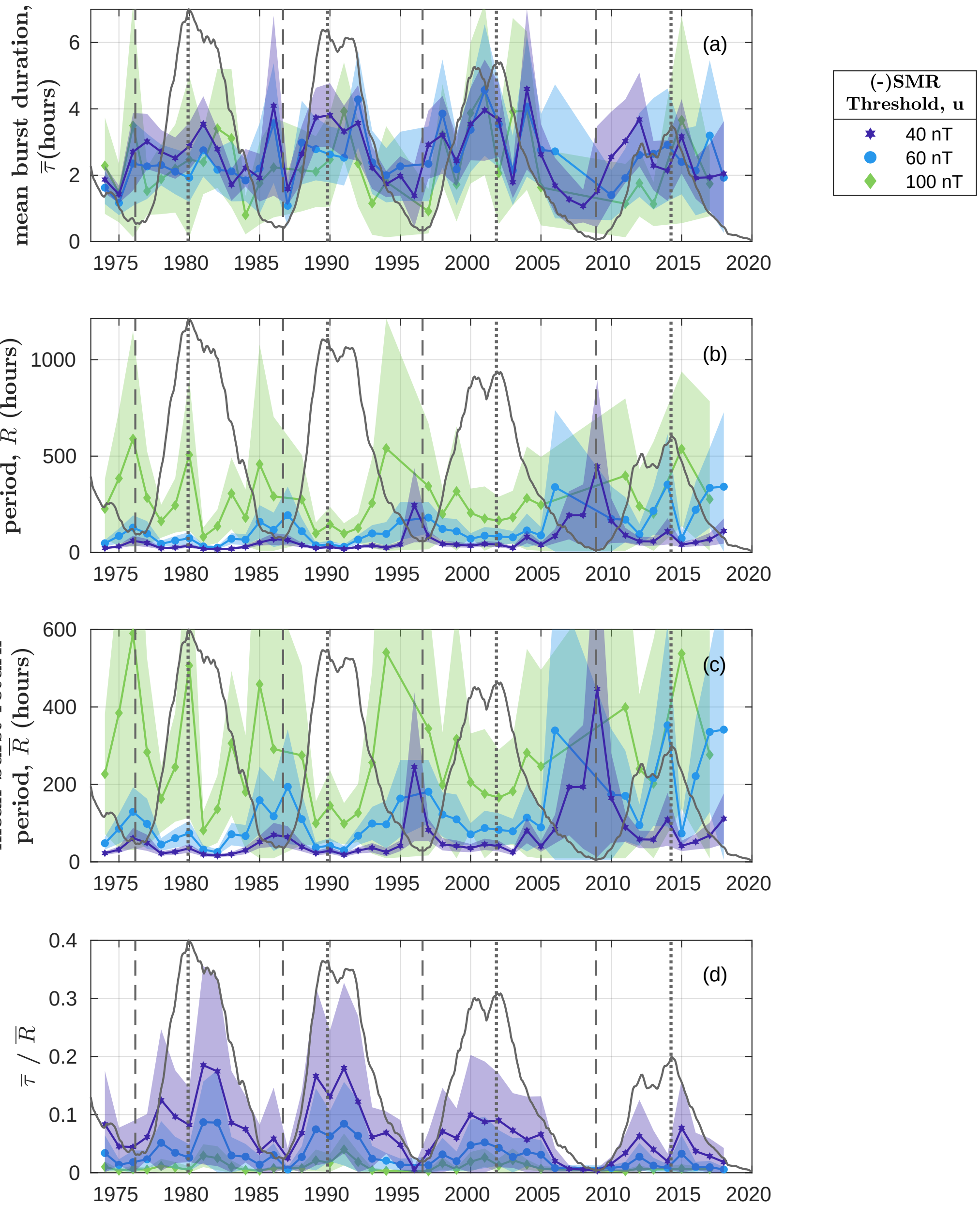


**Figure 5.** (-)$SMR$ mean burst parameters at fixed value thresholds. Burst analysis of the time series is used to plot the mean (a) burst duration, (b,c) burst return period, and (d) ratio of duration to return period for bursts in the (-)$SMR$ time series over threshold of 40 nT (purple), 60 nT (blue) and 100 nT (green). Data is sampled in non-overlapping 1-year periods. Format as in Figure 3.

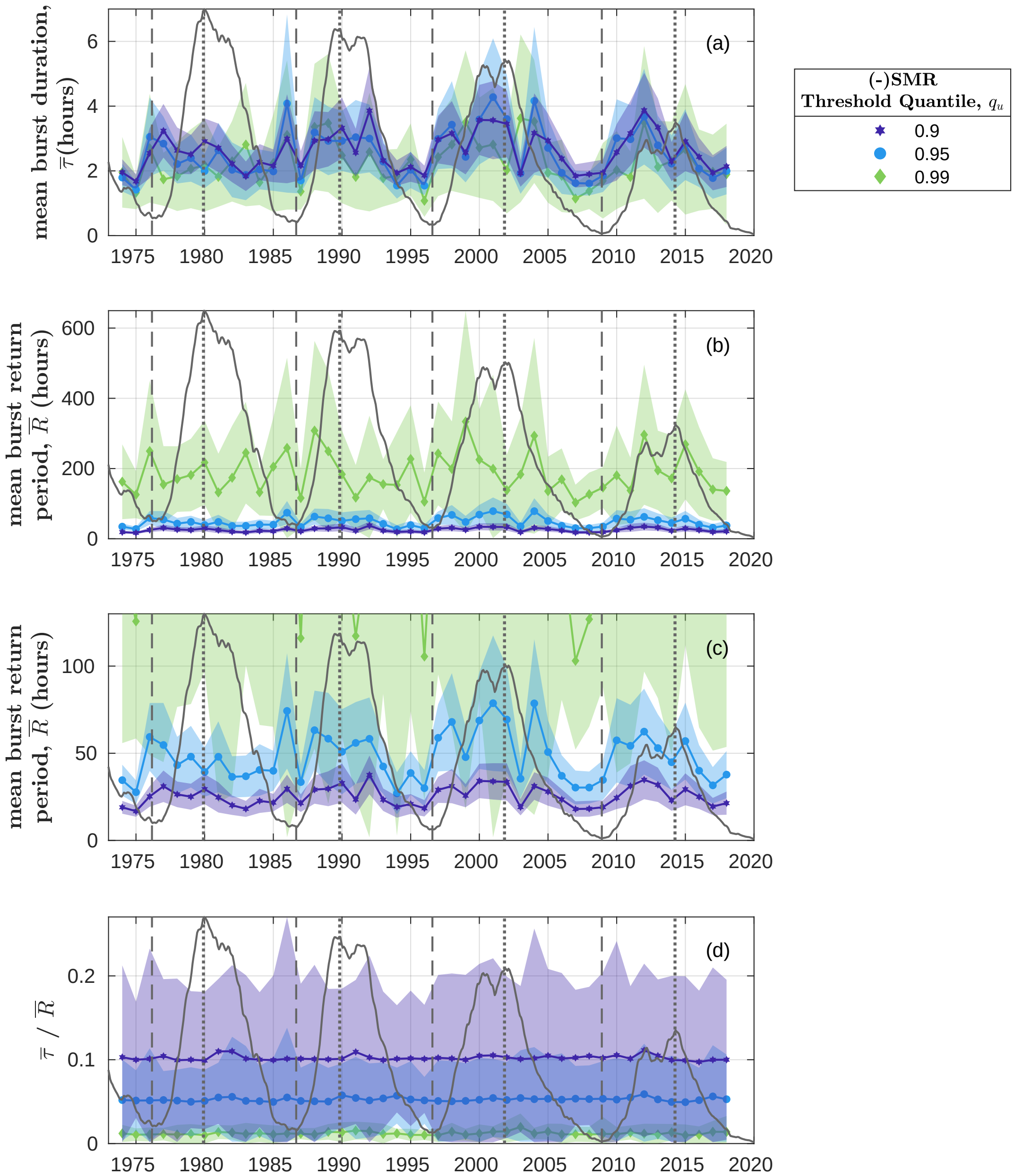


**Figure 6.** (-)$SMR$ mean burst parameters at quantile thresholds. Burst analysis of the time series is used to plot the mean (a) burst duration, (b,c) burst return period, and (d) ratio of duration to return period for bursts in the (-)$SMR$ time series over quantile threshold of 0.9 (purple), 0.95 (blue) and 0.99 (green). Data is sampled in non-overlapping 1-year periods. Format as in Figure 3.

the ratio $\overline{\tau}/\overline{R}$ must track the value of the underlying cdf, $C(u)$ or quantile of the observed *values* of the index. If we construct bursts with a fixed quantile threshold then $\overline{\tau}/\overline{R}$ is constant. If on the other hand, we construct bursts with a fixed threshold value $u$ then $C(u)$ will vary with solar cycle activity level. In Figure 1, we see that the quantiles of the observed index values track SSN quite well. This then constrains the ratio $\overline{\tau}/\overline{R}$ to also track the level of solar cycle activity, for bursts constructed with a fixed value threshold. $\overline{\tau}/\overline{R}$ may thus provide an overall activity parameter that relates changes in overall solar driving, the amplitude of the observed index, and the statistics of the burst parameters.

## 5 Full Distributions of Burst Parameters

In Section 4 we found that the variation in $\overline{R}$, and to a lesser extent $\overline{\tau}$, from one solar maximum to the next that is seen when we threshold bursts at a fixed value, is suppressed when we threshold at a quantile. We now compare the full distribution of burst parameters, not just the mean, for fixed value and quantile thresholds.

We select one year at each solar cycle maximum to compare across the four solar cycles, as described in Section 2.1. We obtain the full probability density functions (pdfs) of burst durations and burst return periods, these are shown in Figure 7 for $AE$ and Figure 8 for (-)$SMR$. Bin widths are determined using the Freedman-Diaconis rule (Freedman & Diaconis, 1981) and the uncertainties are calculated as the square root of the bin count. A log scale is applied to the $x$-axes. We compare the distributions of burst parameters obtained by thresholding at a fixed quantile, and at a fixed value of the index timeseries.

In Figure 7a a fixed value threshold at 500 nT is applied to the $AE$ timeseries and in Figure 7b we use a sample-specific quantile threshold at the 85th quantile. In Figure 8 we present the distribution of return periods and duration of bursts in (-)$SMR$ above (a) fixed value threshold of 40 nT and (b) quantile threshold at the 90th quantile. The distribution of burst return periods and duration clearly vary from one solar maximum to the next when the bursts are constructed using a fixed value threshold, for both $AE$ and (-)$SMR$. However, when the sample-specific quantile threshold is used, the resulting burst distributions show little variation between one solar maximum and the next and in some cases effectively collapse onto each another, within uncertainties.

The variation from one solar maximum to the next in $\overline{R}$, and to a lesser extent $\overline{\tau}$, that is seen when we threshold at a value, is suppressed when we threshold at a quantile, as shown in Section 4. We now have the much stronger result, that this approximately holds for the full distributions of $\tau$ and $R$. Provided future solar cycles behave as in the past, this has the potential to constrain predictions of space weather activity. Given a prediction of the value of the quantile of the observed values of the index, then the distribution of duration and return period of bursts above this quantile may be known, since it simply follows the empirical distributions of burst parameters plotted here. We have attempted single functional form fits to these distributions but no robust functional form was found. We repeated this analysis for samples during minimum and declining phases of the solar cycle. In each case, where a quantile threshold is applied to the $AE$ timeseries sample, the full distributions of $\tau$ and $R$ vary only weakly from one cycle minimum or declining phase to the next. In the case of (-)$SMR$, where bursts are identified above a quantile threshold, distributions of $\tau$ and $R$ show more cycle-to-cycle variability than those found for solar cycle maximum, in Figure 8.

## 6 Rescaling Properties of the Distributions of Observations

The crossing theory identity (Equation 4) relates the underlying distribution of observed values to the ratio of the average duration and return times of bursts obtained by thresholding the timeseries, $\overline{\tau}/\overline{R}$. In this section we focus on how the underlying dis-

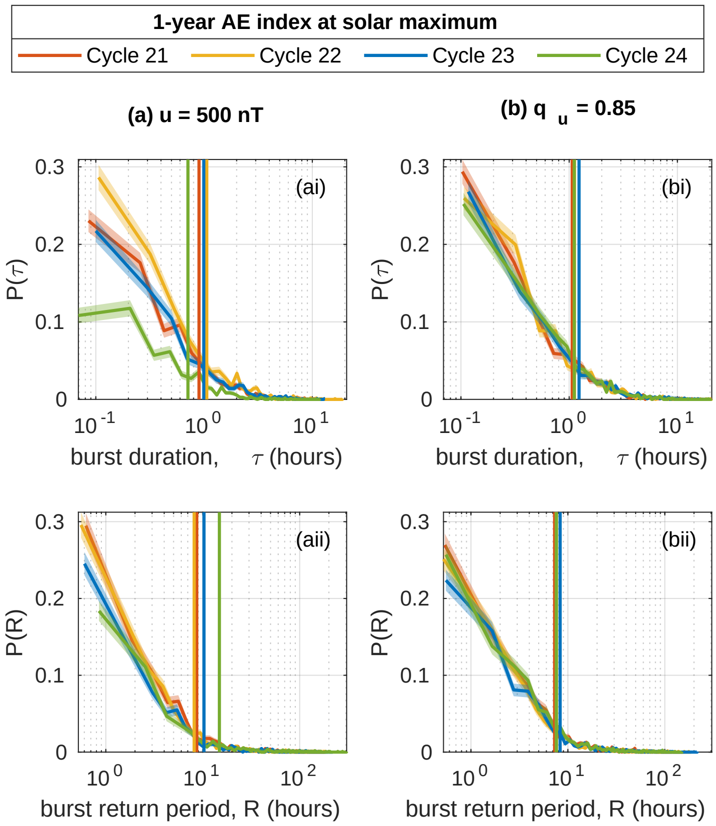


**Figure 7.** Bursts are identified above (a) fixed value threshold of 500 nT and (b) quantile threshold at $q_u = 0.85$ in 1-year samples of the $AE$ index at the maxima of solar cycles 21 (red), 22 (yellow), 23 (blue) and 24 (green). Probability density function of burst (i) return periods, R, and (ii) duration, $\tau$, are plotted on a log x-axis scale. Uncertainties are calculated as the square root of the bin count and are indicated by shading. Overplotted is the distribution mean (vertical line).

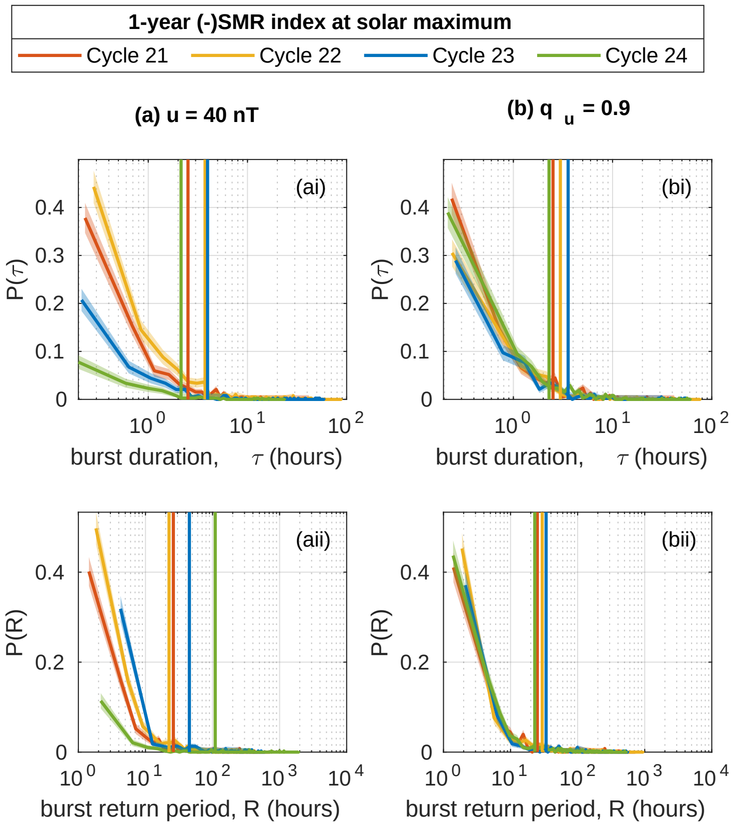


**Figure 8.** Bursts are identified above (a) fixed value threshold of 40 nT and (b) quantile threshold at $q_u = 0.9$ in 1-year samples of the (-)$SMR$ index at the maxima of solar cycles 21 (red), 22 (yellow), 23 (blue) and 24 (green). Probability density function of burst (i) return periods, R, and (ii) duration, $\tau$, are plotted on a log x-axis scale. Uncertainties are calculated as the square root of the bin count and are indicated by shading. Overplotted is the distribution mean (vertical line).

tribution of observed values varies across multiple solar cycles. Chapman et al. (2018) found that the cdf of observations of a variety of parameters that track solar wind - coupling, including OMNI 1-hour $AE$ index and 1-hour (-)$Dst$ index, share the same functional form for large to extreme observations in the near-tail region of the distribution, across successive solar maxima. In this section we apply the same analysis to the Kyoto 1-minute $AE$ index and SuperMAG 1-minute (-)$SMR$ index.

We plot the the survival distribution function (sdf) of $AE$ index observations for 1-year samples at solar cycle maximum. $S(x) = 1 - C(x)$ where $x$ is a set of timeseries observations, $S(x)$ is the sdf and $C(x)$ is the empirical cdf. Uncertainties in the sdfs are estimated using Greenwood's formula (Greenwood, 1926) and are indicated by shaded region in these figures. The tails of the index empirical distributions are identified as exceedences of a quantile threshold, $q_E$. The entire 1-year distribution is rescaled by the mean, $\mu$, and standard deviation, $\sigma$, of the sample observations which exceed $q_E$. The method of determination of the $q_E$ threshold quantile is discussed in detail by Chapman et al. (2018) but in summary, data-data quantile-quantile plots are used to identify two components in the distribution, one relating to the relatively quiet intervals of the timeseries and another during large bursts or storms. This threshold quantile is found at $q_E = 0.75$ for the $AE$ index and $q_E = 0.9$ for the (-)$SMR$ index. The Generalized Pareto Distribution (GPD), commonly used in extreme value theory to characterize the statistics of extreme observations, is here applied as a flexible fitting distribution. The GPD survival function is

$$S(x) = \left(1 + \xi\left(\frac{x-u}{\phi}\right)\right)^{-\frac{1}{\xi}} \tag{5}$$

with threshold parameter, $u$, shape parameter, $\xi$ and scale parameter, $\phi$.

In Figure 9ai, we plot the sdfs of 1-year $AE$ index samples from the period of solar maximum of solar cycles 21-24. In Figure 9aii we show these distributions, rescaled by $\mu$ and $\sigma$ of exceedences of the 0.75 quantile. We see that between the 0.75 quantile and the 0.999 quantile, the distributions lie over one another, within uncertainties. In Figure 9aii it is shown that each individual sdf tail may be fit with a generalized Pareto distribution function that is a good fit up until the 0.999 quantile. Figure 9aiii shows how the four rescaled cycle observation samples may be combined to form an aggregate, where the tail of that aggregate distribution may be fit by one GPD master distribution, at least until the 0.999 quantile. Figure 9b shows the sdfs of 1-year (-)$SMR$ samples from solar maxima 21-24. It is seen that the GPD may be fit as a master distribution to the tail of the distribution between the 0.9 and 0.999 quantile where the distributions have been rescaled by the mean and variance of exceedences of the 0.9 quantile.

We see a roll-off in the tail of the $AE$ and particularly the (-)$SMR$ index observations above the 0.999 quantile where the curves no longer collapse onto a single functional form. These are the most extreme values observed during the 1-year timeseries samples and amount to approximately 10 hours of observations total. In the case of $AE$, these correspond to approximately 100 upcrossings of the 0.999 threshold in the timeseries, that is multiple short duration, high-intensity events. For reference there are approximately 3000 upcrossings of $q_E = 0.75$ in each timeseries sample. In each (-)$SMR$ timeseries samples there are approximately 1000 upcrossings of the $q_E = 0.9$ threshold but only 30 upcrossings of the 0.999 quantile threshold, relating to events of long duration relative to $AE$.

We repeat this analysis for $AE$ and (-)$SMR$ observation samples from solar minimum and the declining phase. We find that, up to the 0.999 quantile, the rescaled sdfs may be fit by master GPD distributions within uncertainties, the parameters of which are tabulated in Table 1. The GPD shape parameter, $\xi$, is indicative of the Fisher-Tippett subclass to which the distribution belongs (Embrechts et al., 2013). We see that for solar minimum, maximum and declining phase the $AE$ index distribution tails are close to the Gumbel distribution class ($\xi$=0) and approach the exponential case ($\xi$=0, $\phi$=1).

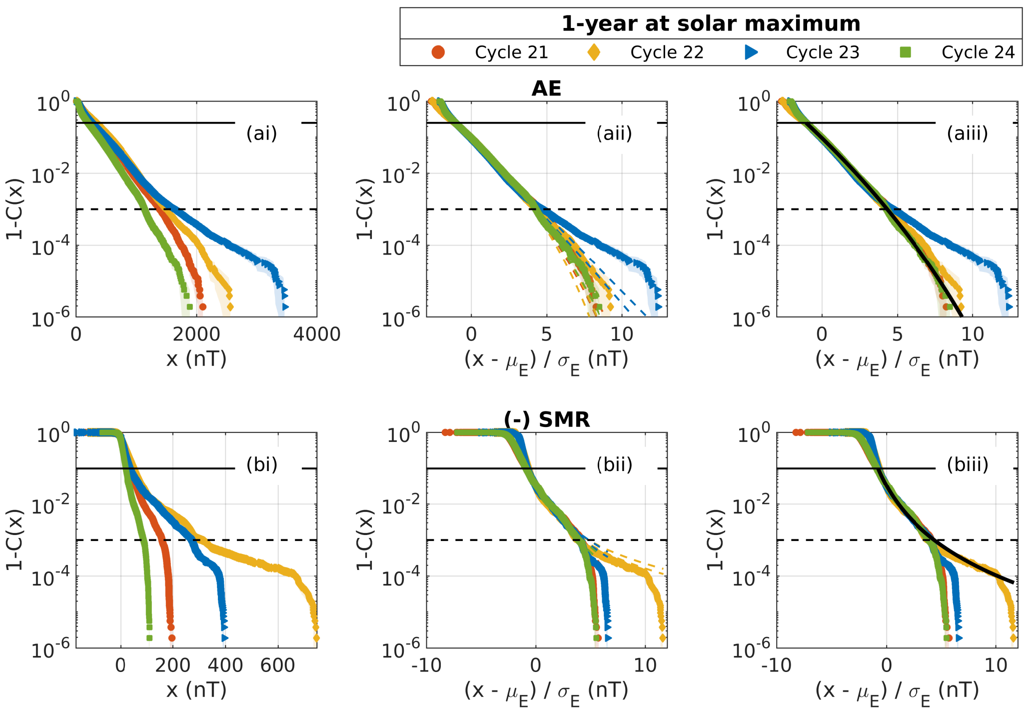


**Figure 9.** Single functional form for the empirical distribution long tail of 1-year intervals of the (a) $AE$ and (b) (-)$SMR$ indices at solar maximum. (i) Survival distributions of the index observations for cycles 21 (red), 22 (yellow), 23 (blue) and 24 (green) are plotted, uncertainties calculated using Greenwood's formula are shaded. (ii) The survival distributions of the index samples, rescaled to the mean and standard deviation of the exceedences of the $q_E$ quantile threshold, are plotted. Overplotted (dashed line) are the 95% confidence intervals of generalized Pareto distribution (GPD) fits for the cycle-specific exceedences. (c) The survival distributions of the rescaled index samples are plotted. Overplotted (solid black curve) is the generalized Pareto distribution (GPD) fit for the exceedences aggregated over all four solar cycles. The quantile $q_E$ (solid black line) is approximately at the transition between two regimes of the distributions. The 0.999 quantile is indicated (dashed black line).

**Table 1.** Generalized Pareto distribution parameters, $\xi$ and $\phi$, for fits to exceedences of the aggregated 1-year $AE$ and (-)$SMR$ index samples from solar cycles 21-24. Exceedences are defined above the quantile threshold, $q_E$.

| | $q_E$ | Phase | Shape ($\xi$) | Scale ($\phi$) |
|---|---|---|---|---|
| $AE$ | 0.75 | Minimum | -0.007 $\pm$ 0.003 | 1.014 $\pm$ 0.004 |
| | | Maximum | -0.048 $\pm$ 0.002 | 1.107 $\pm$ 0.004 |
| | | Decline | -0.042 $\pm$ 0.002 | 1.088 $\pm$ 0.004 |
| (-)$SMR$ | 0.9 | Minimum | 0.224 $\pm$ 0.005 | 0.603 $\pm$ 0.004 |
| | | Maximum | 0.246 $\pm$ 0.006 | 0.602 $\pm$ 0.004 |
| | | Decline | 0.264 $\pm$ 0.006 | 0.601 $\pm$ 0.004 |

This is in agreement with the findings of Chapman et al. (2018) for the 1-hour $AE$ index at 3.5 years of solar maximum ($\xi$=-0.071, $\phi$=1.169). Regarding the (-)$SMR$ index, the shape of the GPD fits fall in the Frechét class ($\xi$ >0) and is subexponential. This result is also in agreement with the results of Chapman et al. (2018) for the 1-hour (-)$Dst$ index at 3.5 years of solar maximum ($\xi$=0.198, $\phi$=0.664). The mean and variance of the distribution is finite when $\xi < 1/2$, this is true for all samples examined here. In summary, these results suggest that the near-tail mean and variance of the $AE$ and (-)$SMR$ indices for any given solar cycle phase sample are sufficient to quantify the full distribution of the cdf of observations. Different solar cycles vary in overall amplitude so that the absolute likelihood of observations of a given size will vary from one sample to the next, but the relative likelihood of large, as compared to small, observed values, does not vary from one cycle to the next. These are not predictions for the most extreme events that might be expected in an annual sample, but rather for bursts that populate to near-tail region of large-to-extreme observations which is operationally useful for long-term planning and system design (Owens et al., 2021).

## 7 Discussion

Geomagnetic activity occurs across a broad range of scales. One method to quantify this is to construct bursts from geomagnetic index timeseries by specifying a threshold. The statistics of bursts then provide an indicator of the overall space weather climate. Bursts identified in this manner will also in a broad sense capture space weather events, although we stress that a clear identification of geomagnetic storms and substorms requires additional diagnostics of magnetospheric activity. With this in mind, we have investigated how the statistical properties of bursts vary across the last four solar cycles, for which we have high time resolution geomagnetic index data. In Figure 1 we see that the variation in quantiles of annual $AE$ and (-)$SMR$ samples track the solar cycle variation in SSN. Therefore, where we identify bursts by thresholding above a fixed value, the burst statistics reflect the overall level of solar activity as captured by the SSN. On the other hand, thresholding at a quantile will tend to suppress solar cycle variation in burst statistics.

We have considered the burst return period, $R$, and burst duration, $\tau$, which are both commonly studied; $\tau$ is also an important factor in terms of burst size (Tindale et al., 2018; Uritsky et al., 2001) and burst time-integrated effects (Haines et al., 2019; Mourenas et al., 2018). Crossing theory stipulates that for a given timeseries sample, the burst distribution averages $\overline{\tau}$ and $\overline{R}$ are not independent quantities; their ratio, $\overline{\tau}/\overline{R}$, can be determined wholly from the quantiles of the underlying empirical distribution of the observations from which the bursts were constructed. We suggest that $\overline{\tau}/\overline{R}$ provides a dimensionless activity parameter which characterizes the fraction of time the magnetosphere spends in an active state for a given period.

Figures 3 and 5 show that when a fixed value threshold is applied, $\overline{\tau}$ and $\overline{R}$ exhibit detailed variation which is not consistent from one solar cycle to the next. However, $\overline{\tau}/\overline{R}$ tracks the variation of SSN. We find that $\overline{\tau}/\overline{R}$ is peaked in the declining phase for $AE$ year-long samples and approximately follows the SSN double peak for (-)$SMR$. We have seen that annual distribution quantiles from these year-long samples track SSN so the aforementioned $\overline{\tau}/\overline{R}$ tracking of SSN is a result from crossing theory. $\overline{\tau}/\overline{R}$ could in principle be predicted, given a prediction for the SSN for an upcoming solar cycle (see e.g. Nandy (2021)).

In Figures 4 and 6, thresholding at a quantile involves a threshold that moves up and down, tracking solar cycle activity, hence we do not see any robust features in $\overline{\tau}$ and $\overline{R}$ and crossing theory constrains $\overline{\tau}/\overline{R}$ to be constant. Qualitatively, when activity is high we see more frequent and larger (i.e. longer duration) events, so that for small $\overline{\tau}$ we observe large $\overline{R}$ and vice versa. Quantitatively, although $\overline{\tau}$ and $\overline{R}$ show detailed variation over the solar cycle, knowledge of one of $\overline{\tau}$ or $\overline{R}$ constrains the value of the other.

For moderate amplitude space weather events, there is sufficient data to directly identify bursts in a timeseries and then to subsequently obtain the mean return period and duration. Extreme space weather events are rare events, and at these correspondingly high thresholds there are insufficient bursts to directly estimate their mean duration or mean return period. In this case, crossing theory provides an estimate of the average duration of a burst of a given occurrence frequency or the average occurrence frequency of a burst of a given duration, at any threshold for which the cdf of the underlying raw observations, (that is, the observed *values* of the indices), may be obtained.

We can obtain the full distribution of burst parameters directly by thresholding the time series. In Figures 7 and 8 we compare across cycles 21-24 the distributions of return period and duration at solar maximum (we repeated this analysis for all solar cycle phases). The bursts can again be defined by thresholding at a fixed value, or at a fixed quantile. If thresholded at a fixed value, the distributions of burst parameters, and their means, vary from one solar cycle to the next. However if thresholded at a quantile, the burst parameter distribution means 'standardise' to a single value for all the distinct solar cycles, this is the case for the $AE$ index at solar minimum, maximum and declining phase and for the (-)$SMR$ index at solar maximum. Furthermore, except at the smallest values, the full burst distributions of durations and return periods of bursts fall roughly on top of each other, that is they tend to collapse onto a single functional form. This result connects knowledge of space weather climate to the overall intensity of space weather; Figures 7 and 8 empirically determine the mapping between a burst duration or return period and its likelihood of occurrence, for bursts at a given quantile threshold. Figure 1 then plots how that quantile translates into the physical value of the index that the burst has exceeded. A corollary is that more intense solar maxima have events that are both more frequent (shorter return period) and longer duration than less intense solar maxima, in a manner that is directly determined by how the quantiles of the index itself, rather than the bursts, vary across solar cycles.

The near-tail region of the distribution of $AE$ and (-)$SMR$ indices, when rescaled by the mean and standard deviation for a given year-long sample, are shown to exhibit collapse to a single master GPD distribution which is unique to solar minimum, maximum or declining phase of solar cycles 21-24, (shown in Figure 9), extending the results of Chapman et al. (2018). Lockwood et al. (2018) also found that the distribution shape of annual $AE$ indices did not differ significantly from one cycle to the next. Results such as these tell us that while the overall amplitude of observations may vary from one solar cycle to the next, if we have knowledge of just the moments of the exceedence distributions of the observed values of the index time series, we can, based on past solar cycles, estimate the full cdf. This in turn allows us to quantify the behaviour of the burst statistics in terms of the overall activity parameter $\overline{\tau}/\overline{R}$ and, in the case of $AE$ or (-)$SMR$ at maximum, estimate the full distribution of burst durations and return periods.

## 8 Summary

We analyse the time-series burst statistics and the empirical distributions for 1-year samples of the 1-minute Kyoto $AE$ index and the 1-minute SuperMAG $SMR$ index over solar cycles 21-24. We find that quantiles of year-long samples of the values of the $AE$ and (-)$SMR$ distributions track the solar cycle variation of the daily sunspot number (SSN). Bursts in the time-series are defined as excursions above a threshold which is either (i) a fixed value or (ii) a quantile of the distribution of the observed index values. We study the solar cycle dependence of the distributions of the burst return periods (time between consecutive threshold upcrossings), $R$, and the burst durations (time between threshold upcrossing and downcrossing), $\tau$.

Our main results are as follows:

1. At fixed value burst thresholds the ratio of the mean burst duration to return period, $\overline{\tau}/\overline{R}$, is peaked in the declining phase for $AE$ annual samples and follows the SSN double peak for (-)$SMR$. At fixed quantile burst thresholds crossing theory constrains $\overline{\tau}/\overline{R}$ to be constant.
2. We obtain the full distribution of burst duration $\tau$ and return period $R$ for bursts identified in year-long samples at three different phases of the solar cycle. Fixed quantile threshold bursts have distributions that fall on single empirical curves for each of (i) the $AE$ index at solar minimum, maximum and declining phase and (ii) the (-)$SMR$ index at solar maximum. This goes beyond the constraint on average $\overline{\tau}/\overline{R}$ from crossing theory.
3. The 'mid-tail' of the empirical cdfs of the observed values of the $AE$ and (-)$SMR$ indices collapse onto common functional forms specific to each index and cycle phase when normalized to the first two moments of their exceedence distributions.

Crossing theory constrains how the ratio $\overline{\tau}/\overline{R}$ of bursts depends on the underlying distribution of the observed quantity, here, the $AE$ and $SMR$ indices. Ordered behaviour in the distribution of the observed quantity then translates to ordered behaviour in the burst $\overline{\tau}/\overline{R}$, suggesting that it is a useful activity parameter to relate overall solar activity to magnetospheric response. Furthermore, there is ordered behaviour in the *full distribution* of bursts which is consistent with, but goes beyond, the constraint of crossing theory. Taken together, these results may combine to offer important constraints in the quantification of overall space weather activity levels.

**Acknowledgments**

The $AE$ index used in this paper was provided by the WDC for Geomagnetism, Kyoto (http://wdc.kugi.kyoto-u.ac.jp/wdc/Sec3.html) and was retrieved from the WDC Kyoto interface (at http://wdc.kugi.kyoto-u.ac.jp/aeasy/index.html). The $SMR$ index was retrieved from the SuperMAG interface (at https://supermag.jhuapl.edu/indices/). We gratefully acknowledge the SuperMAG collaborators (https://supermag.jhuapl.edu/info/?page=acknowledgement). Sunspot data is from the World Data Center SILSO, Royal Observatory of Belgium, Brussels (http://sidc.be/silso/datafiles). S. C. C. acknowledges Air Force Office of Scientific Research (AFOSR) Grant FA9550-17-1-0054 and Science and Technology Facilities Council (STFC) Grant ST/P000320/1. We thank J. Gjerloev for valuable discussions.

## References


Acero, F., Vaquero, J., Gallego, M., & García, J. (2018). A limit for the values of the Dst geomagnetic index. *Geophysical Research Letters*, *45*(18), 9435–9440.

Ahn, B.-H., Akasofu, S.-I., & Kamide, Y. (1983). The Joule heat production rate and the particle energy injection rate as a function of the geomagnetic indices AE and AL. *Journal of Geophysical Research: Space Physics*, *88*(A8), 6275–6287.

Alberti, T., Consolini, G., & De Michelis, P. (2021). Complexity measures of geomagnetic indices in the last two solar cycles. *Journal of Atmospheric and Solar-Terrestrial Physics*, *217*, 105583.

Baker, D. (2000). Effects of the Sun on the Earth's environment. *Journal of Atmospheric and Solar-Terrestrial Physics*, *62*(17-18), 1669–1681.

Barabási, A.-L. (2010). *Bursts: The Hidden Pattern Behind Everything We Do.* New York: Dutton.

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.

Borovsky, J. E. (2020). What magnetospheric and ionospheric researchers should know about the solar wind. *Journal of Atmospheric and Solar-Terrestrial Physics*, *204*, 105271.

Borovsky, J. E., & Denton, M. H. (2006). Differences between CME-driven storms and CIR-driven storms. *Journal of Geophysical Research: Space Physics*, *111*(A7).

Chapman, S. C., Horne, R. B., & Watkins, N. W. (2020). Using the aa Index Over the Last 14 Solar Cycles to Characterize Extreme Geomagnetic Activity. *Geophysical Research Letters*, *47*(3), e2019GL086524.

Chapman, S. C., McIntosh, S., Leamon, R., & Watkins, N. W. (2020). Quantifying the solar cycle modulation of extreme space weather. *Geophysical Research Letters*, *47*(11), e2020GL087795.

Chapman, S. C., Watkins, N. W., Dendy, R. O., Helander, P., & Rowlands, G. (1998). A simple avalanche model as an analogue for magnetospheric activity. *Geophysical Research Letters*, *25*(13), 2397-2400. doi: https://doi.org/10.1029/98GL51700

Chapman, S. C., Watkins, N. W., & Stainforth, D. A. (2019). Warming trends in summer heatwaves. *Geophysical Research Letters*, *46*(3), 1634–1640.

Chapman, S. C., Watkins, N. W., & Tindale, E. (2018). Reproducible aspects of the climate of space weather over the last five solar cycles. *Space Weather*, *16*(8), 1128–1142.

Consolini, G. (1997). Sandpile cellular automata and magnetospheric dynamics. In S. Aiello, N. Iucci, G. Sironi, A. Treves, & U. Villante (Eds.), *Cosmic physics in the year 2000* (Vol. 58). Societa Italia di Fisica.

Cramér, H., & Leadbetter, M. R. (2004). *Stationary and Related Stochastic Processes: Sample Function Properties and Their Applications.* Dover Publications, Inc.

Davis, T. N., & Sugiura, M. (1966). Auroral electrojet activity index AE and its universal time variations. *Journal of Geophysical Research*, *71*(3), 785–801.

Denton, M. H., Borovsky, J. E., Skoug, R., Thomsen, M., Lavraud, B., Henderson, M., … Liemohn, M. (2006). Geomagnetic storms driven by ICME-and CIR-dominated solar wind. *Journal of Geophysical Research: Space Physics*, *111*(A7).

Elvidge, S. (2020). Estimating the Occurrence of Geomagnetic Activity Using the Hilbert-Huang Transform and Extreme Value Theory. *Space Weather*, *18*(8), e2020SW002513.

Embrechts, P., Klüppelberg, C., & Mikosch, T. (2013). *Modelling extremal events: for insurance and finance* (Vol. 33). Springer Science & Business Media.

Freedman, D., & Diaconis, P. (1981). On the histogram as a density estimator: $L_2$ theory. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, *57*(4), 453–476.

Freeman, M. P., & Watkins, N. W. (2002). The heavens in a pile of sand. *Science*, *298*(5595), 979–980.

Freeman, M. P., Watkins, N. W., & Riley, D. (2000). Evidence for a solar wind origin of the power law burst lifetime distribution of the AE indices. *Geophysical Research Letters*, *27*(8), 1087–1090.

Gjerloev, J. (2012). The SuperMAG data processing technique. *Journal of Geophysical Research: Space Physics*, *117*(A9).

Gonzalez, W. D., de Gonzalez, A. L. C., & Tsurutani, B. T. (1990). Dual-peak solar cycle distribution of intense geomagnetic storms. *Planetary and Space Science*, *38*(2), 181–187.

Gonzalez, W. D., Tsurutani, B. T., & De Gonzalez, A. L. C. (1999). Interplanetary origin of geomagnetic storms. *Space Science Reviews*, *88*(3), 529–562.

Greenwood, M. (1926). A report on the natural duration of cancer. *Reports of Public Health and Related Subects*(33).

Haines, C., Owens, M., Barnard, L., Lockwood, M., & Ruffenach, A. (2019). The variation of geomagnetic storm duration with intensity. *Solar Physics*, *294*(11), 1–15.

Hamilton, D. C., Gloeckler, G., Ipavich, F., Stüdemann, W., Wilken, B., & Kremser, G. (1988). Ring current development during the great geomagnetic storm of February 1986. *Journal of Geophysical Research: Space Physics*, *93*(A12), 14343–14355.

Hathaway, D. H. (2015). The solar cycle. *Living reviews in solar physics*, *12*(1), 4.

Hush, P., Chapman, S., Dunlop, M., & Watkins, N. W. (2015). Robust statistical properties of the size of large burst events in AE. *Geophysical Research Letters*, *42*(21), 9197–9202.

Iyemori, T. (1990). Storm-time magnetospheric currents inferred from mid-latitude geomagnetic field variations. *Journal of geomagnetism and geoelectricity*, *42*(11), 1249–1265.

Kakad, B., & Kakad, A. (2020). Characteristics of probability distribution functions of low-and high-latitude current systems during Solar Cycle 24. *Advances in Space Research*, *65*(6), 1559–1567.

Kamide, Y., Baumjohann, W., Daglis, I., Gonzalez, W., Grande, M., Joselyn, J., ... others (1998). Current understanding of magnetic storms: Storm-substorm relationships. *Journal of Geophysical Research: Space Physics*, *103*(A8), 17705–17728.

Lawrance, A., & Kottegoda, N. (1977). Stochastic modelling of riverflow time series. *Journal of the Royal Statistical Society: Series A (General)*, *140*(1), 1–31.

Lockwood, M., Bentley, S. N., Owens, M. J., Barnard, L. A., Scott, C. J., Watt, C. E., ... Freeman, M. P. (2019). The development of a space climatology: 3. Models of the evolution of distributions of space weather variables with timescale. *Space Weather*, *17*(1), 180–209.

Lockwood, M., Owens, M. J., Barnard, L. A., Scott, C. J., Watt, C. E., & Bentley, S. (2018). Space climate and space weather over the past 400 years: 2. Proxy indicators of geomagnetic storm and substorm occurrence. *Journal of Space Weather and Space Climate*, *8*, A12.

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.

Mayaud, P.-N. (1980). Derivation, meaning, and use of geomagnetic indices. *Washington DC American Geophysical Union Geophysical Monograph Series*, *22*,

607.
Milan, S. E., Clausen, L. B. N., Coxon, J. C., Carter, J. A., Walach, M.-T., Laundal, K., ... others (2017). Overview of solar wind-magnetosphere-ionosphere-atmosphere coupling and the generation of magnetospheric currents. *Space Science Reviews*, *206*(1), 547–573.
Moloney, N. R., & Davidsen, J. (2011). Extreme bursts in the solar wind. *Geophysical Research Letters*, *38*(14).
Moloney, N. R., & Davidsen, J. (2014). Stationarity of extreme bursts in the solar wind. *Physical Review E*, *89*(5), 052812.
Mourenas, D., Artemyev, A., & Zhang, X.-J. (2018). Statistics of extreme time-integrated geomagnetic activity. *Geophysical Research Letters*, *45*(2), 502–510.
Nakamura, M., Yoneda, A., Oda, M., & Tsubouchi, K. (2015). Statistical analysis of extreme auroral electrojet indices. *Earth, Planets and Space*, *67*(1), 1–8.
Nandy, D. (2021). Progress in solar cycle predictions: sunspot cycles 24-25 in perspective. *Solar Physics*, *296*(3), 1–27.
Newell, P., & Gjerloev, J. (2011). Evaluation of SuperMAG auroral electrojet indices as indicators of substorms and auroral power. *Journal of Geophysical Research: Space Physics*, *116*(A12).
Newell, P., & Gjerloev, J. (2012). SuperMAG-based partial ring current indices. *Journal of Geophysical Research: Space Physics*, *117*(A5).
Nose, M., Iyemori, T., Sugiura, M., & Kamei, T. (2015). Geomagnetic AE index. *World Data Center for Geomagnetism, Kyoto*, *10*, 15–031.
Nurhan, Y. I., Johnson, J. R., Homan, J. R., Wing, S., & Aschwanden, M. J. (2021). Role of the solar minimum in the waiting time distribution throughout the heliosphere. *Geophysical Research Letters*, *48*(16), e2021GL094348.
Owens, M. J., Lockwood, M., Barnard, L. A., Scott, C. J., Haines, C., & Macneil, A. (2021). Extreme Space-Weather Events and the Solar Cycle. *Solar Physics*, *296*(5), 1–19.
Pulkkinen, T. (2007). Space weather: terrestrial perspective. *Living Reviews in Solar Physics*, *4*(1), 1-60.
Richardson, I., Cane, H., & Cliver, E. (2002). Sources of geomagnetic activity during nearly three solar cycles (1972–2000). *Journal of Geophysical Research: Space Physics*, *107*(A8), SSH–8.
Riley, P. (2012). On the probability of occurrence of extreme space weather events. *Space Weather*, *10*(2).
Sandhu, J. K., Walach, M.-T., Allison, H., & Watt, C. (2019). A global view of storms and substorms. *Astronomy & Geophysics*, *60*(3), 13–19.
SILSO World Data Center. (1973-2020). The International Sunspot Number. *International Sunspot Number Monthly Bulletin and online catalogue*.
Silverman, B. W. (1986). *Density estimation for statistics and data analysis* (Vol. 26). CRC press.
Sugiura, M. (1964). Hourly values of equatorial Dst for the IGY. *Annals of the International Geophysical Year*, *35*(9).
Takalo, J. (1993). *Correlation dimension of AE data.* Ph. Lic. Thesis, Laboratory Report 3, Department of Physics, University of Jyväskylä.
Tindale, E., Chapman, S., Moloney, N., & Watkins, N. W. (2018). The dependence of solar wind burst size on burst duration and its invariance across solar cycles 23 and 24. *Journal of Geophysical Research: Space Physics*, *123*(9), 7196–7210.
Tsubouchi, K., & Omura, Y. (2007). Long-term occurrence probabilities of intense geomagnetic storm events. *Space Weather*, *5*(12).
Uritsky, V. M., Klimas, A. J., & Vassiliadis, D. (2001). Comparative study of dynamical critical scaling in the auroral electrojet index versus solar wind fluctuations. *Geophysical research letters*, *28*(19), 3809–3812.

Vanmarcke, E. (2010). *Random fields: analysis and synthesis, 2nd edition.* World scientific.

Watkins, N. W., Pruessner, G., Chapman, S. C., Crosby, N. B., & Jensen, H. J. (2016). 25 years of self-organized criticality: Concepts and controversies. *Space Science Reviews*, *198*(1), 3–44.

Wheatland, M. S. (2000). The origin of the solar flare waiting-time distribution. *The Astrophysical Journal*, *536*(2), L109-L112. doi: 10.1086/312739