The development of a space climatology: 3. Models of the evolution of distributions of space weather variables with timescale

Article

Accepted Version

Lockwood, M. ORCID: https://orcid.org/0000-0002-7397-2172, Bentley, S. N., Owens, M. J. ORCID: https://orcid.org/0000-0003-2061-2453, Barnard, L. A. ORCID: https://orcid.org/0000-0001-9876-4612, Scott, C. J. ORCID: https://orcid.org/0000-0001-6411-5649, Watt, C. E., Allanson, O. and 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). pp. 180-209. ISSN 1542-7390 doi: https://doi.org/10.1029/2018SW002017 Available at https://centaur.reading.ac.uk/81043/

To link to this article DOI: http://dx.doi.org/10.1029/2018SW002017

Publisher: American Geophysical Union

Final version 16th November 2018

The development of a space climatology: 3. Models of the evolution of distributions of space weather variables with timescale

Mike Lockwood[1], Sarah N. Bentley[1], Mathew J. Owens[1], Luke A. Barnard[1],
Chris J. Scott[1], Clare E. Watt[1], Oliver Allanson [1] and Mervyn P. Freeman[2]

[1]Department of Meteorology, University of Reading, Earley Gate, P.O. Box 243, Reading, Berkshire, RG6 6BB, UK.

[2] British Antarctic Survey, High Cross, Madingley Road, Cambridge, CB3 0ET, UK.

Corresponding author: M. Lockwood (m.lockwood@reading.ac.uk)

Key Points:

- Core distributions and extreme events of geomagnetic activity are studied as a function of averaging timescale τ
- The autocorrelation is shown to have a dominant role determining how these core distributions vary with averaging timescale τ
- Models for computing the distribution of geomagnetic activity for a given timescale τ and annual mean are presented

Abstract

We study how the probability distribution functions of power input to the magnetosphere P_α and of the geomagnetic *ap* and *Dst* indices vary with averaging timescale, τ, between 3 hours and 1 year. From this we develop and present algorithms to empirically model the distributions for a given τ and a given annual mean value. We show that lognormal distributions work well for *ap*, but because of the spread of *Dst* for low activity conditions, the optimum formulation for *Dst* leads to distributions better described by something like the Weibull formulation. Annual means can be estimated using telescope observations of sunspots and modelling, and so this allows the

distributions to be estimated at any given τ between 3 hour and 1 year for any of the past 400 years, which is another important step towards a useful space weather climatology. The algorithms apply to the core of the distributions and can be used to predict the occurrence rate of "large" events (in the top 5% of activity levels): they may contain some, albeit limited, information relevant to characterizing the much rarer "superstorm" events with extreme value statistics. The algorithm for the *Dst* index is the more complex one because, unlike *ap*, *Dst* can take on either sign and future improvements to it are suggested.

1. Introduction

This paper is the third of a series of three that is aimed at putting in place some of the key elements that will be needed to build a space weather climatology that covers both grand solar maximum and grand solar minimum conditions. As discussed in the introductions to Papers 1 and 2 [*Lockwood et al.,* 2018b; c], information on space climate over an interval long enough to cover both a grand solar minimum and a grand solar maximum (of order 400 years) is available only in the form of modelled annual means of some key variables [*Owens et al.,* 2017]. Hence developing a climatology giving the probability of space weather events of a given geoeffectiveness that covers both these extremes of the long-term solar variation requires us to develop an understanding of relationships between these annual means and the distributions of event amplitudes, quantified over the relevant timescales. Because space weather events come in bursts, the integrated value of any activity index X over the most relevant timescale τ, I_X, is a useful metric [*Echer et al*., 2008; *Lockwood et al.,* 2016*; Borovsky,* 2017; *Tindale et al.,* 2018], and this equals the arithmetic mean value times τ (i.e., $I_X = \int_\tau X dt = \tau \langle X \rangle_\tau$). Hence it is important to study how $\langle X \rangle_\tau$ varies with τ and how it relates to the annual arithmetic mean value $\langle X \rangle_{\tau=1yr}$. *Lockwood et al.* [2018a] have demonstrated how annual means can be used to quantify the frequency of geomagnetic disturbance events above a given (large but not extreme) threshold for the past 400 years, but they studied only hourly and daily means (τ = 1 hr and τ = 1 day) which, in general, will not be the most relevant timescales for all space weather phenomenon. For example, *Lockwood et al.* [2016] recently studied the interplanetary conditions leading to large geomagnetic storms as detected in the *Dst* index and found $\tau \approx$ 6 hrs (with a 2σ uncertainty range of 4-12 hrs) was optimum for predicting the maximum of the storm (i.e., the minimum *Dst*) but $\tau \approx$ 4.5 days was needed to best predict the integrated *Dst* over the duration of the storm. Paper 1 [*Lockwood et al.*, 2018b] studied energy coupling from the solar wind into the magnetosphere and showed that neglecting the effects of gaps in interplanetary data has, in the past, introduced serious errors into derived solar wind-magnetosphere coupling functions. Paper 1 also used near-continuous data to show that there is no evidence that the coupling function varies with averaging timescale τ between 1 minute and 1 year. Paper 2 [*Lockwood et al.,* 2018c] used this result to study the distribution of power input into the magnetosphere P_α and why the probability density function (p.d.f.) of $<P_\alpha>_\tau$ (i.e., P_α averaged over intervals of duration τ) has the form it does at $\tau = 1$ min. Paper 2 also showed how this p.d.f. evolves with increasing τ up to 3 hours, giving the observed p.d.f.s of 3-hourly geomagnetic indices. In the present paper, we study how the distributions of power input into the magnetosphere, and of the geomagnetic indices, continue to evolve with increasing τ between 3 hours and 1 year, allowing us to study the relationships of the p.d.f. at any relevant τ to the annual mean. These are key relationships that can make it possible to construct a climatology of space weather events based on observations of solar variability over the past 400 years.

1.1 Core distributions of space weather variables and extreme events

In this section, we make clear the distinctions between the "core" distribution of space weather events, "large events" (for example, *Lockwood et al.* [2017a; 2018a] studied events in the top 5%) and "extreme events". Our aim is to investigate how much information on the extreme events could potentially be gleaned from the annual means and the core distribution. We use the 3-hourly *ap* planetary geomagnetic range index which are available continuously since 1932. This index is used because of the longevity of the data series and because it is more robust than the *aa* index as it employs more than just two observatories. Appendix B shows that the *ap* index has a marked tendency to exaggerate the semi-annual variation in average values by having a larger response to events occurring at the equinoxes and also has a lower response to large events during northern-hemisphere winter. We here use a version of *ap*, ap_C, that includes a correction for the effect of this uneven response in *ap*, as described in Appendix B. To compare to any events before 1932 we use the *aa* geomagnetic index, using inter-calibration curves that are also presented in Appendix B.

Allen [1982] pointed out that averages of *ap* over a calendar day (by convention referred to as $Ap = <ap>_{\tau=1\text{day}}$) are not appropriate for defining storm days because an isolated storm that spans midnight UT would be recorded as two moderately disturbed days rather than a single large storm day. Hence *Allen* proposed using 24-hour boxcar (running) means of *ap*, which he termed *Ap**. These have been employed by *Kappenman*, [2005] and *Cliver and Svalgaard* [2004]. For the purposes of identifying and ranking storm days we take the largest value of the 8 such running-means of the corrected *ap* index in each calendar day, $[Ap_C{*}]_{MAX}$. A rank-order listing of the largest events defined this way is given in the Supporting Information file, along with available references.

Many papers have found variables of near-Earth interplanetary space and the magnetosphere approximately follow a lognormal (or similar) distribution for the great majority of the time [*Hapgood et al.*, 1991; *Dmitriev et al.*, 2009, *Vaselovsky et al.*, 2010; *Farrugia et al.*, 2012; *Lockwood and Wild*, 1993, *Weigel and Baker*, 2003; *Vörös et al.*, 2015, *Love et al.*, 2015; *Lotz and Danskin*, 2017; *Riley and Love*, 2017, *Xiang and Qu*, 2018]. This mathematical formulation describes the "core" of the distribution, but often fails to match the occurrence of very large or extreme events [e.g., *Riley*, 2012; *Baker et al.,* 2013, *Cliver and Dietrich*, 2013; *Lotz and Danskin*, 2017]. Hence such cases are often described by substituting a distribution to the large-event tail that is different to that which fits the core of the distribution. Extreme Value Statistics (EVS) [e.g., *Kotz and Nadarajah* 2000; *Beirlant et al.* 2004; *Coles*, 2004] has been widely applied, initially in studies of hydrology but subsequently to extreme terrestrial weather events and many other areas such as in engineering, insurance and finance. The "extremal types theorem" (also called the "Fisher–Tippett–Gnedenko" theorem) [*Coles*, 2004], states that extreme maxima follow one of three types of distribution ("Gumbel", "Fréchet" and "(negative) Weibull", which are encapsulated in a family of continuous probability distributions called the Generalized Extreme Value (GEV) distribution. In the "block maxima" (BM) approach to extreme values, the observation period is divided into non-overlapping periods of equal size and attention given to the maximum observation in each period to which the GEV distribution applies. In the "peaks-over-threshold" (POT) approach, observations that exceed a certain high threshold are selected. The second theorem in extreme value theory is the Pickands–Balkema–de Haan theorem and states that the threshold excesses have an approximate distribution within the "Generalized Pareto Distribution" (GPD) family. EVS has been applied to geomagnetic indices (for example by *Siscoe* [1976], *Tsubouchi and Omura*, [2007], *Silbergleit* [1996; 1999], *Chapman et al.* [2018] and *Mourenas et al*. [2018]), to the occurrence of very large Geomagnetically-Induced Currents (GICs) [*Thompson et al.*, 2011;

Lotz and Danskin, 2017], and to the fluxes of energetic magnetospheric particles [*Koons*, 2001; *O'Brien et al.*, 2007].

Figure 1 places into context the relationship of the extreme event tail to the core distribution for geomagnetic activity as measured by the (corrected) *ap* index, ap_C. The plot shows (top) some selected annual distributions of the Ap_C* index and (bottom) the corresponding distributions of Ap_C* as ratio of the annual mean value, $Ap_C*/<ap_C>_{\tau=1yr}$. The gray histograms are for all available Ap_C* data (i.e. covering the years 1932-2016). Note that we here quote *ap*, and hence ap_C, Ap_C* and $[Ap_C*]_{MAX}$, as indices without units (the standard *ap* values are an index in units of 2nT and hence the values in nT would be double those given here [*Menvielle and Berthelier*, 1991]). The black vertical dashed line shows *Apo*, the 95th percentile of all available samples. The year 1960 (shown in red) was one year after the maximum of the largest sunspot cycle (number 9) of the recent grand solar maximum [*Lockwood et al.*, 2009] and gave the largest annual mean value since *ap* measurements began ($<ap_C>_{\tau=1yr} = 23.65$) and also contained the largest observed event since 1932, as determined by a daily $[Ap_C*]_{MAX}$ value of 249 on 13 November of that year. The year 2009 (in blue) was at the low sunspot minimum (between cycles 23 and 24) gave the smallest annual mean in the record ($<ap_C>_{\tau=1yr} = 3.93$). The year 1859 (in orange) has been chosen because between 28 August and 5 September of that year, the Carrington event took place (see contemporary reports by E. Loomis, collected together by *Shea and Smart* [2006]), which is thought to be the largest terrestrial space weather event to have been observed as it happened [*Nevanlinna*, 2006; *Cliver and Dietrich*, 2013; *Ngwira et al.,* 2014]. The mean $<ap>_{\tau=1yr}$ for 1859 has been estimated to have been 10.98 by *Lockwood et al.* [2018a]. The distribution of daily *Ap* occurrence for 1859 shown in Figure 1 has been generated from the estimated mean value for that year using a model that will be developed in the present paper and is described in Appendix A. The distribution for 2012 is included (in green, $<ap_C>_{\tau=1yr} = 9.20$) because on 23 July of that year a very large and very rapid Coronal Mass Ejection (CME) erupted, an event which would have generated extreme terrestrial space weather (a "superstorm") had it hit the Earth. It was observed as it passed over the STEREO-A spacecraft and, from modelling based on the measurements taken by that craft and by solar instruments, it is estimated it would have caused a terrestrial event as large as the Carrington event, had the eruption taken place just one week earlier such that the CME would have hit Earth's magnetosphere instead of STEREO-A [*Baker et al.*, 2013; *Ngwira et al.,* 2013].

From available magnetometer data, *Nevanlinna* [2006] has estimated that the daily *aa* geomagnetic index reached *Aa* = 400nT during the Carrington event. This estimate allows for missing data, but may still be an under-estimate and *Cliver and Svalgaard* [2004] estimated the peak value of the running mean of a corrected version of the *aa* index over 24 hours of *Aa** to be 425nT. The *aa* index was designed by *Mayaud* [1972, 1980] to act as an equivalent to the *ap* index using data from just two stations: however, the data since 1932 show that the two are not linearly related, with *ap* at large *aa* being significantly lower than would be obtained from a linear fit. Polynomial fits of daily means, *Ap*, as a function of the daily means in *aa* (by convention termed *Aa*) are given in Appendix B for the four quarter-year intervals around the equinoxes and solstices. Taking the peak *Aa* to be 425nT for the Carrington event, the relevant equation (B3) gives an estimated maximum *Ap** value of 284±30. Because it is considered that the STEREO event would have given a storm comparable to the Carrington event, we here take this *Ap** to apply to it as well. These values of $[Ap^*]_{\mathrm{MAX}}$ of 284 are shown by the vertical dot-dash lines. Applying the time-of year correction given in Appendix B, this yields $[Ap_{\mathrm{C}}^*]_{\mathrm{MAX}}$ of 215±23 and 211±23, respectively, for these two events. These estimated $[Ap_{\mathrm{C}}^*]_{\mathrm{MAX}}$ values for the Carrington and STEREO events are shown in Figure 1a by, respectively, the solid vertical orange and green vertical lines. By way of comparison, the largest daily-mean in the observed $[Ap_{\mathrm{C}}^*]_{\mathrm{MAX}}$ record (since 1932) is 249, recorded on 13 November 1960.

The list of storm days, since 1932 ranked by their $[Ap_{\mathrm{C}}^*]_{\mathrm{MAX}}$ values is given in the Supporting Information file. It has similarities to other such lists [e.g., *Nevanlinna et al.*, 2006; *Kappenman*, 2005; *Cliver and Svalgaard*, 2004], but there are differences because we have made allowance for the variation with time-of-year of the *Ap** response and, in the case of the Carrington event, the relationship between the *Aa** and *Ap** indices. Even quite small changes in the estimated magnitude of the storm day can have a large effect on its ranking order. The major surprise is that the positions of both the Carrington and STEREO events in the list is somewhat lower than in other lists if we correct for the tme-of-year dependence of *ap* (estimate 1, $[Ap_{\mathrm{C}}^*]_{\mathrm{MAX}}$). This raises the question as to whether this correction should be applied to these events or not. Logically, there is no doubt that it should be as equation B-3 converts the *Aa** estimate into an *Ap** value which should then need correcting to become Ap_{C}^*. The main argument for not applying the correction is that the original *Aa** estimate is a proxy compiled from other sources. That these sources are largely

European-sector mid-latitude observatories and *Ap* is heavily weighted to mid-latitude European station data, does argue that this correction should indeed be applied. However, there remains great uncertainty in the true magnitude of the Carrington event. We also note that $[Ap_C^*]_{MAX}$ is almost certainly not a fully adequate metric of this superstorms because it does not take account of the fact that the Carrington event on 3 September was in the middle of an extended interval of very high geomagnetic activity between 28 August and 5 September and this almost certainly drove excessively large negative *Dst* values through the integrated effect on the ring current population, giving the famously large negative deflection recorded at the lower-latitude Colaba observatory in Mumbai.

Lockwood et al. [2017a] estimated the annual mean power input into the magnetosphere $<P_\alpha>_\tau$ from the reconstructed solar wind and interplanetary field variables derived by *Owens et al.* [2017], and from this *Lockwood et al.* [2018] have estimated that the annual mean of *ap* for 1859 was 10.98. Hence the estimated peak $[Ap_C^*]_{MAX}$ /$<ap_C>_{\tau=1yr}$ for the Carrington event is 19.5±2.1 (shown by the solid orange line in the lower panel of Figure 1) for the corrected data and $[Ap^*]_{MAX}/<ap_C>_{\tau=1yr}$ = 25.9±2.7 for the uncorrected value (the orange dot-dashed line). From the observed $<ap_C>_{\tau=1yr}$ of 9.20 for 2012 the $[Ap_C^*]_{MAX}$ /$<ap_C>_{\tau=1yr}$ for the STEREO event would have been 23.0±2.5 (shown by the green line in the lower panel of Figure 1) and $[Ap^*]_{MAX}$ /$<ap_C>_{\tau=1yr}$ = 30.9±3.3 for the uncorrected data (green dot-dash line). These ratio estimates are much larger values than for the observed 13 November 1960 event, for which $[Ap_C^*]_{MAX}/<ap_C>_{\tau=1yr}$ is considerably lower, being 10.51 because it occurred during the most active geomagnetic year on record. Table S-7 of the Supporting Information shows that the largest value of $[Ap_C^*]_{MAX}/<ap_C>_{\tau=1yr}$ in the observational record (since 1932) is 16.27 for 8 February 1986 (for which $[Ap_C^*]_{MAX}$ = 203, the 7th largest value). This is the outstanding example in the observational record of a big storm being observed very close to sunspot minimum; however, its $[Ap_C^*]_{MAX}/<ap_C>_{\tau=1yr}$ ratio is still very much smaller than that estimated for the Carrington and STEREO events. In their absolute corrected Ap_C^* values or uncorrected Ap^* values, the Carrington and STEREO events appear to be comparable with, or somewhat larger than, the largest events seen since 1932; however, they arose in years of relatively low average activity and so are wholly exceptional in their $Ap_C^*/<ap_C>_{\tau=1yr}$ and $Ap^*/<ap_C>_{\tau=1yr}$ values.

Figure 1 demonstrates why the description of superstorms requires more than an extrapolation of the core and hence needs the application of EVS. However, there may still some valuable information on extreme events to be obtained from the core distribution, as *Love* [2012] and *Love et al.* [2015] have demonstrated for large geomagnetic storms (as defined and quantified using the *Dst* geomagnetic index). The points in Figure 2 show the available 31040 24-hour $[Ap_C*]_{MAX}$ samples as a function of the annual mean of the year in which they occur: the cyan points are the top 100 days (0.32%) in terms of $[Ap_C*]_{MAX}$ value (shown by the short vertical cyan lines in Figure 1); the mauve points are the top 6 days (0.02%, shown by the short vertical mauve lines in Figure 1); and the grey points are the remaining 99.68%. Figure 2 stresses how much our understanding rests rather critically on the estimates of the 1859 and 2012 superstorm values of (the orange and green squares being the uncorrected values and the triangles being the corresponding corrected values). If we do not consider these two events and look just at the observed record since 1932, we see a quite strong relationship between the largest value seen in the year and the average value for that the year with the data points falling in the bottom right half of the plot. The corrected $[Ap_C*]_{MAX}$ values for the 1859 and 2012 superstorms (the orange and green triangles) are close to being in line with this relationship, especially the lower values of the uncertainty range. These values suggest that the occurrence of extreme superstorms is (weakly) related to the average activity in those years and that the extreme events are forming something like the negative Weibull distribution "pile up" towards a maximum possible value not much greater than that for the November 1960 event. However the uncorrected values, $[Ap*]_{MAX}$ (shown by the green and orange squares) appear to be a completely different class of event from the events seen after 1932 and not obeying any sort of relationship between the peak and mean values. We should here also note that it is possible that even these uncorrected values are underestimates (being based on the *Cliver and Svalgaard* [2004] estimate of $Aa*$) that have been limited by procedure of quantizing the available data into k-index bands [see *Lockwood et al*., 2018d]. Thus the uncertainty in the estimated severity of the Carrington and STEREO events becomes crucial. On the other hand, the lower estimates for the Carrington and STEREO events suggest that the annual mean value and the core distribution could be helpful in quantifying the probability of the extreme events.

Even if the former proves to be the case and annual means or of no assistance in predicting superstorms, characterizing the core of the distribution (as opposed to the extreme tail) is, however, still important in space weather applications where the integral of the space weather activity is of relevance and the threshold to the effect is not in the extreme tail. Examples would include the effect of GICs on pipeline corrosion [*Boteler*, 2000; *Pulkkinen et al.*, 2001; *Gummow*, 2002; *Cole*, 2003; *Pirjola*, 2005; *Pirjola et al.,* 2005, *Viljanen et al*., 2006*; Ingham and Rodger,* 2018]; the effect of GICs on power grid transformer degradation [*Kappenman, and Radasky*, 2005; *Gaunt*, 2016]; the effect of energy deposition in the upper atmosphere on the orbits of LEO satellites and space debris [*Doornbos and Klinkrad*, 2006]; and the effect of integrated radiation dose on the degradation of spacecraft electronics [*Baker*, 2000; *Fleetwood et al*., 2000]. In all these examples, although the extreme superstorm events have a large effect, they are rare and a much larger number of smaller events, described by the core distribution, can also have a significant integrated effect. Lastly, we note that *Chapman et al.* [2018] have recently studied the extreme event tails in several terrestrial disturbance indices during recent maxima of the solar cycle and fitted Generalized Pareto Distributions. They found that if the mean and variance of the large-to-extreme observations can be predicted for a given solar maximum, then a relationship between the core distribution and the extreme tail can be found giving a description of the full distribution. Thus it does appear possible that the study of the core of the distributions presented here could be extended to characterize the extreme tails: this will be the subject of a future study.

As mentioned above, the $[Ap_C^*]_{MAX}$ values are unlikely be the best indicators of all storm characteristics, in particular in relation to the ring current and the *Dst* geomagnetic index. This gives another reason why we should study the core of the distributions, associated with storm “pre-conditioning” and the fact that the best predictors of large *Dst* storm occurrence are time-integrated over long intervals (several days) [*Lockwood et al.,* 2016; *Borovsky*, 2017]. The largest and most disruptive geomagnetic storms tend to be the longest lived [*Balan et al.*, 2016; *Echer et al.*, 2008; *Mourenas et al.*, 2018]. Many large and long-lived storms show a “two-step” development [*Tsurutani et al.*, 1999; *Xie et al*., 2006]; however, these multistep storms have been shown not originate from just a simple superposition of individual events [*Chen et al.*, 2000; *Kozyra et al.*, 1998, 2002] and it is not yet fully clear how the implied pre-conditioning originates. *Kozyra et al.* [1998] argued that prior energetic particle injections are swept out of the dayside magnetopause as the second population from the plasma sheet moves into the inner magnetosphere and so suggested that the preconditioning occurs in a multistep storm through the cumulative effects of the successive storms on the population in the source plasma sheet [*Chen et al.*, 2000; *Kozyra et al.,* 1998, 2002]. Alternatively, it has been suggested that prior storms prime the inner magnetosphere through O^+ ions injected from the ionosphere [*Hamilton et al.*, 1988; *Daglis*, 1997]. *Lockwood et al.* [2016] have shown that the key element in driving the largest storms (as measured by the *Dst* index) is not so much the peak magnitude of the interplanetary coupling function, rather the timescale over which it applies – large storms being a response to forcing that is both large and sustained over several days. (In other words, very large interplanetary coupling function values do not drive major storms if they persist for only short intervals). *Borovsky* (2016) reached the same conclusion in relation to the damaging relativistic electron fluxes generated in the largest storms. Thus there is likely to be some information in the core of the distributions that could be exploited to predict the occurrence of the long-lived and extreme events. Lastly, we also note that *Kauristie et al.* [2017] have also looked at the core distributions of *ap*, *Dst* (as well as *am* and d*Dst*/d*t*), not with a view to identifying highly disturbed periods and large and extreme events, rather the opposite - to find the quietest intervals that could be used to generate an empirical model of the undisturbed main field.

1.2 Construction of a Space-Weather Climatology

A number of techniques that have been developed and refined for terrestrial meteorological and climate studies are now being deployed in the field of space weather. In addition to EVS discussed above, these include: NWP (Numerical Weather Prediction) [*Pizzo et al*., 2015]; DA (Data Assimilation) [*Siscoe and Solomon,* 2006*; Schunk et al.,* 2014*; Barnard et al.,* 2017; *Lang*, 2017]; cost-loss analysis [*Owens and Riley*, 2017]; ensemble forecasts [*Knipp*, 2016]; climate analogue forecasts [*Barnard et al.,*2011]; ensemble climate reconstructions [*Owens et al*., 2016a; b], skill scores [*Balch,* 2008]; cost-loss analysis [*Henley and Pope*, 2017]; and several others. In meteorology, many of these techniques are used in conjunction with a “climatology” which describes statistically the probability of a relevant variable at key locations having one of the full potential range of values. “Climatological forecasts” assume that the future of a system can be determined from these statistical properties of the past behavior of that system. These will clearly often be rendered invalid by long-term changes in the system that are not covered by the climatology. This limitation to climatological forecasts can actually be useful because deviations from climatological forecasts (“anomalies”) can be used to detect and quantify the effects of the long-term changes. Note that long-term changes can also generate false conclusions about, for example, skill scores or event occurrence, if they are neglected [e.g., *Hamill and Juras*, 2006].

There are four elements that we need to generate a useful climatology of space weather for each of the key variables: (1) the mean value (over a convenient period such as a year); (2) the core distribution of values about that mean; (3) the extreme tail of the distribution (giving the repeat period of superstorms); and (4) the autocorrelation function, ACF. All these would be available to us, if we possessed the time series at high enough temporal resolution and over an interval long enough that adding any more data does not significantly alter the distribution. This approach has been employed by *Matthes et al.* [2017] to build a space climatology using the *aa* index geomagnetic that extends back to 1868. Unfortunately, as discussed below, this does not include the grand minima conditions such as existed during the Maunder minimum [*Usoskin et al*., 2015] that we know from cosmogenic isotopes to have prevailed for extended periods roughly 30 times in the last 9000 years [e.g., *Barnard et al.,* 2011]. These four elements would enable us to evaluate integrated deterioration of systems influenced by space weather, the probability of an event over a certain size and the probability of multiple events that may have a greater effect than the sum of the effect of the events individually. There is great emphasis in space weather on protecting systems from the largest events or, at least, evaluating the risk posed by those events. However, evaluating the distribution core and mean and the probabilities of quiet conditions is also important to avoid the cost and other wasted resources associated with "over-engineering" systems (such that they become obsolete long before they are lost or degraded) and so ensuring that the designs are cost-effective. As pointed out by *Henley and Pope* [2017], the development of a useful space weather climatology, as with forecasting procedures, requires a detailed dialogue with the system design engineers and end-users.

The biggest problem in trying to assemble a space-weather climatology is the long timescales of the variations [*Henley and Pope*, 2017]. The primary periodicity in space weather is the solar cycle oscillation the period of which averages about 11 years. Since *in-situ* observations of the near-Earth space environment began, we have accrued direct space-weather data for just four such cycles. To put this in context, consider a terrestrial tropospheric weather climatology: the dominant periodicity is one year and a climatology based on just four years would not be of much value for most applications. Hence, as pointed out by *Lockwood* [2003], we need to extend the interval by using other measurements and inferring the space weather variables, rather than just using the directly measurments.

The most direct way of doing this is to employ geomagnetic activity observations, as used by *Matthes et al.* [2017]. In theory these could extend back to 1832, when Gauss established the first well-calibrated geomagnetic observatory in Göttingen. Reviews of the development of the observation of geomagnetic activity have been given by *Stern* [2002] and *Lockwood* [2013]. Some composites have used geomagnetic activity data from soon after the establishment of Gauss' observatory; for example, *Svalgaard and Cliver* [2010] used regressions with different types of geomagnetic data to extend the sequence back to 1835. However, there are concerns about the calibration, stability and homogeneity of the earliest data [*Lockwood*, 2013].

Geomagnetic activity on annual timescales depends on both the solar wind speed V_{SW} and the IMF field strength, B, and the first separation of the two was made by *Lockwood et al.* [1999] using two different geomagnetic indices (the *aa* index and Sargent's recurrence index derived from *aa*). Later, *Lockwood* [2014] used 4 different pairings of different indices to derive V_{SW}, B and the open solar flux, with a full Monte-Carlo uncertainty analysis, back to 1845. From this date, the geomagnetic data give us almost 17 full solar cycles, considerably more useful than the 4 available in direct observations but still not enough for a full climatology that allows for centennial scale solar change. Crucially, this interval does not include the Maunder minimum (or even the lesser Dalton minimum) and hence a climatology based on geomagnetic data would not cover grand minimum conditions or even periods like the Dalton minimum.

Recent advances allow us to start to construct a climatology based on sunspot numbers which are available with reasonable regularity from about 1612, soon after the invention and patenting of the telescope in 1608. *Owens et al.* [2017] have used the sunspot number data in conjunction with modelling to reconstruct the solar wind number flux N_{SW}, as well as B and V_{SW} from 1615 onwards. This has enabled *Lockwood et al.* [2017] to reconstruct the annual mean power input into the magnetosphere from 1615 and from this *Lockwood et al.* [2018a] have estimated the annual means of the *ap* index. These advances make it possible to construct elements of a climatology which extends over 30 clear solar cycles as well as the 50-year break to normal solar cycles during the Maunder minimum. During the Maunder minimum, the modelling predicts 8 small-amplitude, smaller-period cycles which show a different phase relationship with the weak cycles in sunspot numbers. *Owens et al.* [2012] have shown evidence for these small Maunder-minimum cycles in galactic cosmic ray fluxes.

In addition to the increased number of solar cycles, these reconstructions that extend back to the early 17th century cover both a grand minimum (the Maunder minimum [*Usoskin et al*., 2015]) and the recent grand solar maximum [*Lockwood et al.*, 2009]. There is also potential to even extend the climatology to cover up to 9000 years, covering 24 grand maxima and 30 grand minima, using cosmogenic isotope abundance measurements which generally require decadal averages or which are smoothed by the time constants of the isotope deposition into the terrestrial reservoirs where they are measured. *Barnard et al.* [2011] have discussed a method for temporal scale-changing from these decadal-scale data to annual means. At the present time we are lacking one key element, namely a way to determine the times of solar cycle minimum and/or maxima and hence the phase of the solar cycle of each year.

In paper 1 of this series of 3 papers [*Lockwood et al*., 2018b], we showed that the total power input into the magnetosphere P_α can be computed using a constant coupling exponent α that does not depend on the averaging timescale τ (previous studies that had suggested it did were adversely influenced by data gaps). Paper 2 [*Lockwood et al*., 2018c] studied how the core distributions of P_α on timescales of 3 hours and less arise. In the current paper we study how and why these distributions in P_α evolve with averaging timescale τ and the subsequent evolution with τ of the *ap* (section 2) and *Dst* (section 3) geomagnetic indices. In each of these two sections we develop an algorithm that allows the core distribution for that geomagnetic index to be evaluated for a given mean value and at a required timescale, τ. The formulae required to implement these algorithms are given in Appendix A.

2. Distributions of power input to the magnetosphere and geomagnetic indices

Figure 3 studies the evolution with averaging timescale τ of the distribution of three space weather indicators. The left-hand panels show the power input into the magnetosphere, computed from the near-continuous interplanetary data for 1996-2016 (inclusive) and normalized to the mean value over the calendar year, $<P_\alpha>_\tau/<P_\alpha>_{1yr}$. The central panels show the normalized geomagnetic *ap* index, $<ap>_\tau/<ap>_{1yr}$ from the full dataset available (for 1932-2016) and the right-hand panels how the normalized negative geomagnetic *Dst* index, $<Dst'>_\tau/<Dst>_{1yr}$, (where *Dst'* is defined below), again using all the available data (for 1957-2016).

The coupling function of $\alpha = 0.44$, shown in Paper 1 [*Lockwood et al.,* 2018b] to apply at all τ, is used with the equation of *Vasyluinas et al.* [1982] to generate P_α [described in *Lockwood et al*., 2017, 2018a;b]. The *ap* index responds primarily to the substorm current wedge [see *Lockwood*, 2013] and the *Dst* index primarily to the ring current. However, *Dst* is importantly also influenced by other currents [e.g. *Turner et al*., 2000] such as the Chapman-Ferraro currents in the magnetopause and so also varies with compressions of the dayside magnetosphere by solar wind dynamic pressure enhancements. The ring current effect dominates meaning that *Dst* is increasingly negative as activity increases but the dynamic pressure effect mean that positive *Dst* value can occur. Corrections for the effect of solar wind dynamic pressure on *Dst*, via magnetopause currents, have been developed [*O'Brien and McPherron*, 2000; *Consolini et al.,* 2008] but we do not use them here, mainly because it reduces the available dataset to after 1996 (when quasi-continuous interplanetary data are available) but also because a great many papers have used the uncorrected *Dst* index to characterize magnetic storms in the past. The fact that *Dst*, unlike *ap* (or P_α), can have either sign generates a fundamental difference between the *ap* and *Dst* indices when trying to formulate a long-term climatology: when activity is low *ap* tends to a limiting value of zero whereas *Dst* tends towards a distribution of values spread around zero. Half-wave rectifying *Dst* so that positive values are put to zero is not an option as this generates a large number of samples at zero that distorts the distribution. Instead we here treat $Dst \geq 0$ as data gaps (we here call the index so derived *Dst'*) which yields an index that correlates much better with multiplicative interplanetary coupling functions [*Lockwood*, 2013]. However, such samples are still included in the total number when computing the occurrence probability of a large negative *Dst* value. Note that using *Dst'* instead of *Dst* is purely a measure that gives us a unipolar activity index to work with (which makes the modeling required much less complex) and is not, in any way, a correction for magnetopause currents. Of course, even strongly negative *Dst* values will still be influenced by magnetopause currents to some extent, which is why *Dst* is an imperfect metric of ring current storms. In a later paper we will present a separate model for predicting the distributions of the pressure-corrected index, *Dst**, as a function of τ. Note that *Dst** also has both positive and negative values (see Figures 1 and 2 of *Consolini et al.* [2008]) and so the same sort of techniques will be required for the construction of a model for *Dst** as are used here for *Dst*.

To summarize the procedure employed here: we make normalized values of the variable X, where X is one of the observed variables P_α, ap, and Dst for a given averaging timescale τ (also done for the synthesized variables X_R and X_{RF} that are used below to clarify the behavior of the observed variables). We normalise by dividing by the arithmetic mean for the calendar year of the sample $<X>_{\tau=1yr}$. From these normalized values we derive the distribution of $X/<X>_{\tau=1yr}$ for all 22 years studied. This distribution has an arithmetic mean $m = 1$ which is the "grand mean" or (the "mean-of-means") of the 22 annual normalized data subsets and which applies because we have, to a good approximation, the same number of samples in each year. We then fit model p.d.f.s so that we can empirically model the probability of $X/<X>_{\tau=1yr}$ which is the probability of X for a given $<X>_{\tau=1yr}$, i.e. $P(X\,|<X>_{\tau=1yr})$. Hence this enables us to achieve our goal of empirically modelling the distribution of X for a given $<X>_{\tau=1yr}$. We wish this fitted distribution to reproduce the observed one as closely as possible so we use model distributions of means of $\mu = m = 1$ and find the optimum variance v using Maximum Likelihood Estimation. Some of the distributions fitted are described by shape and scale parameters instead of μ and v and these are constrained so that μ is unity. The procedure is repeated for the full range of averaging timescales, τ.

The blue histograms in Figure 3 are the observed distributions, the black lines shows the best-fit lognormal distributions and the mauve lines are the best-fit Weibull distributions (both with mean value $\mu = 1$ in the cases of P_α and ap and $\mu = R_m(\tau)$ for Dst' (where R_m deviates from unity because in Dst' we treat each $<Dst>_\tau \geq 0$ sample as a data gap: the factor $R_m(\tau)$ is discussed further later). The blue histograms were generated by counting the number of samples in 150 contiguous bins centered on on $k.x_{98}/100$,where k is varied between 0.5 and 149.5 in steps of 1 and x_{98} is the 98th percentile of the distribution. The numbers of samples n in each bin then normalized so that $\Sigma n(x_{98}/100)$ is unity. Fitting directly a distribution to these histograms gives results which, in general, depend on the bin width adopted [e.g., *Woody et al.,* 2016] and so we here fit distributions using Maximum Likelihood Estimation (MLE) which does not require prior binning of the data into bins of arbitrarily-chosen width. A basic description of MLE fitting, and of goodness of fit metrics (both absolute and relative) is given in the Supporting Information file. Plots of best-fit probability density functions (p.d.f.s) and cumulative distribution functions (c.d.f.s), and tables of best-fit distribution parameters and goodness of fit metrics are also given in the Supporting Information file for seven standard distribution forms: the normal (Gaussian) distribution, the Lognormal distribution, the

Weibull distribution; the Burr distribution, the Gamma distribution, the Log-logistic (Fisk) distribution, and the Rician distribution. For all these distributions the number of degrees of freedom is $d_f = 2$, except the Burr for which $d_f = 3$.

The top row in Figure 3 is for averaging timescale τ = 1yr and the rows beneath are, successively for τ of 0.5 year, 27 days, 7 days, 1 day and 3 hours (0.125 day). The omission of positive $<Dst>_\tau$ samples has no effect for τ = 1 year (as all values are negative), but the number of Dst' samples is 99.17%, 94.08%, 88.42%, 80.60%, and 78.48% of all *Dst* samples for τ of, respectively, 0.5 year, 27days, 7 days, 1 day and 6 hours. Because of the normalization, the distributions for τ = 1yr are, by definition, delta functions at unity. At general τ, the distributions for $<ap>_\tau/<ap>_{1yr}$ are always close to lognormal in form (the black lines) the variance increasing with decreasing τ (see Supporting Information file for goodness-of-fit evaluations). At the larger τ, the low variance lognormal distributions are essentially Gaussian in form. On the other hand, the Dst' distributions are equally well fitted by the Weibull, Gamma or Log-logistic families of distributions (see Supporting Information) and in Figure 3 we show the Weibull distributions (the mauve lines), again with variance increasing with decreasing τ. Note that for Dst', significantly better fits could be obtained using a distribution with an extra degree of freedom, such as the Burr (see supporting information). The difference between *ap* and Dst' is caused by the flatter and broader distribution at small $< Dst' >_\tau/<Dst>_{1yr}$ values. The $<P_\alpha>_\tau/<P_\alpha>_{1yr}$ distributions are lognormal in form for τ greater than about 2 days, but at lower τ these distributions are increasingly Weibull-like in form. The origin of a Weibull form at low τ was discussed in Paper 2 [*Lockwood et al*., 2018c] and is associated with the variability of the Interplanetary Magnetic Field (IMF) orientation factor on these timescales, via the quasi "half-wave rectification" effect of the southward component of the IMF on solar wind – magnetosphere coupling. Note that because of the smoothing effect of the magnetospheric energy storage/release system, the Weibull distribution of power input to the magnetosphere for small τ yields a log-normal distribution in power input on the timescales relevant to *ap* and hence in *ap* itself.

The evolution of the distributions shown by the different rows of Figure 3 reveal the "Central Limit Theorem" (hereafter CLT) in action [*Heyde*, 2006; *Fischer*, 2011; *Wilks*, 1995]. This states that when independent random variables are added, their properly normalized sum tends toward a normal distribution. It applies in this context because the key operation in taking an average value is summation and because, as τ is increased in relation to the correlation timescale, an increasing fraction of the samples are independent.

2.1. The evolution of the distributions with timescale for *ap* and $P\alpha$

Figure 4 looks in more detail at the evolution of the distributions of $<P_\alpha>_\tau/<P_\alpha>_{1yr}$ (for $\alpha = 0.44$) as a function of the logarithm of the averaging interval. The upper plot shows the probability density function (pdf) color-coded as a function of $\log_{10}(\tau)$ and $<P_\alpha>_\tau/<P_\alpha>_{1yr}$ such that the distributions shown in the left-hand plots of Figure 3 are vertical slices of Figure 4. The blue line in the lower panel shows the corresponding variation of the distribution variance *v* (also on a logarithmic scale). Figure 5 is the corresponding plot for $<ap>_\tau/<ap>_{1yr}$.

In the Supporting Information file, the distributions shown in Figure 3 are fitted with seven distribution forms, six or which are characterized by two parameters (either the mean, *m*, and variance, *v*, or a pair of parameters that are defined by *m* and *v*). (Note the seventh distribution form used, the Burr, has an additional shape parameter and is included to test if this gives a statistically significant improvement to the fit). Two of the distributions, the Gaussian and the Rician, do not give good fits at low τ but do quantify the evolution of the distributions towards a Gaussian-like form as τ is increased towards 1 year. Because we here look at the distributions of normalized disturbance metrics $<X>_\tau/<X>_{1yr}$ (in this paper we consider *X* of P_α, *ap* and *Dst*) the mean *m* is, by definition, always unity and hence we only need to study the behavior of the variance, *v*, shown in Figure 4b for $<P_\alpha>_\tau/<P_\alpha>_{1yr}$ and in Figure 5b for $<ap>_\tau/<ap>_{1yr}$.

2.2. The effect of autocorrelation on the evolution of distributions

To help understand Figures 4 and 5, Figure 6 shows the evolution with increased τ for a synthesised variable X_R that is selected at random at time resolution $\tau = 3$ hrs from a Weibull distribution with *k* of 1.0625 and λ of 1.0240 (giving a mean $m = 1$) which in Paper 2 [*Lockwood et al.,*2018c] was shown to be good fit to the distribution of $<P_\alpha>_\tau/<P_\alpha>_{1yr}$ at that timescale. The general pattern of evolution of the pdfs of $<X_R>_\tau/<X_R>_{1yr}$ in Figure 6a is like that in Figure 4a and 5a, other than that the distributions evolve towards a delta function at unity with increasing τ rather more rapidly for

X_R. This is also reflected by the mauve line in figure 6b, which shows that the variance, v, falls more rapidly than the blue and red lines in Figures 4b and 5b for P_α and *ap*, respectively. The initial distribution in Figure 6 is a Wiebull form but even at τ as low as 9 hrs it has evolved into a lognormal form, which it keeps at all greater τ (but the variance falls so it approaches a Gaussian near $\tau = 1$ yr). This evolution of the distribution form is the same sequence that $P\alpha$ follows.

The mauve line in Figure 7 shows the autocorrelation function (the autocorrelation at lags of 3 hours, the resolution of the synthetic data) of the random variable X_R employed in Figure 6. It can be seen that X_R is indeed completely random at the autocorrelation function falls to zero at lag 1. To investigate the effect of autocorrelation we generate a second random distribution which we then pass through a smoothing filter to give it autocorrelation. This generates a synthetic data series X_{Rf}. Because the filter has a similar effect on the distribution as averaging we have to draw the original random distribution from a higher-variance Weibull. By iteration we find that for the filter we use, an initial Weibull random distribution with k of 0.2800 and λ of 0.0778 (giving $m = 1$) generates an almost identical distribution at $\tau = 3$hr after filtering to that of X_R used in Figure 6. The filter used is a triangular-weighting moving-average filter with two response peaks. The first is a [1-3-5-3-1], around lag $\delta t = 0$ which adds short-range correlation into the X_{Rf} data series. The second is a [1-2-3-4-5-6-7-8-7-6-5-4-3-2-1]×(5/8) triangular response peak centered on lag 216 (for the 3-hour resolution X_{Rf} data series, this second peak is at lag 27 day). The black line in Figure 7 shows the autocorrelation function of X_{Rf} and it can be seen that the filter has introduced short-term autocorrelation on lags up to about 1 day, and a 27-day (the mean solar rotation period seen from earth) recurrence.

Figure 8 shows the equivalent plot to Figure 6 for the X_{Rf} data series. Figure 8a shows that the effect of the autocorrelation is to slow the progression towards the delta function at unity. This is expected from the CLT because the autocorrelation means that larger averaging timescales are needed before samples are sufficiently uncorrelated for the CLT to apply. Figure 8b shows the variation of the variance v for X_{Rf} in black, and compares it with that for X_R (in mauve) from Figure 6b. It can be seen that at the τ where autocorrelation has been introduced into the X_{Rf} series by the filter, the variance fall less quickly than for the random series, X_R. At all τ the distribution of X_{Rf} is lognormal in form and mirrors the evolution for *ap*. Note that Figures 7 and 8, and the results for a random and a smoothed-random data series (X_R and X_{Rf}), are included here only to illustrate how autocorrelation influences the form of the evolution of the distribution with τ and also influences the dependence of variance v on τ. They are not used again in the derivation of a model of the distribution at a given τ. Instead we fit the $v(\tau)$ variation derived directly from data with a polynomial in τ .

2.3. Modelling the evolution of distribution of *ap* with increasing timescale

The section describes how we model the evolutions of the distributions of $<ap>_\tau / <ap>_{1yr}$ with increasing τ and Figure 9a presents the results for that modelling, aimed at reproducing Figure 5a. Figure 9(b) shows the log-log plots of variance v, as a function of τ from Figures 4b, 5b and 6b using the same color scheme, i.e. for P_α in blue, for *ap* in red and for the random variable, X_R in mauve. Also shown, in cyan, is the variation for the 150-year data series of the *aa* geomagnetic index. The black line is a polynomial fit to the *ap* variation, given by equations (A11) and (A12) of Appendix A which yield the variance, $v(\tau)$. The Maximum Likelihood analysis given in the Supporting Information (on which Figure 3 is based) shows that for $<ap>_\tau / <ap>_{1yr}$ the observed distribution at all τ is best fitted with a lognormal form with mean $m = 1$. (That is until τ approaches 1 year when the distribution becomes nearly Gaussian in form and the goodness-of-fit metrics for all 7 distributions become very similar). Figure 9a shows the modelled lognormal distributions using the polynomial fit to the variance variation shown in Figure 9b. The equations for reproducing the distribution for a given τ are given in part (i) of Appendix A. From this, the p.d.f. of $<ap>_\tau$ (and hence that of the time-integral of the activity $\tau<ap>_\tau$) at a given τ can be computed for a known annual mean $<ap>_{1yr}$.

The cyan line in Figure 9b is for all the full *aa* index data set which covers the interval 1868-2017. The close similarity of the $v(\tau)$ relationship to that for the *ap* data (1932-2017, the red line) strongly indicates that this relationship has not varied significantly over the past 150 years. To check this in more detail, the *aa* data have been divided into three 50-year intervals (1868-1917, 1918-1967 and 1968-2017, inclusive), and the $v(\tau)$ relationship for these three data subsets are plotted in Figure

10b as green, blue and red lines, respectively, and can be seen to be very similar (and to that for the overall *aa* plot in Figure 9b). Figure 10a studies the autocorrelation function (ACF) of the $aa/\langle aa\rangle_{\tau=1\mathrm{yr}}$ data for these three intervals. The three are again very similar showing the persistence effect at low τ (up to about 5 days), a recurrence peak at 27 days, plus some weak harmonics of the 27 day variation, and hence are very similar to that for the smoother random variable, X_{Rf}, in Figure 7. In fact, the ACF for X_{Rf} could easily be made to match the observed ACFs for *aa* shown in figure 10 very closely, if the smoothing filter used were adjusted to give slightly lower persistence at low τ (< 1day) and the response peak around 27 days were to be broadened somewhat. There is also a small but marked and persistent diurnal signal visible in Figure 10a. The main difference between the three intervals is that the 27-day peak is a little bit larger for the earliest interval (1868-1917) and the low-τ persistence a little bit weaker. These differences cannot be identified in the $v(\tau)$ plots. The only other data that are continuous and high enough time resolution to potentially investigate this further back in time are the daily values international sunspot number *R,* which are almost continuous since 1818. However, sunspot numbers behave very differently to geomagnetic activity indices, showing sudden increases/decreases as spot groups rotate onto onto/off the visible disk of the Sun and rises and falls as the groups wax or wane as they rotate across the visible solar disc: they do not have the bursty nature of Earth-directed interplanetary disturbances and hence of geomagnetic disturbances. Hence they cannot help us investigate the ACF, and the associated $v(\tau)$ relationship for near-Earth space and geomagnetic activity before the start of regular, well-calibrated geomagnetic observations.

Figure 11 investigates if ACFs and variances for *aa* shown in figure 10 vary with sunspot number. We use the international sunspot number *R*, derived and distributed by WDC-SILSO, Brussels. We take 3-year averages of the data to keep sample numbers high. For each period we evaluate the mean sunspot number, $\langle R\rangle_{\tau=3\mathrm{yr}}$, and the ACF of $aa/\langle aa\rangle_{\tau=1\mathrm{yr}}$. These ACFs were then averaged together for contiguous bins of $\langle R\rangle_{\tau=3\mathrm{yr}}$ that are centred on values between 10 and 200 in steps of 20. In addition the variance v of the distribution of all $\langle aa\rangle_{\tau}/\langle aa\rangle_{\tau=1\mathrm{yr}}$ samples in each band of $\langle R\rangle_{\tau=3\mathrm{yr}}$ was computed for each averaging timescale τ. The top panel of Figure 12 shows a surface plot of the ACF as a function of $\log_{10}(\tau)$ and $\langle R\rangle_{\tau=3\mathrm{yr}}$. On timescales below about $\tau = 25$ days the ACFs hardly varies at all with the sunspot number. The major effect is on the peak at 27 days (and its harmonics) which has a larger amplitude when the sunspot number is low. The lower panel gives the corresponding surface plot of $\log_{10}(v)$: note that sample numbers do not allow this analysis to extend to as great a sunspot number as for the ACF analysis. As would be expected from the ACFs, there is almost no variation in the v-τ relationship with sunspot number at τ below about 25 days but above this the larger ACF peak at 27 days for low sunspot number causes v to fall with τ slightly less rapidly than it does at higher sunspot numbers. There are some slight but persistent ridges and dips in the surface shown in figure 11b at certain $<R>$ but the surface is remarkably independent of R. Note that the lack of any dependence of the v-τ relationship on sunspot number (at low τ) was also revealed by Figure 8c of *Lockwood et al.* [2018a], which plots distributions of $<aa>_{\tau=1\text{day}}/<aa>_{\tau=1\text{yr}}$ as a function of year and no solar cycle variation can be detected.

It is tempting to argue that we should modify the model form of the v-τ relationship at $\tau > 25$ days to allow for the (weak) sunspot number variation seen at large τ in the lower panel of Figure 12. The major reason is that during the Maunder minimum the persistently low sunspot number might make this a factor. However, this is not necessarily the case because a prolonged (grand) sunspot activity minimum is in many ways quite different to a sunspot activity minimum between solar cycles: one major reason being that for the cycle minima there is residual open flux generated during the previous cycle out of which fast solar wind flows. The 27-day ACF peak is largely caused by CIRs (Co-rotating Interaction Regions) caused by fast-solar wind emanating from coronal holes reaching down to low latitudes, catching up with Earth-bound slow solar wind of the streamer belt. Modelling for the Maunder minimum predicts that the streamer belt will have been considerably wider than in modern times with coronal holes restricted to high heliographic latitudes [*Lockwood and Owens*, 2014a; b; *Owens et al.,* 2017], making CIRs that hit Earth less, rather than more, common. Hence it is not at all clear that that the effect noted in low sunspot years at $\tau > 25$ days in Figure 12 will also apply to the Maunder minimum. For the present paper we assume that the $v(\tau)$ relationship does not change and we fit it with a single polynomial form. However, should a long-term changes in the $v(\tau)$ relationship be discovered at some point in the future, it could be readily accommodated by making the fit polynomial coefficients a function of time.

Figure 12 shows that the modelled distributions shown in Figure 9a can explain the variation of occurrence of large events, as a function of the annual means discussed in Paper 2. The points in Figure 12a show probability that 3-hourly values of *ap* are in the *ap* top 5% of the overall distribution (for 1932-2016, 252152 samples), $f[ap>ap\mathrm{o}]$ (i.e., *ap* exceeds its 95-percentile of 3-hourly *ap* values, $ap\mathrm{o} = 47.91$), as a function of the annual mean value $<ap>_{\tau=1\mathrm{yr}}$. The mauve line is the prediction for τ=3hrs for the model values displayed in Figure 9a. The fit can be seen to be close. The family of model predictions of $f[ap>ap\mathrm{o}]$ as a function of $<ap>_{\tau=1\mathrm{yr}}$ is shown in Figure 12b for timescales of 1day (in blue), 7 days (in orange) and 27 days (in black). Hence the model is reproducing the behavior noted in Figure 1 of Paper 2, namely that, with some scatter, the number of events in any one year that are in the top 5% of the overall distribution, increases hyperbolically with the mean value for that year.

2.4. The evolution of the distributions with timescale for *Dst'*

Figure 13 is the equivalent plot to Figure 4 for the *Dst'* data which extend from 1957-2016. Here the pdf is shown as a function of τ and $< Dst' >_{\tau} / <\mathrm{Dst}>_{1\mathrm{yr}}$. Generating a model fit to this plot is more complex because *Dst* do not converge to zero for low activity and we have to use *Dst'* instead, where *Dst'* is the same as *Dst*, but all positive values are treated as data gaps. In annual mean data, this makes no difference, because all annual means are negative, but with decreasing τ the number of *Dst'* samples falls compared to the number of *Dst* samples, and the mean R_m of the distribution of $< Dst' >_{\tau}/<Dst>_{1\mathrm{yr}}$, although unity at τ = 1 year, is greater than unity at lower τ because negative values of $< Dst' >_{\tau}/<Dst>_{1\mathrm{yr}}$ (i.e., positive values of $< Dst' >_{\tau}$) are neglected. Figure 14a shows in red the variation with $\log_{10}(\tau)$ of f_neg ($= N_{Dst'}/N_\mathrm{Dst}$), the fraction of *Dst* samples that are negative (the subset termed *Dst'*). The black line is a polynomial fit to this variation which is given by equation (A12) of appendix A. The green line shows the corresponding variation of R_m, the mean of $< Dst' >_{\tau}/<Dst>_{1\mathrm{yr}}$. Again the black line is best polynomial fit given by equation (A13) of Appendix A. Appendix A-ii gives the algorithm for computing the pdf of *Dst'* or a given *Dst* and timescale τ which allow for these two factors. Figure 15 corresponds to Figure 9 for the *Dst* index. As shown by Figure 3, the distributions of $< Dst' >_{\tau}/<Dst>_{1\mathrm{yr}}$ follow the Weibull family of distributions and these are derived from the best fit to the observed $\log_{10}(\nu)$-$\log_{10}(\tau)$ variation (shown in green in Figures 13b and 15b), using the polynomial fit given in black which is given by equations (A10) and (A11) of Appendix A. For comparison, Figure 15b also shows the $\log_{10}(\nu)$-$\log_{10}(\tau)$ variations for P_α (in blue), *ap* (in red) and the random variable, X_R (in mauve).

Figure 16 corresponds to figure 12 and shows how the model can reproduce the occurrence of *Dst* below its 95 percentile value (*Dst*o = −55.142nT), as a function of the annual mean value. Figure 16b shows the family of such variations for different values of τ.

3. Discussion and Conclusions

It is noticeable that the $\log_{10}(v)$-$\log_{10}(\tau)$ variation for *ap* (in red in Figure 9b) flattens off as averaging timescale τ falls below about 1 day, whereas the variance *v* continues to rise with decreasing τ for power input into the magnetosphere, P_α (in blue). Using a synthesized random time series and a filter we have demonstrated how the flattening off is caused by autocorrelation in the time series. Hence there is autocorrelation in the *ap* time series at τ between 3hrs and 1day that is greater than that in P_α. As P_α is the driver of *ap*, this means that the geomagnetic response seen in *ap* is a smoothed response. This is not surprising, given the currents that the index is sensitive to and their associated time constants. The *ap* index is primarily influenced by the substorm current wedge [*Lockwood*, 2013] which is initiated only after a substorm growth phase lasting typically 30-40 minutes. Hence the rapid variations in the energy input into the magnetosphere, which are mainly associated with IMF orientation changes, are smoothed as energy (and open magnetic flux) are accumulated in the tail.

The same effect is even more clear for *Dst*, for which *v* flattens off as τ falls below about 3 days (the green line in Figure 15b which is again compared to the behavior for P_α in blue). Hence the smoothing effect on the response of *Dst* has a longer time constant than that for *ap*. The (negative) *Dst* index is responding primarily to the ring current [*Turner et al*., 2000] which shows greater time constants, responding to the integral of solar wind forcing on timescales of order of a day or more [*Lockwood et al.,* 2016; *Borovsky*, 2016]. (Note that below we discuss the implications of the fact that even large negative *Dst* can be influenced by other factors, in particular, the magnetopause currents). This is not to say that P_α is the best coupling function explaining the solar wind influence on the ring current, not least because the coupling exponent α has been tuned to 0.44 to make P_α reproduce *ap*, not *Dst*. Nevertheless, the importance of southward IMF in driving disturbed *Dst* means that the same conclusions would be valid for any other coupling function that might better predict *Dst*.

Breaking down the power input into the magnetosphere P_α into its component factors, Paper 2 showed that the factors dependent on solar wind velocity and mass flux and on the IMF (F_V, F_N and F_B) do not vary much on short timescales and the distribution of power input into the magnetosphere is set by the variation in the IMF orientation factor F_θ which, although it can stay stable for several days, is typically changing on minute timescales. Thus the shape of distribution is set by F_θ, at very short timescales, much shorter than the timescale of the geomagnetic index response – it then evolves with τ according to the CLT, making the shape of the distribution a function of τ only.

A climatology is a statistical description that would enable us to evaluate the probability of space weather events of a given magnitude and we are working toward one that applies to the full range of solar conditions from grand solar minimum to grand solar maximum. In particular, there is value in knowing the integrated level of activity over an extended period τ, which equals the average value times the duration. Hence we investigate algorithms that can give us the probability of a given average value for a given τ. These algorithms will be of great value in generating a long-term climatology because they can compute the probabilities for a given annual mean and we have annual means from the past 400 years from recent modelling work based on telescopic sunspot observations [Owens et al., 2017]. The approach outlined in this paper is based on the finding that the shape of the distribution of the normalized values (normalized by dividing by the annual mean value) only depends on the averaging timescale τ. This was used by *Lockwood et al.* [2018a] to look at the occurrence of "large" events (defined as in the top 5% since records began) over 400 years. The constancy of the shape of the distributions was just taken by *Lockwood et al.* [2018a] as an empirical observation that could be exploited. The present series of three papers provide greater understanding of why this empirical result applies and why the distributions have the form that they do. This is important because it means the result can be applied with greater confidence to periods when inference are only made from proxy data, and in particular, to grand minima like the Maunder minimum.

We have developed methods that enable computation of the core distribution of both the *ap* and (negative) *Dst* geomagnetic indices for a given annual mean value at a required averaging timescale τ. The algorithms for doing this are detailed in parts (i) and (ii), respectively, of Appendix A. The complications caused by the fact that the *Dst* index, unlike *ap*, does not tend to zero when activity is quiet have led to the algorithm for *Dst* being somewhat more involved than that for *ap*, and the distributions are best fitted with a Weibull family of distributions, as opposed to the lognormal family for *ap*.

The model distributions for the *ap* index make use of the lognormal form which, as shown in the Supporting Information, gives the best MLE fit of all the distribution forms with two free parameters. The Burr distribution gives slightly better fits according to the absolute goodness-of-fit metrics (least squares and modified Kolmogorov–Smirnov) but the relative metrics that allow for the degrees of freedom (AIC and BIC) show the extra degree of freedom is not justified. (Note that as τ approaches one year and the observed distribution tends towards a Gaussian all the distributions are good fits and differences are minimal). Thus there is no question that the *ap* model employs the best form of distribution (i.e., the lognormal). The model is also relatively straightforward because the *ap* index is unipolar and tends to zero at the quietest activity levels. The largest uncertainty in using the model in even the Maunder minimum relates to the occurrence of CIRs and recurrent disturbances which may influence the model at averaging timescales τ greater than about 25 days.

For the *Dst* model these considerations are less straightforward. Firstly, the Weibull, Gamma and log-logistic distributions all perform similarly, and none of them are ideal fits to the observed distribution. Furthermore, the extra degree of freedom of the Burr distribution gives fits that are better by a statistically significant degree. This means the added complexity of using two shape parameters (in addition to the mean $m = 1$) would be worthwhile. However, at this point it is worth remembering that the *Dst* index is, intrinsically, and imperfect metric and hence the additional fit accuracy is unlikely to justify the additional complexity. Hence we propose, in a later paper, to generate a model for the pressure-corrected index *Dst**. Because *Dst** can, like *Dst*, have both positive and negative values and approach similar to that adopted here for Dst will be needed.

Acknowledgments and Data

The authors are grateful to the staff of Space Physics Data Facility, NASA/Goddard Space Flight Centre, who prepared and made available the OMNI2 dataset used. The data were downloaded from http://omniweb.gsfc.nasa.gov/ow.html. They are also grateful to the staff of GeoForschungsZentrum (GFZ) Potsdam, Adolf-Schmidt-Observatorium für Geomagnetismus, Niemegk, Germany who generate the *ap* data. The *ap* and aa data were downloaded from the UK Space Science Data Centre from https://www.ukssdc.ac.uk/ with updating of recent data from BGS

Edinburgh http://www.geomag.bgs.ac.uk/data_service/data/magnetic_indices/apindex.html. The international sunspot data were compiled and made available by WDC-SILSO, Royal Observatory of Belgium, Brussels http://www.sidc.be/silso/ . The work presented in this paper is supported by STFC consolidated grant number ST/M000885/1, the work of ML and MJO is also supported by the SWIGS NERC Directed Highlight Topic Grant number NE/P016928/1/ and of OA by NERC grant NE/P017274/1. SB is supported by NERC as part of the SCENARIO Doctoral Training Partnership NE/L002566/1.

References

Allen, J. H. (1982) Some Commonly Used Magnetic Activity Indices: Their Derivation, Meaning, and Use, in Proceedings of a Workshop on Satellite Drag, March 18-19, 1982, Boulder, Colorado, Ed. J.A.C. Joselyn, pp. 114-134

Baker, D.N. (2000) The occurrence of operational anomalies in spacecraft and their relationship to space weather. *IEEE Transactions on Plasma Science*, **28**(6), 2007-2016, doi: 10.1109/27.902228

Baker, D. N., X. Li, A. Pulkkinen, C. M. Ngwira, M. L. Mays, A. B. Galvin, and K. D. C. Simunac (2013), A major solar eruptive event in July 2012: Defining extreme space weather scenarios, *Space Weather,* **11**, 585–591, doi:10.1002/swe.20097

Balan, N., I. S. Batista, S. Tulasiram, and P. K. Rajesh (2016) A new geomagnetic storm parameter for the severity of space weather, *Geosci. Lett.*, **3**, 3, doi:10.1186/s40562-016-0036-5.

Balch, C. C. (2008) Updated verification of the Space Weather Prediction Center's solar energetic particle prediction model, *Space Weather*, **6**, S01001, doi:10.1029/2007SW000337

Barnard, L.A., M. Lockwood, M.A. Hapgood, M.J. Owens, C.J. Davis, and F. Steinhilber (2011) Predicting Space Climate Change, *Geophys. Res. Lett*., **38**, L16103, doi: 10.1029/2011GL048489

Barnard, L. A., C. A. de Koning, C. J. Scott, M. J. Owens, J. Wilkinson, and J. A. Davies (2017), Testing the current paradigm for space weather prediction with heliospheric imagers, *Space Weather*, 15, 782–803, doi: 10.1002/2017SW001609.

Beirlant, J., Y. Goegebeur, J. Segers, and J. Teugels (2004), Statistics of Extremes, John Wiley, Chichester, U.K., doi: 10.1002/0470012382

Borovsky, J.E. (2017) Time-integral correlations of multiple variables with the relativistic electron flux at geosynchronous orbit: The strong roles of substorm injected electrons and the ion plasma sheet, *J. Geophys. Res.*, **122**, doi:10.1002/2017JA024476.

Boteler, D.H. (2000) Geomagnetic effects on the pipe-to-soil potentials of a continental pipeline, In *Adv. Space Res*., **26** (1), 15-20, doi: 10.1016/S0273-1177(99)01020-0

Chapman, S.C., N. W. Watkins, and E. Tindale, (2018) Reproducible aspects of the climate of space weather over the last five solar cycles, *Space Weather*, **16** (8), 1128–1142. doi: 10.1029/2018SW001884

Chen, M. W., L. R. Lyons, and M. Schulz (2000) Stormtime ring-current formation: A comparison between single-and double-dip model storms with similar transport characteristics, *J. Geophys. Res*., **105** (A12), 27755–27765, doi:10.1029/1999JA000440.

Cliver E.W. and W.F. Dietrich (2013), The 1859 space weather event revisited: limits of extreme activity, *J. Space Weather Space Clim*., **3**, 2013, A31, pp. 1-15, doi: 10.1051/swsc/2013053

Cliver, E.W. and L. Svalgaard (2004) The 1859 solar–terrestrial disturbance and the current limits of extreme space weather activity, *Solar Physics*, **224** (1-2), 407-422, doi: 10.1007/s11207-005-4980-z

Cole, D.G. (2003) Space weather: its effects and predictability, *Space Science Reviews*, 107 (1/2), 295-302, doi: 10.1023/A:1025500513499

Coles, S. (2004) An Introduction to Statistical Modelling of Extreme Values, Springer, London

Consolini, G., P. De Michelis, and R. Tozzi (2008), On the Earth's magnetospheric dynamics: Nonequilibrium evolution and the fluctuation theorem, *J. Geophys. Res.,* **113**, A08222, doi:10.1029/2008JA013074.

Daglis, I. A. (1997) The Role of Magnetosphere-Ionosphere Coupling in Magnetic Storm Dynamics, in *Magnetic Storms* (eds B. T. Tsurutani, W. D. Gonzalez, Y. Kamide and J. K. Arballo), American Geophysical Union, Washington, D. C.. doi: 10.1029/GM098p0107

Dmitriev, A. V., A.V. Suvorova, and I.S. Veselovsky (2009) Statistical Characteristics of the Heliospheric Plasma and Magnetic Field at the Earth's Orbit during Four Solar Cycles 20-23, in Handbook on solar wind; in *Handbook on Solar Wind: Effects, Dynamics and Interactions*, Ed. Hans E. Johannson, NOVA Science Publishers, Inc., New York, p. 81-144. eprint arXiv:1301.2929

Doornbos, E., and H. Klinkrad (2006) Modelling of space weather effects on satellite drag, *Adv. Space Res.*, **37** (6), 1229-1239, doi: 10.1016/j.asr.2005.04.097

Echer, E., W. D. Gonzalez, and B. T. Tsurutani (2008) Interplanetary conditions leading to superintense geomagnetic storms (Dst ≤ -250 nT) during solar cycle 23, *Geophys. Res. Lett.*, **35**, L06S03, doi:10.1029/2007GL031755.

Farrugia, C.J., B. Harris, M. Leitner, C. Möstl, A.B. Galvin, K.D.C. Simunac, R.B. Torbert, M.B. Temmer, A.M. Veronig, N.V. Erkaev, A. Szabo, K.W. Ogilvie, J.G. Luhmann, and V.A. Osherovich (2012) Deep Solar Activity Minimum 2007 – 2009: Solar Wind Properties and Major Effects on the Terrestrial Magnetosphere, *Solar Phys*, **281**, 461–489, 2012, doi: 10.1007/s11207-012-0119-1

Fischer, H. (2011) A History of the Central Limit Theorem, from Classical to Modern Probability Theory*,* book in series "*Sources and Studies in the History of Mathematics and Physical Sciences"*, Springer, New York, doi:10.1007/978-0-387-87857-7

Fleetwood, D.M. P.S. Winokur, P.E. Dodd (2000) An overview of radiation effects on electronics in the space telecommunications environment, *Microelectronics Reliability*, 40 (1), 17-26, doi: 10.1016/S0026-2714(99)00225-5.

Gaunt, C. T. (2016), Why Space Weather Is Relevant to Electrical Power Systems, *Space Weather*, **14**, 2–9, doi:10.1002/2015SW001306.

Gummow, R.A. (2002) GIC effects on pipeline corrosion and corrosion control systems, *J. Atmos. and Sol.-Terr. Phys*., **64** (16), 1755-1764, doi: 10.1016/S1364-6826(02)00125-6.

Hamilton, D. C., G. Gloeckler, F. M. Ipavich, W. Stüdemann, B. Wilken, and G. Kremser (1988) Ring current development during the great geomagnetic storm of February 1986, *J. Geophys. Res*., **93** (A12), 14343–14355, doi: 10.1029/JA093iA12p14343.

Hapgood, M.A., G. Bowe, M. Lockwood, D.M. Willis, and Y. Tulunay (1991) Variability of the interplanetary magnetic field at 1 A.U. over 24 years: 1963 – 1986, *Planet. Space Sci.*, **39**, 411-423, doi: 10.1016/0032-0633(91)90003-S

Hamill, T. M. and Juras, J. (2006) Measuring forecast skill: is it real skill or is it the varying climatology?, *Q.J.R. Meteorol. Soc.*, **132**, 2905–2923, doi: 10.1256/qj.06.25

Henley, E. M., and E.C.D. Pope (2017) Cost-loss analysis of ensemble solar wind forecasting: Space weather use of terrestrial weather tools, *Space Weather*, **15**, doi: 10.1002/2017SW001758

Heyde, C. (2006) Central Limit Theorem, in *Encyclopedia of Actuarial Science*, 1, John Wiley and Sons, doi: 10.1002/9780470012505.tac019

Ingham, M., & Rodger, C. J. (2018). Telluric field variations as drivers of variations in cathodic protection potential on a natural gas pipeline in New Zealand. *Space Weather*, **16**, 1396-1409. https://doi.org/10.1029/2018SW001985.

Kappenman, J. G. (2005), An overview of the impulsive geomagnetic field disturbances and power grid impacts associated with the violent Sun-Earth connection events of 29–31 October 2003 and a comparative evaluation with other contemporary storms, *Space Weather*, 3, S08C01, doi: 10.1029/2004SW000128.

Kappenman, J. G. and Radasky, W. A. (2005), Too Important to Fail, *Space Weather*, 3: doi:10.1029/2005SW000152

Kauristie, K., A. Morschhauser, N. Olsen, C.C. Finlay, R.L. McPherron, J.W. Gjerloev, and H.J. Opgenoorth (2017) On the Usage of Geomagnetic Indices for Data Selection in Internal Field Modelling, *Space Sci. Rev.,* **206** (1-4), 61-90, doi: 10.1007/s11214-016-0301-0

Knipp, D. J. (2016), Advances in Space Weather Ensemble Forecasting, *Space Weather*, **14**, 52–53, doi: 10.1002/2016SW001366.

Koons, H. C. (2001), Statistical analysis of extreme values in space science, *J. Geophys. Res*., **106**, 10,915–10,921, doi: 10.1029/2000JA000234

Kotz, S. and S. Nadarajah (2000) Extreme Value Distributions: Theory and Applications, World Scientific, doi: org/10.1142/p191

Kozyra, J. U., J. E. Borovsky, M. W. Chen, M.-C. Fok, and V. K. Jordanova (1998) Plasma sheet preconditioning, enhanced convection and ring current development, *in Substorms-4*, edited by S. Kokubun, and Y. Kamide, pp. 755–760, Terra Sci., Tokyo, doi: 10.1007/978-94-011-4798-9_158

Kozyra, J. U., M. W. Liemohn, C. R. Clauer, A. J. Ridley, M. F. Thomsen, J. E. Borovsky, J. L. Roeder, V. K. Jordanova, and W. D. Gonzalez (2002) Multistep Dst development and ring current composition changes during the 4–6 June 1991 magnetic storm, *J. Geophys. Res.*, **107** (A8), 1224, doi: 10.1029/2001JA000023.

Lang, M., P. Browne, P.J. van Leeuwen and M.J. Owens (2017) Data Assimilation in the Solar Wind: Challenges and First Results, *Space Weather* **15**, 1490–1510, doi: 10.1002/2017SW001681

Lefèvre, L., S. Vennerstrøm, M. Dumbović, B. Vršnak, D. Sudar, R. Arlt, F. Clette, N. Crosby (2016) Detailed Analysis of Solar Data Related to Historical Extreme Geomagnetic Storms: 1868 – 2010, *Solar Phys*, **291**, 1483–1531, doi: 10.1007/s11207-016-0892-3

Lockwood, M. (2003) Twenty-three cycles of changing open solar flux, *J. Geophys. Res.,* **108** (A3), 1128, doi: 10.1029/2002JA009431.

Lockwood, M. (2013) Reconstruction and Prediction of Variations in the Open Solar Magnetic Flux and Interplanetary Conditions, *Living Reviews in Solar Physics*, **10**, 4, doi: 10.12942/lrsp-2013-4

Lockwood, M. and M.J. Owens (2014a) Implications of the recent low solar minimum for the solar wind during the Maunder minimum, *Astrophys. J. Lett.,* **781**, L7, doi:10.1088/2041-8205/781/1/L7

Lockwood, M. and M.J. Owens (2014b) Centennial variations in sunspot number, open solar flux and streamer belt width: 3. Modelling, *J. Geophys. Res. Space Physics*, **119** (7), 5193-5209, doi: 10.1002/2014JA019973

Lockwood, M., and M.N. Wild (1993) On the quasi-periodic nature of magnetopause flux transfer events, *J. Geophys. Res.*, **98**, 5935-5940, doi: 10.1029/92JA02375.

Lockwood, M., R. Stamper and M.N. Wild (1999) A doubling of the sun's coronal magnetic field during the last 100 years, *Nature*, **399**, 437-439, doi: 10.1038/20867

Lockwood, M., A.P. Rouillard, and I.D. Finch (2009) The rise and fall of open solar flux during the current grand solar maximum, *Ap. J.*, **700** (2), 937-944, doi: 10.1088/0004-637X/700/2/937, 2009

Lockwood, M., H. Nevanlinna, L. Barnard, M.J. Owens, R.G. Harrison, A.P. Rouillard, and C.J. Scott (2014) Reconstruction of Geomagnetic Activity and Near-Earth Interplanetary Conditions over the Past 167 Years: 4. Near-Earth Solar Wind Speed, IMF, and Open Solar Flux, *Annales. Geophys.,* **32**, 383-399, doi: 10.5194/angeo-32-383-2014

Lockwood, M., M.J. Owens, L.A. Barnard S. Bentley, C.J. Scott, and C.E. Watt (2016) On the Origins and Timescales of Geoeffective IMF, *Space Weather*, **14**, 406–432, doi: 10.1002/2016SW001375

Lockwood, M., M.J. Owens, L.A. Barnard, C.J. Scott, and C.E. Watt (2017a) Space Climate and Space Weather over the past 400 years: 1. The Power input to the Magnetosphere, *J. Space Weather Space Clim.*, **7**, A25, doi: 10.1051/swsc/2017019

Lockwood, M., M.J. Owens, L.A. Barnard, C.J. Scott, C.E. Watt and S. Bentley (2018a) Space Climate and Space Weather over the past 400 years: 2. Proxy indicators of geomagnetic storm and substorm occurrence, *J. Space Weather Space Clim.*, in press, doi: 10.1051/swsc/2017048

Lockwood, M., S. Bentley, M.J. Owens, L.A. Barnard, C.J. Scott, C.E. Watt and O. Allanson (2018b) The development of a space climatology: 1. Solar-wind magnetosphere coupling as a function of timescale and the effect of data gaps, *Space Weather*, in press

Lockwood, M., S. Bentley, M.J. Owens, L.A. Barnard, C.J. Scott, C.E. Watt and O. Allanson (2018c) The development of a space climatology: 2. the variation of space weather parameters with timescale, *Space Weather*, in press

Lockwood, M., A. Chambodut, L.A. Barnard, M.J. Owens, E. Clarke, and V. Mendel (2018d) A homogeneous aa index: 1. Secular variation, J. Space Weather Space Clim., in press, 10.1051/swsc/2018038

Lockwood, M., I.D. Finch, A. Chambodut, L.A. Barnard, M.J. Owens, and E. Clarke (2018e) A homogeneous aa index: 2. hemispheric asymmetries and the equinoctial variation, J. Space Weather Space Clim., in press. 10.1051/swsc/2018044

Lotz, S.I. and D.W. Danskin (2017) Extreme value analysis of induced geoelectric field in South Africa, *Space Weather,* **15**, doi:10.1002/2017SW001662.

Love, J. J. (2012), Credible occurrence probabilities for extreme geophysical events: Earthquakes, volcanic eruptions, magnetic storms, *Geophys. Res. Lett.*, 39, L10301, doi:10.1029/2012GL051431.

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. G*eophysical Research Letters*, **42**, 6544–6553, doi: 10.1002/2015gl064842

Matthes, K., B. Funke, M.E.Andersson, L. Barnard, J. Beer, P. Charbonneau, M.A. Clilverd, T. Dudok de Wit, M. Haberreiter, A. Hendry, C.H. Jackman, M. Kretzschmar, T. Kruschke, M.

Kunze, U. Langematz, D.R. Marsh, A. Maycock, S. Misios, C.J. Rodger, A.A. Scaife, A. Seppälä, M. Shangguan, M. Sinnhuber, K. Tourpali, I. Usoskin, M. van de Kamp, P.T. Verronen, and S.Versick (2017) Solar Forcing for CMIP6 *Geosci. Model Dev*., **10**, 2247-2302, doi: 10.5194/gmd-10-2247-2017

Mayaud, P.-N. (1972) The *aa* indices: A 100-year series characterizing the magnetic activity, *J. Geophys. Res*., **77**, 6870–6874, doi: 10.1029/JA077i034p06870

Mayaud, P.-N. (1980) Derivation, Meaning and Use of Geomagnetic Indices, Geophysical Monograph, **22**, American Geophysical Union, Washington, DC, doi: 10.1029/GM022

Menvielle, M., and A. Berthelier (1991) The K-derived planetary indices: Description and availability, *Rev. Geophys*., **29** (3), 415–432, doi: 10.1029/91RG00994.

Mourenas D., A.V Artemyev, and X.-J. Zhang (2018), Statistics of extreme time integrated geomagnetic activity, *Geophys. Res. Lett*., **44**, doi:10.1002/2017GL076828.

Nevanlinna, H. (2006) A study on the great geomagnetic storm of 1859: Comparisons with other storms in the 19th century, *Adv. Space Res*., **38** (2), 180–187, doi: 10.1016/j.asr.2005.07.076.

Ngwira, C.M., A. Pulkkinen, M. Leila Mays, M.M. Kuznetsova, A.B. Galvin, K. Simunac, D.N. Baker, X. Li, Y. Zheng, and A. Glocer (2013), Simulation of the 23 July 2012 extreme space weather event: What if this extremely rare CME was Earth directed?, *Space Weather*, **11**, 671–679, doi:10.1002/2013SW000990.

Ngwira, C.M., A. Pulkkinen, M.M. Kuznetsova, and A. Glocer (2014), Modeling extreme "Carrington-type" space weather events using three-dimensional global MHD simulations, *J. Geophys. Res. Space Physics*, **119**, 4456–4474, doi: 10.1002/2013JA019661.

O'Brien, T. P., and R. L. McPherron (2000) An empirical phase space analysis of ring current dynamics: Solar wind control of injection and decay, *J. Geophys. Res*., **105** (A4), 7707–7719, doi: 10.1029/1998JA000437.

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

Owens, M. J., and P. Riley (2017) Probabilistic solar wind forecasting using large ensembles of near-Sun conditions with a simple one-dimensional "upwind" scheme. *Space Weather*, **15**, doi: 10.1002/2017SW001679

Owens, M.J. M. Lockwood and I. Usoskin (2012) Heliospheric modulation of galactic cosmic rays during grand solar minima: Past and future variations, *Geophys. Res. Lett*., **39**, L19102, 2012. doi:10.1029/2012GL053151

Owens, M.J., E. Cliver, K. McCracken, J. Beer, L.A. Barnard, M. Lockwood, A.P. Rouillard, D. Passos, P. Riley, I.G. Usoskin, Y.-M. Wang (2016a) Near-Earth Heliospheric Magnetic Field

Intensity since 1800. Part 1: Sunspot and Geomagnetic Reconstructions, *J. Geophys. Res.*, **121** (7), 6048-6063, doi: 10.1002/2016JA022529

Owens, M.J., E. Cliver, K. McCracken, J. Beer, L.A. Barnard, M. Lockwood, A.P. Rouillard, D. Passos, P. Riley, I.G. Usoskin, Y.-M. Wang (2016b) Near-Earth Heliospheric Magnetic Field Intensity since 1800. Part 2: Cosmogenic Radionuclide Reconstructions, *J. Geophys. Res.*, **121** (7), 6064-6074, doi: 10.1002/2016JA022550

Owens, M.J., M. Lockwood, P. Riley (2017) Global solar wind variations over the last four centuries, *Scientific Reports.*, **7**, Article number 41548, doi:10.1038/srep41548

Pirjola, R. (2005) Effects of space weather on high-latitude ground systems, *Adv. Space Res.*, **36**, Issue 12, 2005, Pages 2231-2240, ISSN 0273-1177, doi: 10.1016/j.asr.2003.04.074.

Pirjola, R., K. Kauristie, H. Lappalainen, A. Viljanen, and A. Pulkkinen (2005), Space weather risk, *Space Weather*, **3**, S02A02, doi:10.1029/2004SW000112

Pizzo, V.J., C. de Koning, M. Cash, G. Millward, D.A. Biesecker, L. Puga, M. Codrescu, and D. Odstrcil (2015) Theoretical basis for operational ensemble forecasting of coronal mass ejections, *Space Weather*, **13**, 676–697, doi:10.1002/2015SW001221.

Pulkkinen, A., R. Pirjola, D. Boteler, A. Viljanen, I. Yegorov (2001) Modelling of space weather effects on pipelines, *J. Applied Geophysics*, **48** (4), 233-256, doi: 10.1016/S0926-9851(01)00109-4.

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

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

Schunk, R. W., et al. (2014) Ensemble Modeling with Data Assimilation Models: A New Strategy for Space Weather Specifications, Forecasts, and Science, *Space Weather*, **12**, 123–126, doi:10.1002/2014SW001050.

Shea, M.A. and D.F. Smart (2006) Compendium of the eight articles on the "Carrington Event" attributed to or written by Elias Loomis in the American Journal of Science, 1859-1861, *Adv. Space Res.*, **38** (2), 313-385, doi: 10.1016/j.asr.2006.07.005.

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

Silbergleit, V. M. (1999), Forecast of the most geomagnetically disturbed days, *Earth Planets Space*, **51**, 19–22, doi: 10.1186/bf03352205

Siscoe, G. L. (1976), On the statistics of the largest geomagnetic storms per solar cycle, *J. Geophys. Res*., **81**(25), 4782–4784, doi: 10.1029/JA081i025p04782

Siscoe, G., and S.C. Solomon (2006) Aspects of data assimilation peculiar to space weather forecasting, *Space Weather*, **4**, S04002, doi:10.1029/2005SW000205.

Stern, D. (2002) A millennium of geomagnetism, *Rev. Geophys*., **40** (3), 1007, doi: 10.1029/2000RG000097

Svalgaard, L., and E. W. Cliver (2010) Heliospheric magnetic field 1835–2009, *J. Geophys. Res.*, **115**, A09111, 2010, doi: 10.1029/2009JA015069

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

Tindale, E., Chapman, S. C., Moloney, N. R., & Watkins, N. (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*, **16** (8), 1128-1142, doi:10.1029/2018JA025740

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

Tsurutani, B.T., Y. Kamide, J.K. Arballo, W.D. Gonzalez, R.P. Lepping (1999) Interplanetary causes of great and superintense magnetic storms, *Physics and Chemistry of the Earth, Part C*: *Solar, Terrestrial & Planetary Science*, **24** (1-3), 101-105, doi: 10.1016/S1464-1917(98)00015-4.

Turner, N. E., D. N. Baker, T. I. Pulkkinen, and R. L. McPherron (2000) Evaluation of the tail current contribution to Dst, *J. Geophys. Res*., **105** (A3), 5431–5439, doi:10.1029/1999JA000248.

Usoskin, I.G., R. Arlt, E. Asvestari, E. Hawkins, M. Käpylä, G.A. Kovaltsov, N. Krivova, M. Lockwood, K. Mursula, J. O'Reilly, M. Owens, C. J. Scott, D. D. Sokoloff, S. K. Solanki, W. Soon, and J. M. Vaquero (2015) The Maunder minimum (1645-1715) was indeed a Grand minimum: A reassessment of multiple datasets, *Astron. and Astophys*., **581**, A95, doi: 10.1051/0004-6361/201526652

Vaselovsky, I.S., A.V. Dmitriev, and A.V. Suvorova (2010) Lognormal, Normal and Other Distributions Produced by Algebraic Operations in *The Solar Wind*, AIP Conference Proceedings 1216, 152 -155, doi: 10.1063/1.3395824

Vasyliunas, V. M., J. R. Kan, G. L. Siscoe, and S.-I. Akasofu (1982) Scaling relations governing magnetospheric energy transfer, *Planet. Space Sci*., **30**, 359–365, doi: 10.1016/0032-0633(82)90041-1

Viljanen, A., A. Pulkkinen, R. Pirjola, K. Pajunpää, P. Posio, and A. Koistinen (2006), Recordings of geomagnetically induced currents and a nowcasting service of the Finnish natural gas pipeline system, *Space Weather*, **4**, S10004, doi:10.1029/2006SW000234

Vörös Z., M. Leitner, Y. Narita, G. Consolini, P. Kovács, A. Tóth, and J. Lichtenberger (2015) Probability density functions for the variable solar wind near the solar cycle minimum, *J. Geophys. Res. Space Physics*, **120**, 6152–6166, doi:10.1002/2015JA021257

Weigel, R.S. and D.N. Baker (2003) Probability distribution invariance of 1-minute auroral-zone geomagnetic field fluctuations, *Geophys. Res. Lett.*, **30** (23), 2193, doi:10.1029/2003GL018470

Wilkes, D.S. (1995) Statistical methods in the atmospheric sciences, second edition, *International Geophysics Series*, Vol 59, Academic Press, 464pp. ISBN-10: 0127519653. ISBN-13: 978-0127519654.

Woody, M.S., J.H. Lewis, M.J. Greenberg, Y.E. Goldman, E.M. Ostap (2016) MEMLET: An Easy-to-Use Tool for Data Fitting and Model Comparison Using Maximum-Likelihood Estimation, *Biophysical J.,* **111** (2), 273 to 282, doi: 10.1016/j.bpj.2016.06.019

Xie, H., N. Gopalswamy, P. K. Manoharan, A. Lara, S. Yashiro, and S. T. Lepri (2006) Long-lived geomagnetic storms and coronal mass ejections, *J. Geophys. Res*., **111**, A01103, doi: 10.1029/2005JA011287.

Xiang, N.B. and Z. N. Qu (2018) Evolutionary Characteristics of the Interplanetary Magnetic Field Intensity, *Ap. J.*, **156** (4), 152. doi: 10.3847/1538-3881/aadb91

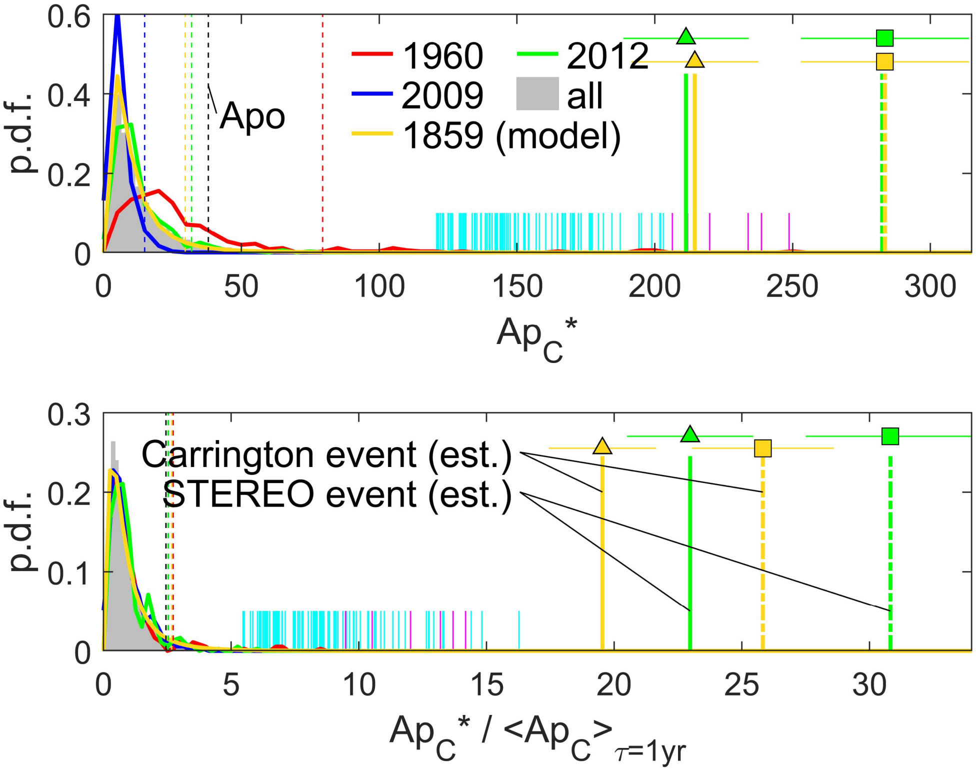

Figure 1. Distributions of ap_C, the *ap* index corrected for the annual variation in its response function, (see Appendix B). Annual distributions of (top) 8-point running (boxcar) means of the three-hourly ap_C values, Ap_C*, and (bottom) of those means as a ratio of the annual mean value for the calendar year in queetion, $Ap_C*/<ap_C>_{\tau=1yr}$, for:- (red) 1960; (blue) 2009; (green) 2012; and (orange) modelled for 1859. The gray histograms in the background are the distributions for all 248368 Ap_C* values available from the interval 1932-2016. The vertical orange lines mark the estimated value for the peak of the 1859 Carrington event: the solid orange line is “estimate 1”, $[Ap_C*]_{MAX}$ which makes allowance for the time-of-year response of the *ap* index (also marked by an orange triangle), the dot-dash orange line is $[Ap*]_{MAX}$ which does not make this correction (“estimate 2”, also marked by an orange square). The uncertainty bars arise only from the conversion of *Aa** to *Ap** and do not include the uncertainty in the *Aa** estimate. The distributions for 2012 are shown because in that year an event, that it is estimated would have caused an extreme event almost as large as the Carrington event, passed over the STEREO A craft but missed the Earth: the vertical green lines show the estimated maximum for that event, had it hit Earth: the solid

green line and green triangle is for the $[Ap_C*]_{MAX}$ (estimate 1) value and the dot-dash green line and green square are for the $[Ap*]_{MAX}$ (estimate 2) value. The vertical coloured dashed lines give the 95-percentiles of the annual distributions, using the same color scheme and the vertical black dashed lines are the equivalent for *Apo*, the 95th percentile of all Ap_C* values. The short vertical cyan lines show the top 100 (0.32%) of the maximum Ap_C* values in a calendar day, $[Ap_C*]_{MAX}$, and the short vertical mauve lines $[Ap_C*]_{MAX}$ values are the 6 days in the top 0.02%. The top 100 events, with further details, are listed in Part 3 of the Supporting Information file.

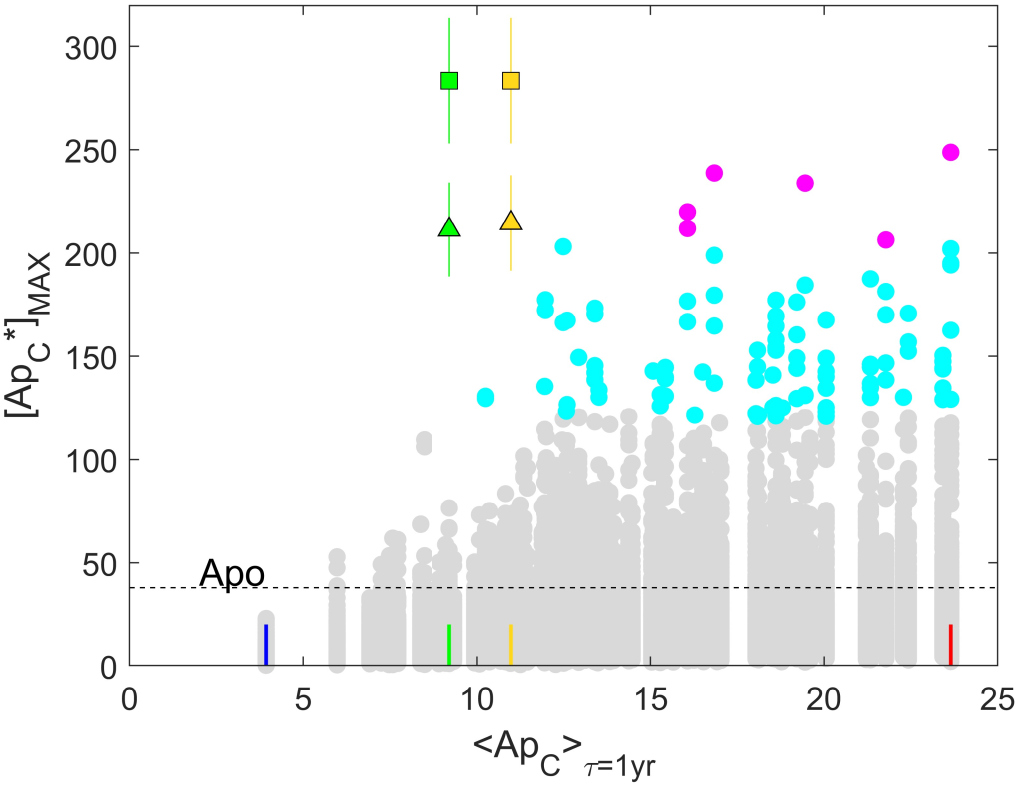

Figure 2. The largest $Ap_C{}^*$ values in a calendar day, $[Ap_C{}^*]_{MAX}$, as a function of the annual mean for the calendar year of that day $<ap_C>_{\tau=1yr}$ for 1932-2016 (inclusive). The grey points make up 99.68% of the available 31047 daily $[Ap_C{}^*]_{MAX}$ samples, the cyan points being in the top 100 days in terms of their $[Ap_C{}^*]_{MAX}$ value (also shown by the short vertical cyan lines in Figure 1) and the mauve points the 6 days in the top 0.02% (shown by the short vertical mauve lines in Figure 1). The top 100 days are listed in the Supporting Information file. The orange and green triangles show the estimated $[Ap_C{}^*]_{MAX}$ values for the Carrington and STEREO-A events (in 1859 and 2012, respectively, see text for details) and the orange and green squares show the corresponding uncorrected $[Ap^*]_{MAX}$ values. The uncertainty bars arise only from the conversion of Aa^* to Ap^* and do not include any uncertainty in the Aa^* estimate. The horizontal dashed line is *Apo*, the 95th percentile of all $Ap_C{}^*$ values. The colored tickmarks along the x axis mark the annual means of the four annual distributions shown in Figure 1 (from left to right 2009, 2012, 1859 and 1960), using the same color scheme

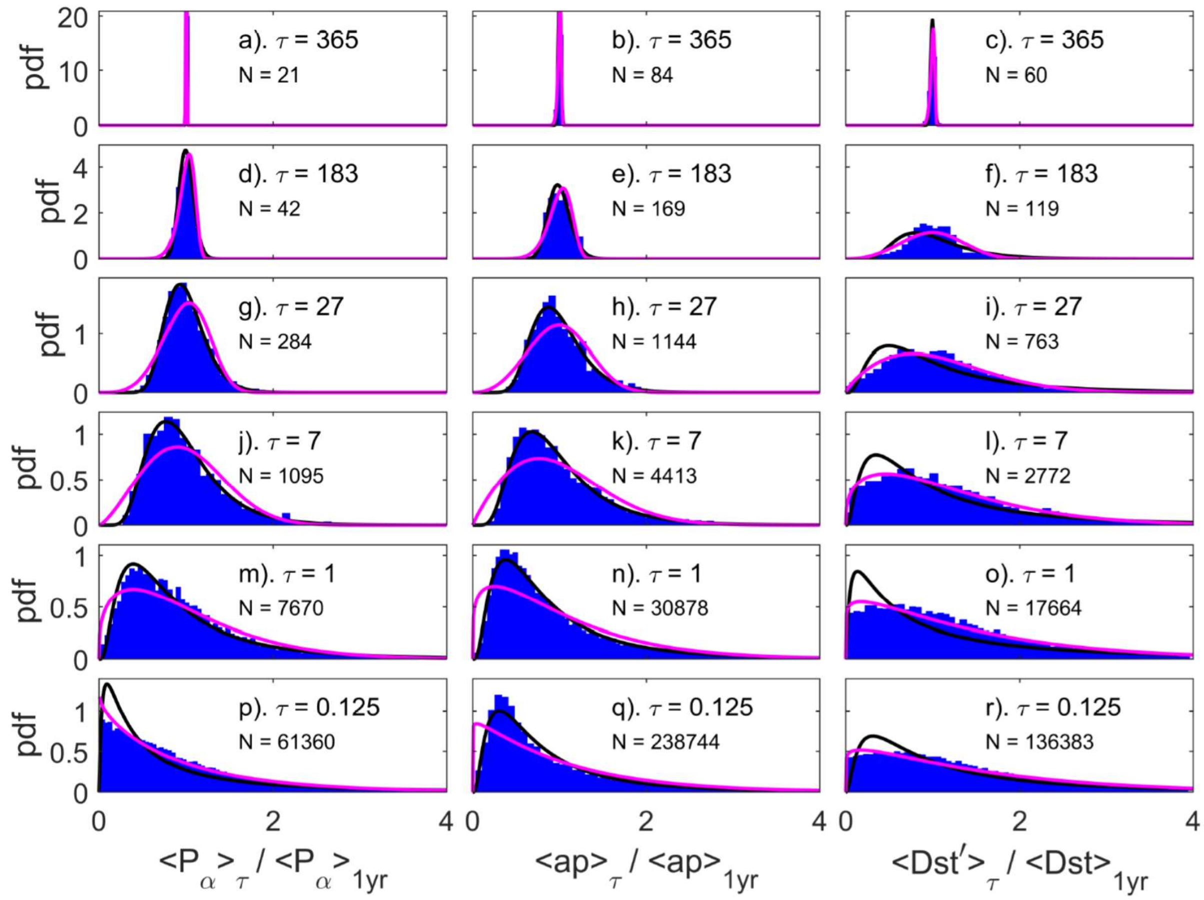

Figure 3. Distributions of (left hand panels) normalized power input into the magnetosphere, $<P_\alpha>_\tau/<P_\alpha>_{1yr}$; (central panels) normalized geomagnetic *ap* index, $<ap>_\tau/<ap>_{1yr}$; and (right hand panels) normalized negative geomagnetic *Dst* index, $< Dst' >_\tau/<Dst >_{1yr}$. The coupling function of $\alpha = 0.44$, shown in Paper 1 to apply at all τ, is used to generate P_α. The distributions are of the means taken over intervals τ long, divided by the annual mean of all samples in that year. The blue histograms are the observed distributions, with samples binned into 150 contiguous bins centered on $k.x_{98}/100$ where k is varied between 0.5 and 149.5 in steps of 1 and x_{98} is the 98th percentile of the c.d.f. and the numbers of samples n are then normalized such that $(x_{98}/100)\Sigma n$ is unity. The black lines shows the best-fit lognormal distributions and the mauve lines are the best-fit Weibull distributions (with mean value $m = 1$ in the cases of P_α and *ap* and $m = Rm(\tau)$ for Dst'). Fits are made using Maximum Likeliood Estimation (see the Supporting Information file). The total number of available samples, N, is given in each panel. (a), (b) and (c) are for $\tau = 1$yr; (d), (e) and (f) for $\tau = 0.5$yr; (g), (h) and (i) for $\tau = 27$dy; (j), (k) and (l) for $\tau = 7$ day; (m), (n) and (o) for $\tau = 1$

day; and (p), (q) and (r) for τ = 3hr. The P_α data are from 1996-2016 (inclusive), the *ap* data for 1932-2016 (inclusive) and the Dst' data are for 1957-2016 (inclusive). $<Dst>_\tau \geq 0$ samples are omitted giving Dst' (so because all $<Dst>_{1yr}$ values are negative, these give $<Dst'>_\tau/<Dst>_{1yr} \geq 0$) in histograms and distribution fits: as a result N for Dst' is 100%, 99.17%, 94.08%, 88.42%, 80.60%, and 78.48% of all *Dst* samples for τ of, respectively, 1yr. (panel c), 0.5yr. (panel f), 27days (panel i), 7 days (panel l), 1 day (panel o), and 3hr. (panel r). The best-fit distribution parameters, goodness-of-fit metrics and c.d.f and p.d.f plots are given in the Supporting Information file for these two fitted distributions and 5 others.

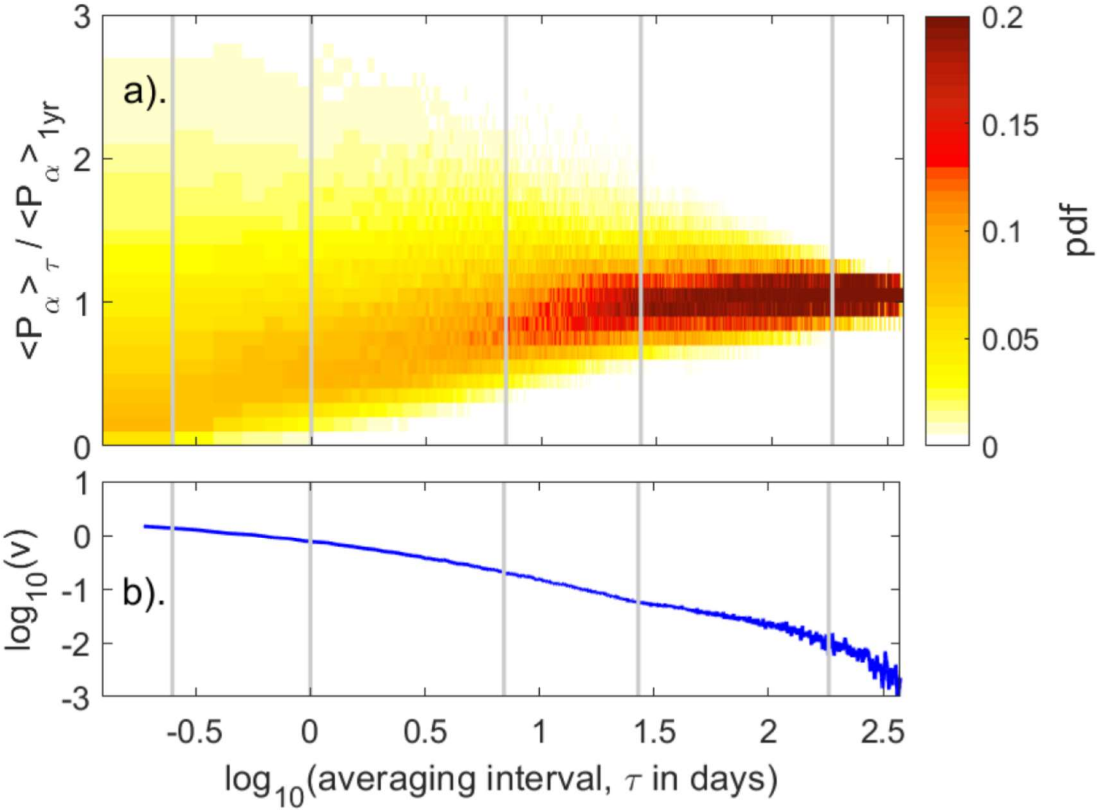

Figure 4. The variation of the observed distributions of the normalized power input into the magnetosphere $<P_\alpha>_\tau/<P_\alpha>_{1yr}$ for $\alpha = 0.44$) as a function of the logarithm of the averaging interval, $\log_{10}(\tau)$. The left hand edge of the plot is at $\tau = 3$ hrs, the right hand edge at $\tau = 1$ yr., and the vertical black lines show τ of 6 hours, 1 day, 7 days, 27 days and 0.5 year. (b) The logarithm of the best-fit variance of the lognormal distribution (of mean value $m = 1$), $\log_{10}(v)$, also as a function of $\log_{10}(\tau)$.

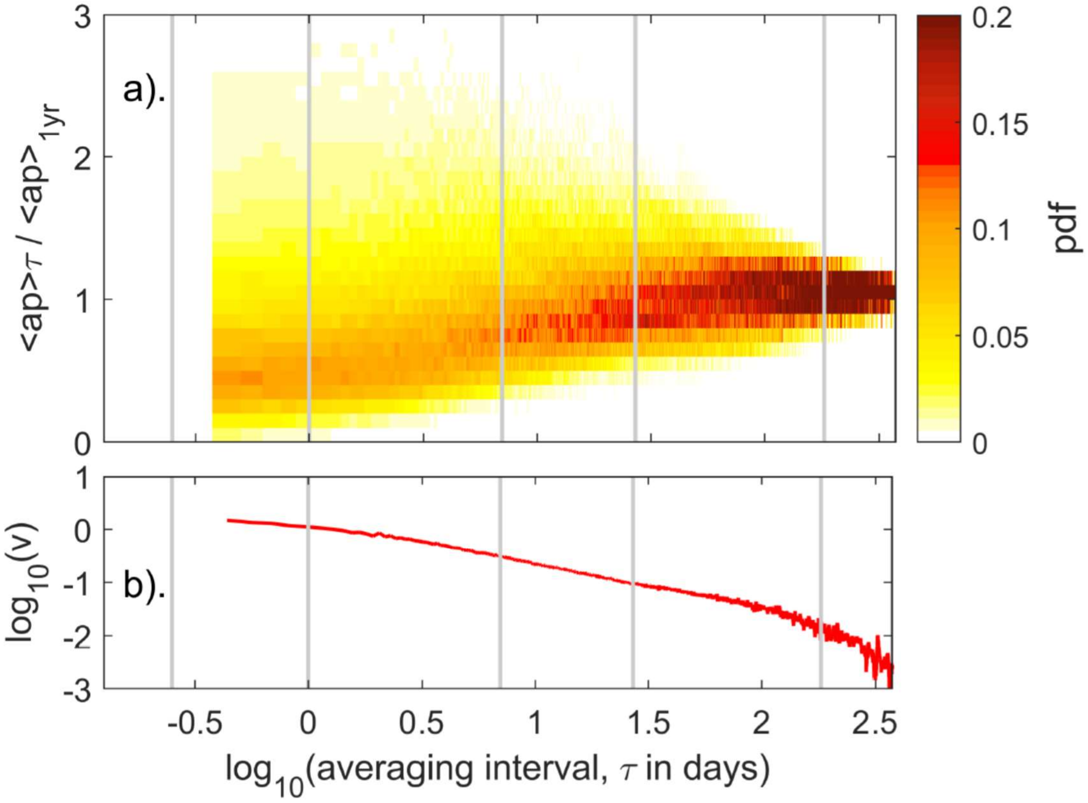

Figure 5. Same as Figure 4 for the normalized *ap* geomagnetic index, <*ap*>$_{\tau}$ / <*ap*>$_{1yr.}$ The distributions for τ < 9 hr. are not shown as the quantization of 3-hourly *ap* levels becomes a factor.

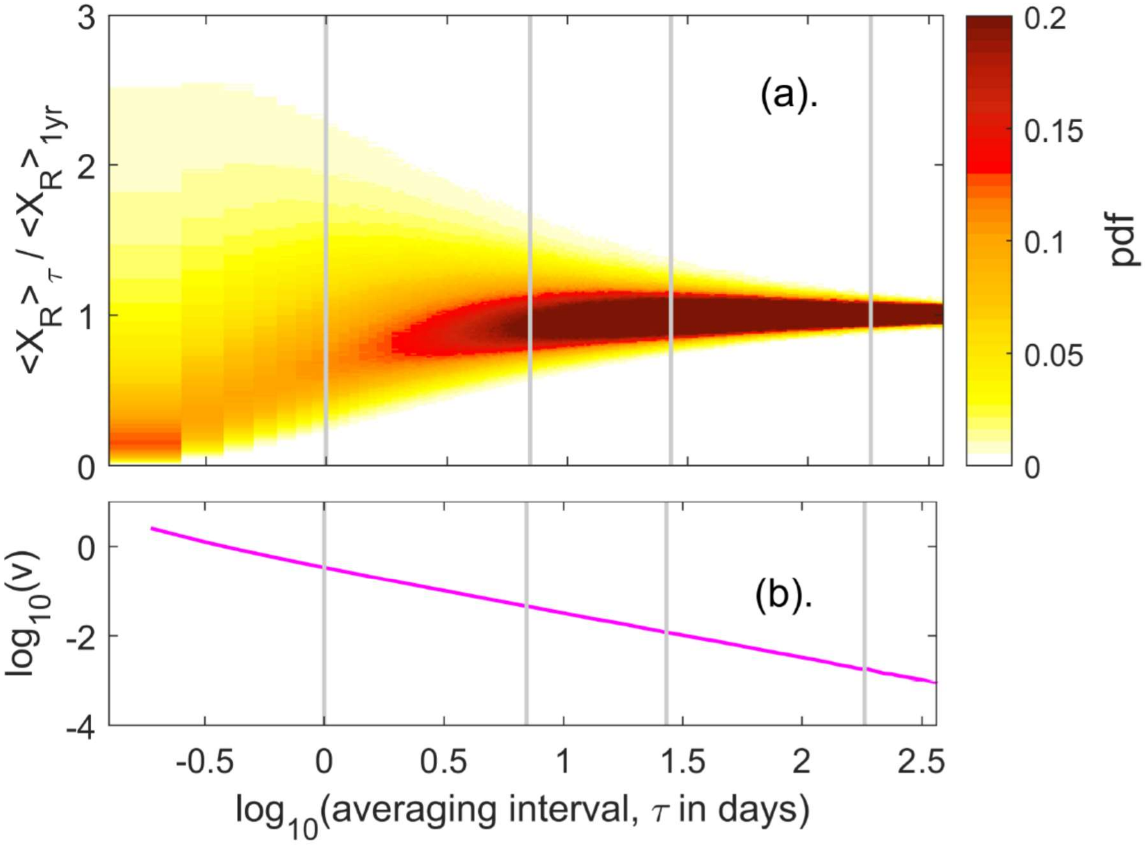

Figure 6. The same as Figure 4 for a random variable X_R of the same length and time resolution as the P_α data series and which for τ = 3 hr. is drawn from a Weibull distribution with k of 1.0625 and λ of 1.0240, which in Paper 2 [*Lockwood et al.,* 2018c] was shown to be good fit to the distribution of P_α at that timescale.

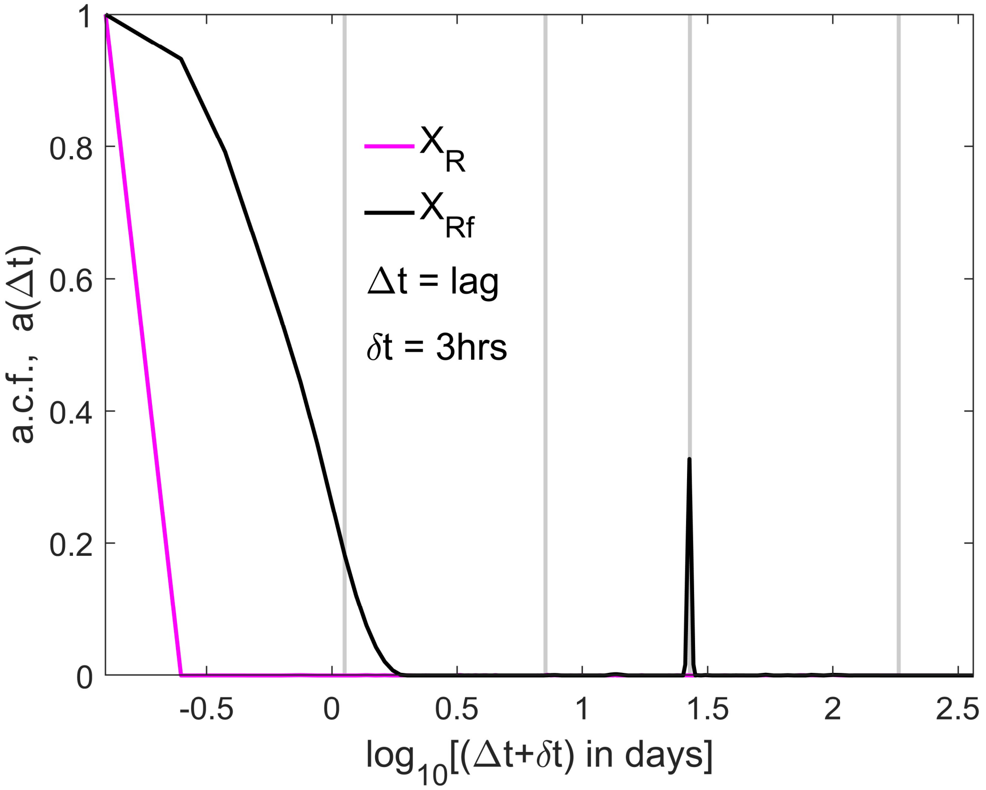

Figure 7. The autocorrelation functions (ACFs) of (in mauve) the random variable X_R employed in Figure 6 and (in black) the filtered random variable X_{Rf} employed in Figure 8. The ACF, $a(\Delta t)$ is computed for lags Δt between zero and 1 year in steps of the data resolution (δt=3 hrs) and are shown as a function of $\log_{10}(\Delta t+\delta t)$ where Δt and δt are both in units of days. (The δt is added to Δt to allow the zero lag point to be shown on a logarithmic scale). The left hand edge of the plot is at Δt =0 and the right hand edge at Δt = 1 year and the vertical grey lines are at lags Δt of 1 day, 7 days, 27 days and 0.5 year. Lag 1 (Δt=δt) is at −0.602 on the x axis.

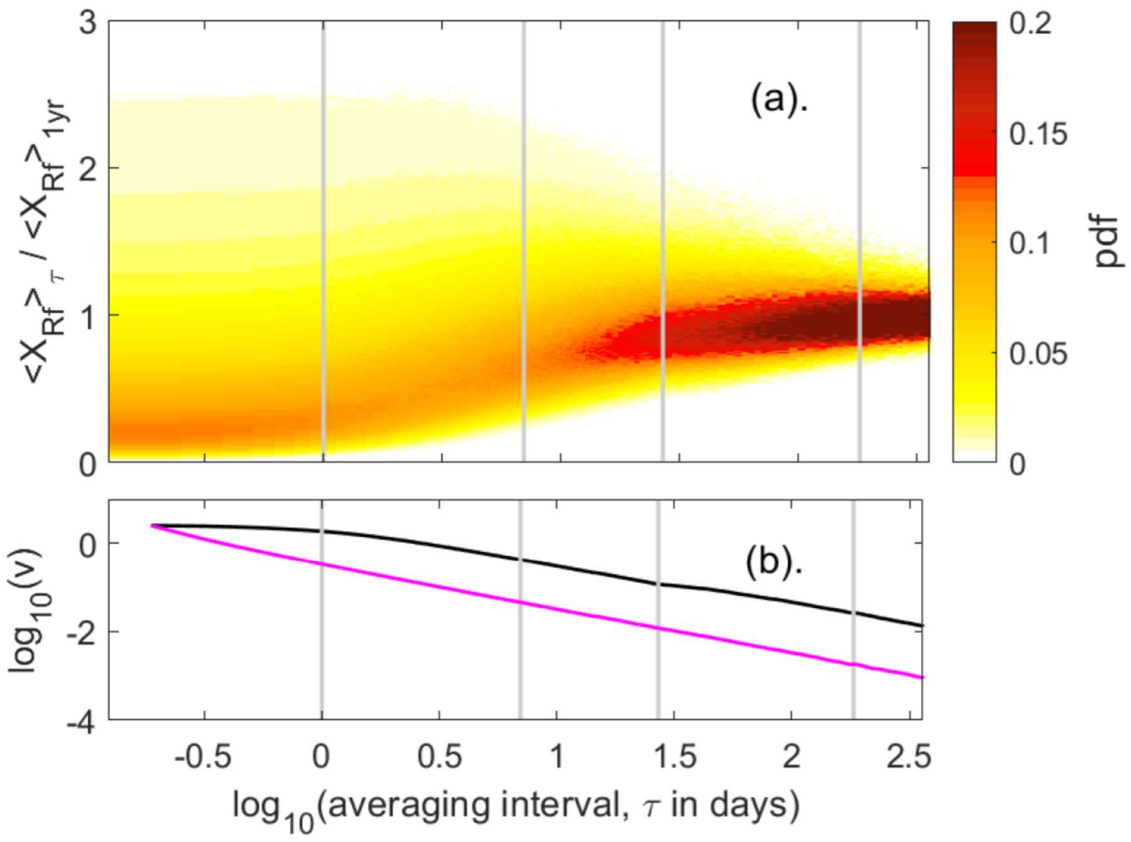

Figure 8. The same as Figure 6 for a random variable X_{Rf} which has been drawn from a Weibull distribution and then passed through a filter to generate the short-term persistence and the 27-day recurrence shown by the autocorrelation function in blue in Figure 7 (see text for details of the filter). In order that the distributions of X_{Rf} and X_{R} have the same variance at $\tau = 3$ hrs (with unity mean), the effect of the filter means that before filtering the distribution must be drawn from a higher-variance Weibull distribution (with unity mean) than X_{R} with k of 0.2800 and λ of 0.0778. The black line in (b) shows the evolution of the variance, v, (on a logarithmic scale) with τ for X_{Rf} and the blue line is the same variation for X_{R}, as shown in Figure 6(b).

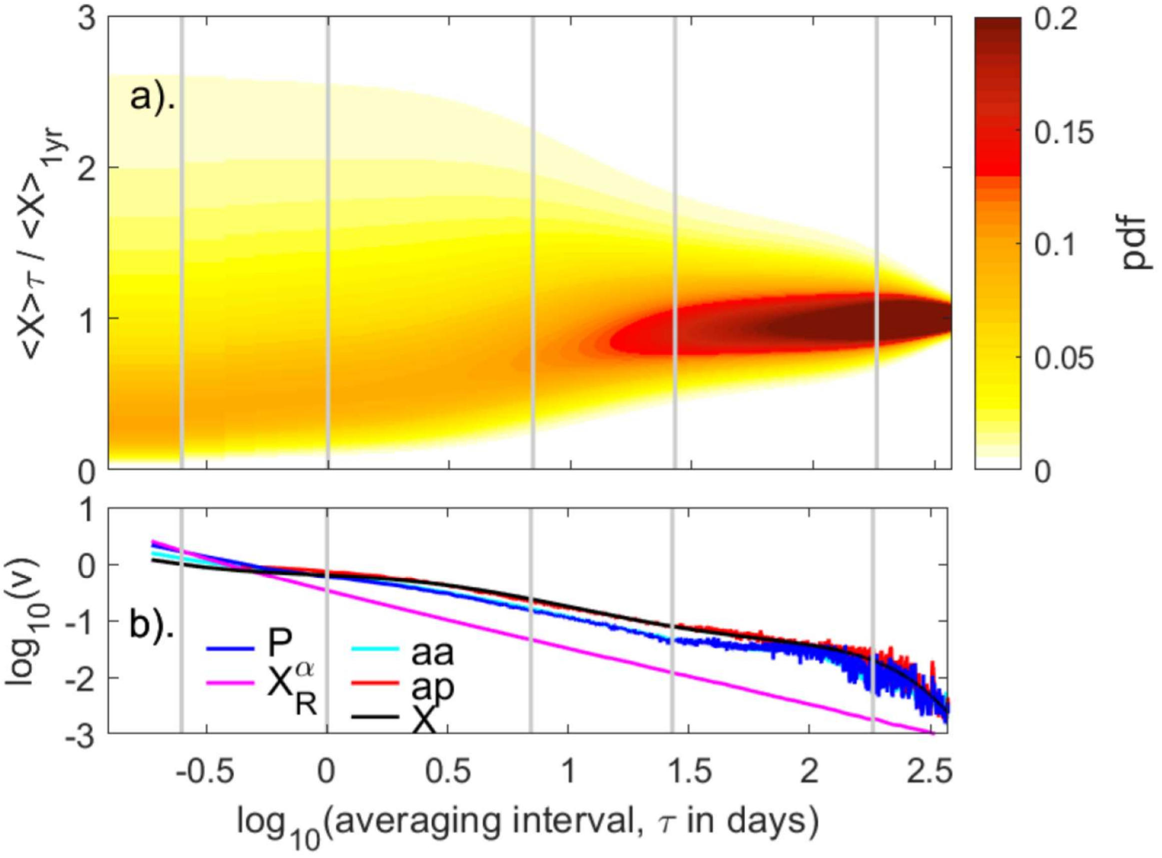

Figure 9. Same as Figure 5 for a model *X* based on lognormal distributions and a 6th-order polynomial fit to the variance of *ap*, $v(\tau)$. In (b) the red line shows v (τ) for *ap* (on a logarithmic scale) and the black line is the polynomial fit (see Appendix A for the polynomial coefficients and formulae for the lognormal distribution family). Also shown are the $v(\tau)$ variations for other variables using the same color scheme as used in Figures 4b, 5b and 6b: P_α (in blue); random variable, X_R (in mauve), plus the *aa* geomagnetic index (in cyan).

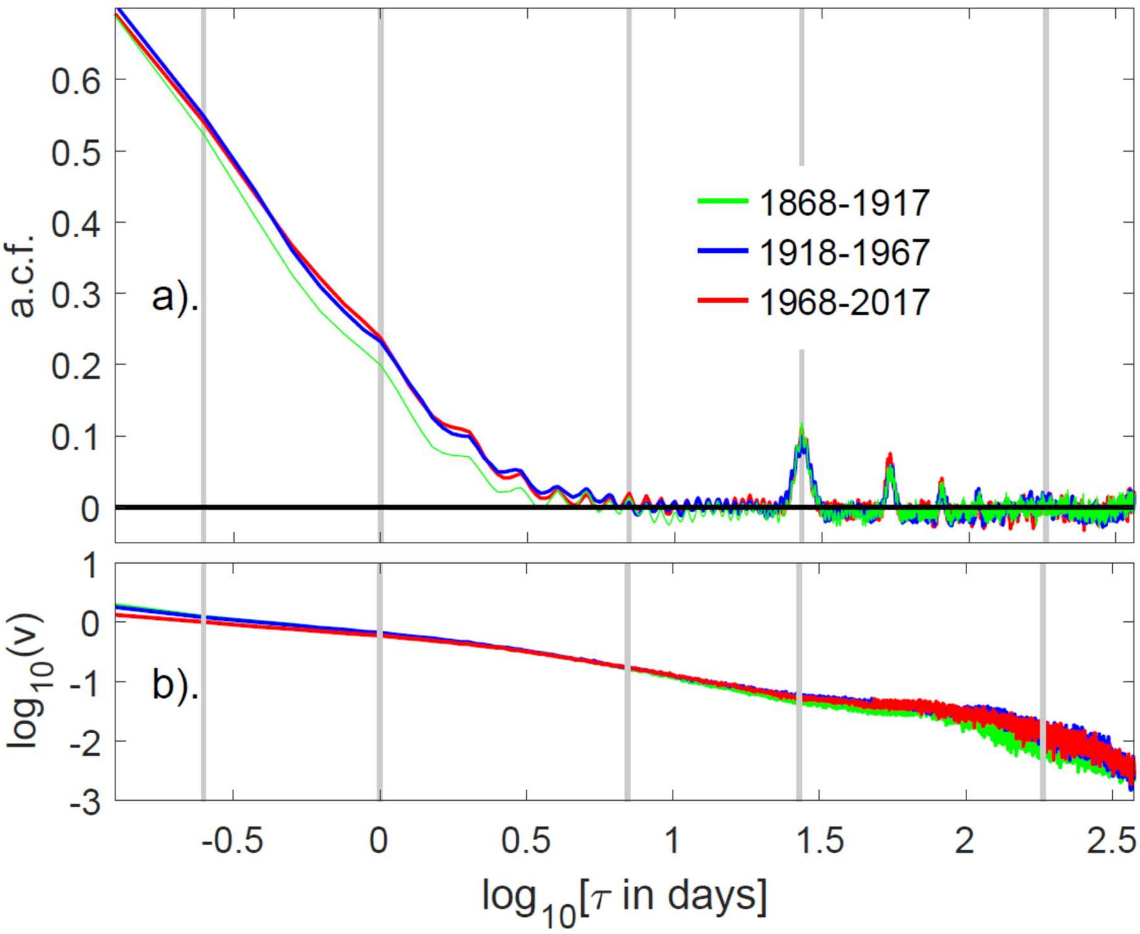

Figure 10. (Top) The autocorrelation function of the 3-hourly *aa* index, divided into three 50-year intervals: (red) 1968-2017 (inclusive); (blue) 1918-1967; and (green) 1868-1917. The lower panel shows the relationship of the variance *v* of the lognormal distribution of $<aa>_{\tau}/<aa>_{\tau=1yr}$ as a function of the averaging timescale (on the log-log plot format used in part (b) of Figures 4-9).

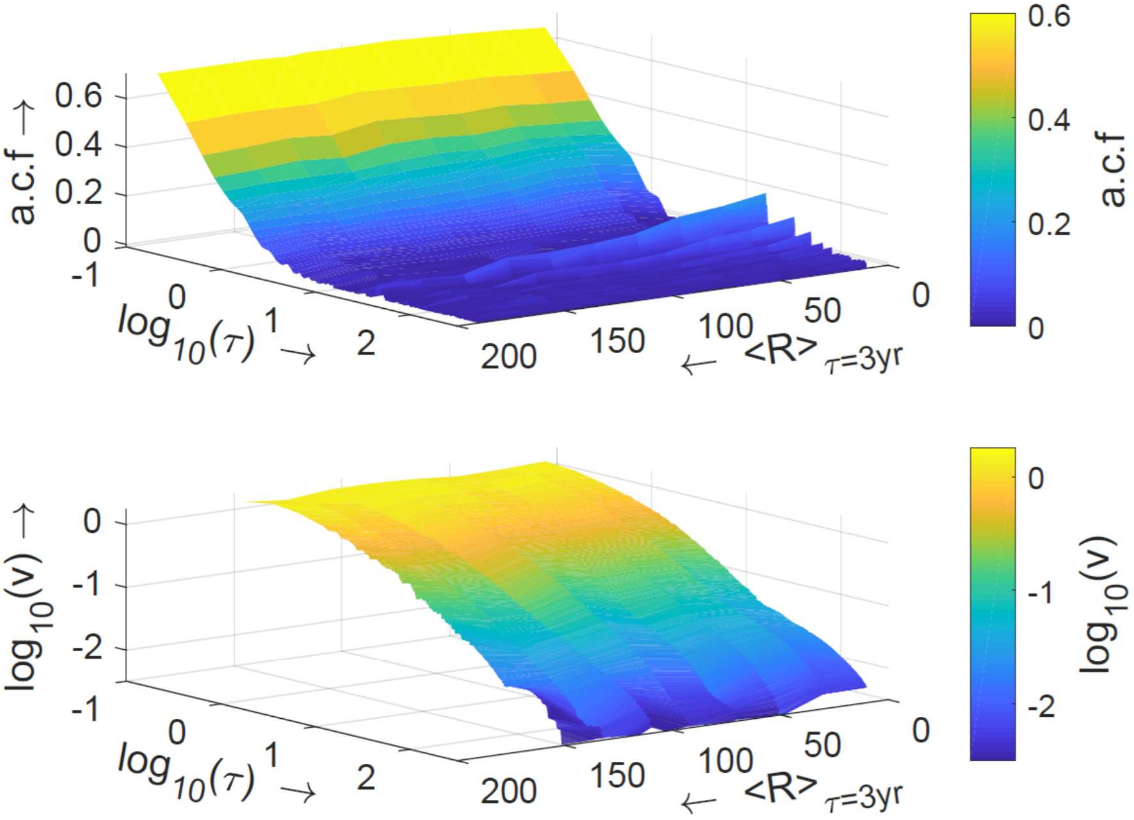

Figure 11. Surface plots of (top) The autocorrelation function, ACF, and (bottom) the logarithm of the variance, $\log_{10}(v)$, for all the *aa* index data (1868-2017) as a function of the logarithm of the averaging timescale, $\log_{10}(\tau)$., and the mean international sunspot number, averaged over a 3-year interval, $\langle R \rangle_{\tau=3\mathrm{yrs}}$.

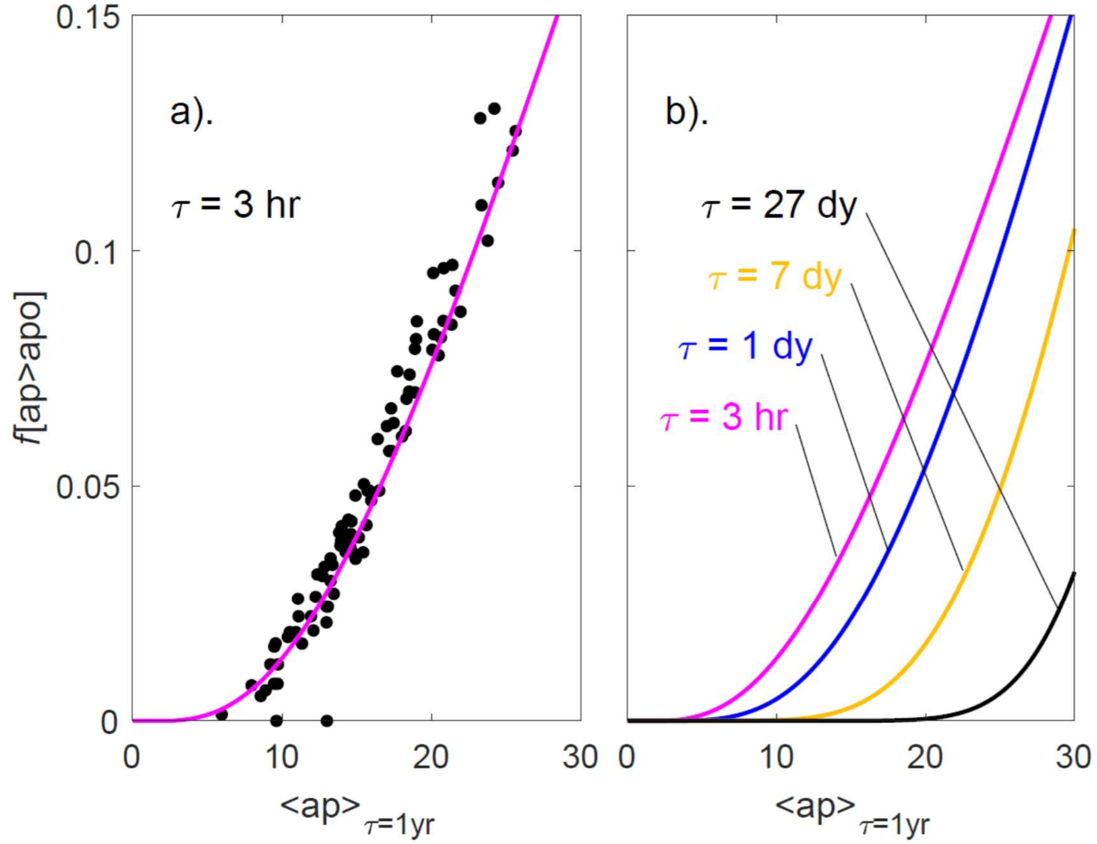

Figure 12. Predictions by the model fit to the *ap* distributions with τ shown in Figure 9. (a) The points show probability that 3-hour values of *ap* are in the top 5% of the overall distribution (for 1932-2016, 252152 samples), *f* [*ap* > *apo*] (i.e., *ap* exceeds its 95-percentile of 3-hourly *ap* values, *ap*o = 47.91), as a function of the annual mean value <*ap*>$_{\tau=1\mathrm{yr}}$. The mauve line is the model prediction for τ=3hrs. (b). The family of model predictions of *f* [*ap* > *ap*o] as a function of <*ap*>$_{\tau=1\mathrm{yr}}$ for timescales τ of 3 hours (in mauve), 1day (in blue), 7 days (in orange) and 27 days (in black).

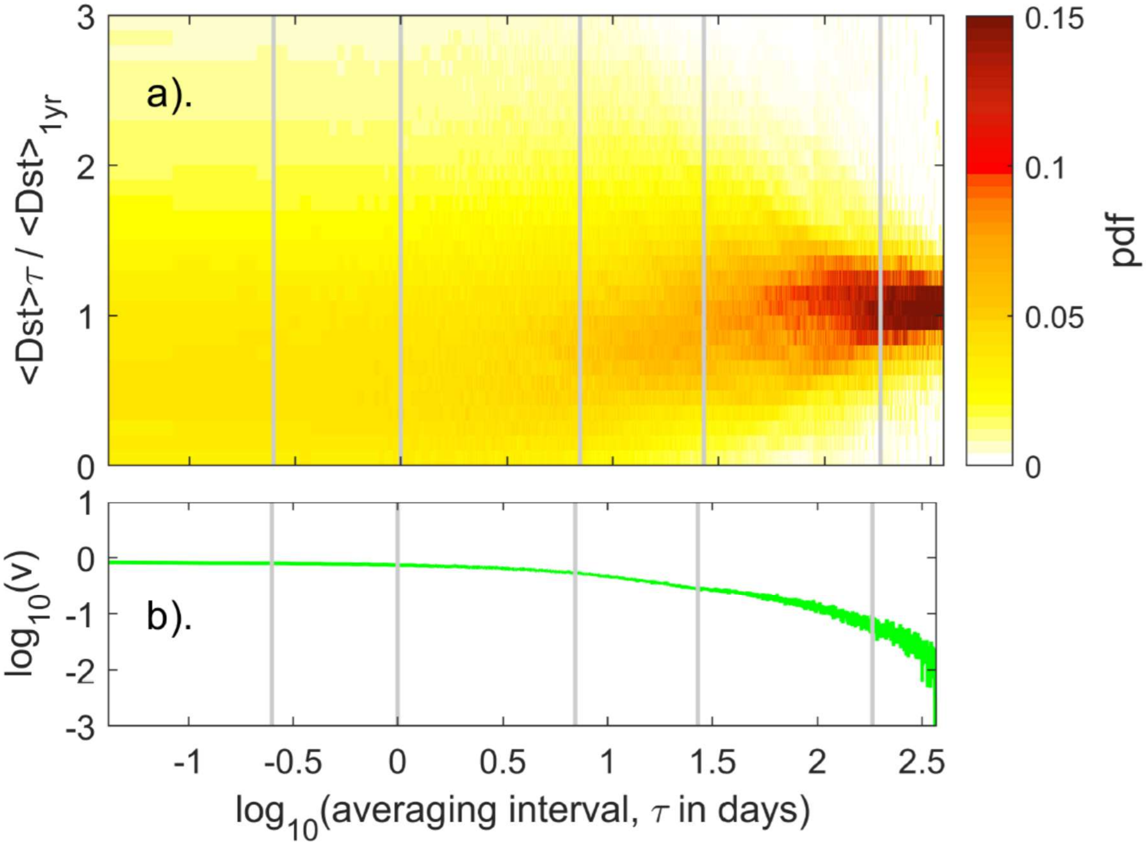

Figure 13. Same as Figure 4 for the normalized Dst geomagnetic index, $< Dst' >_{\tau} / <Dst>_{1yr}$ where Dst' is the subset of Dst values that are negative.

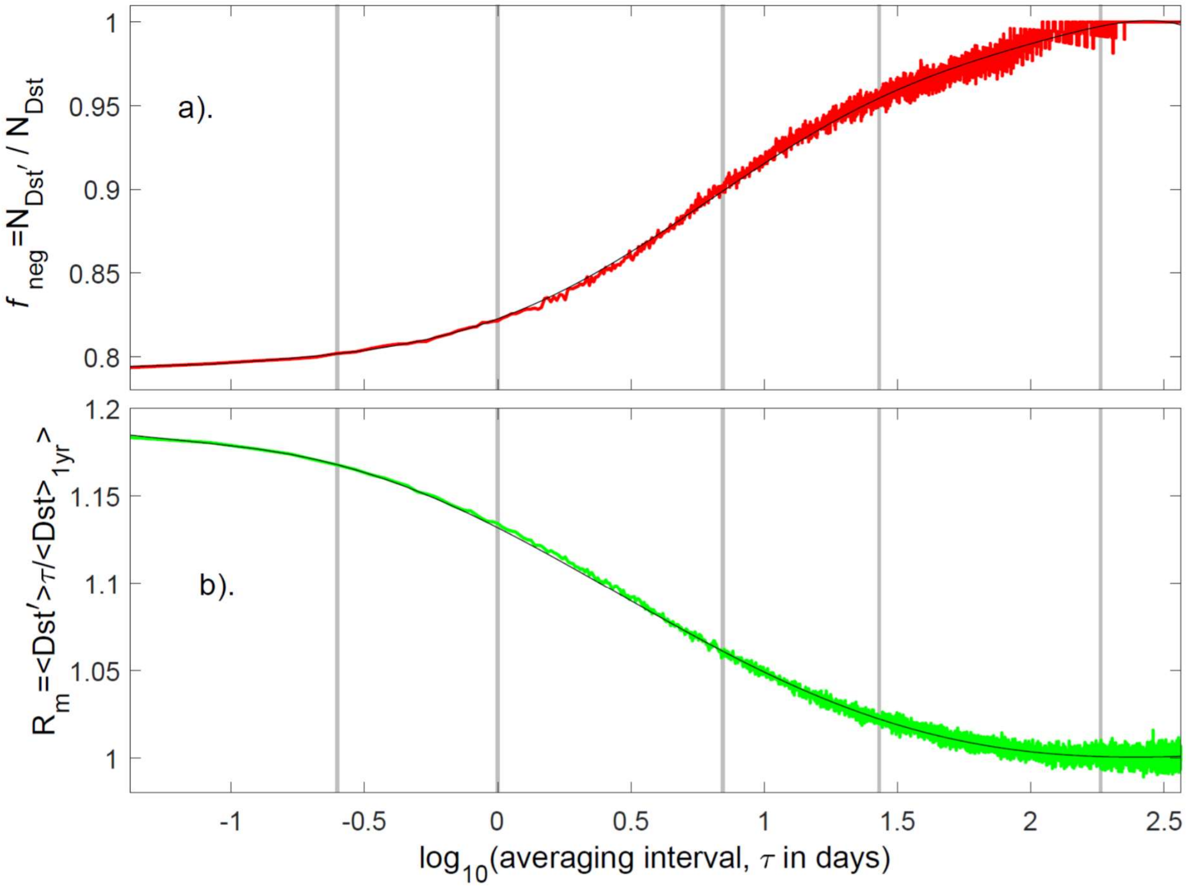

Figure 14. The variation with averaging interval τ of (top) the fraction of Dst samples that are negative (the subset termed *Dst'*) and (bottom) the mean of the ratio of the mean value of *Dst'* in intervals of duration τ, to the annual mean values of *Dst*. (a) $f_{neg} = N_{Dst'} / N_{Dst}$ is shown as a function of $\log_{10}(\tau)$, where $N_{Dst'}$ is the number of samples at that τ for which Dst ≤ 0 and N_{Dst} is the number of *Dst* samples of either sign. The red line is the mean for all *Dst* samples (from 1957-2016), the black line is best polynomial fit (see Appendix A for details). (b) $R_m = <Dst>_\tau/<Dst>_{1yr}$ is shown as a function of $\log_{10}(\tau)$. The green line shows the result for all the data (from 1957-2016), the black line is best polynomial fit (see Appendix A for details).

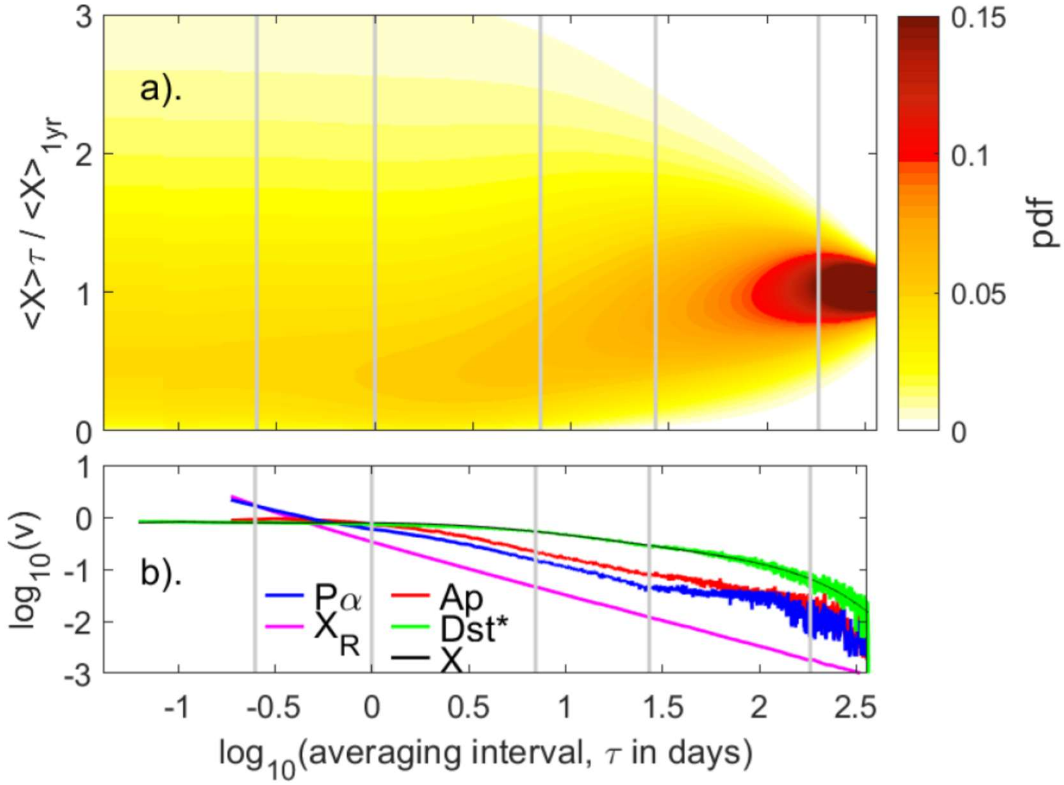

Figure 15. Same as Figure 9 for a model X based on Weibull distributions and a 6th-order polynomial fit to the variance of Dst', $v(\tau)$. Note that by only considering the negative *Dst* values (Dst') the mean values of the fitted distributions are $R_m(\tau)$ rather than unity and pdfs have also been multiplied by f_{neg} to allow for existence of positive values – in both cases, the values used here from the polynomial fits shown in Figure 14. In (b) the green line shows $v(\tau)$ for Dst' (on a logarithmic scale) and the black line is the polynomial fit (see Appendix A for the polynomial coefficients and *formulae* for the Weibull distribution family). Also shown are the $v(\tau)$ variations for other variables using the same color scheme as used in Figures 4b, 5b and 6b: P_α (in blue); *ap* (in red); random variable, X_R (in mauve).

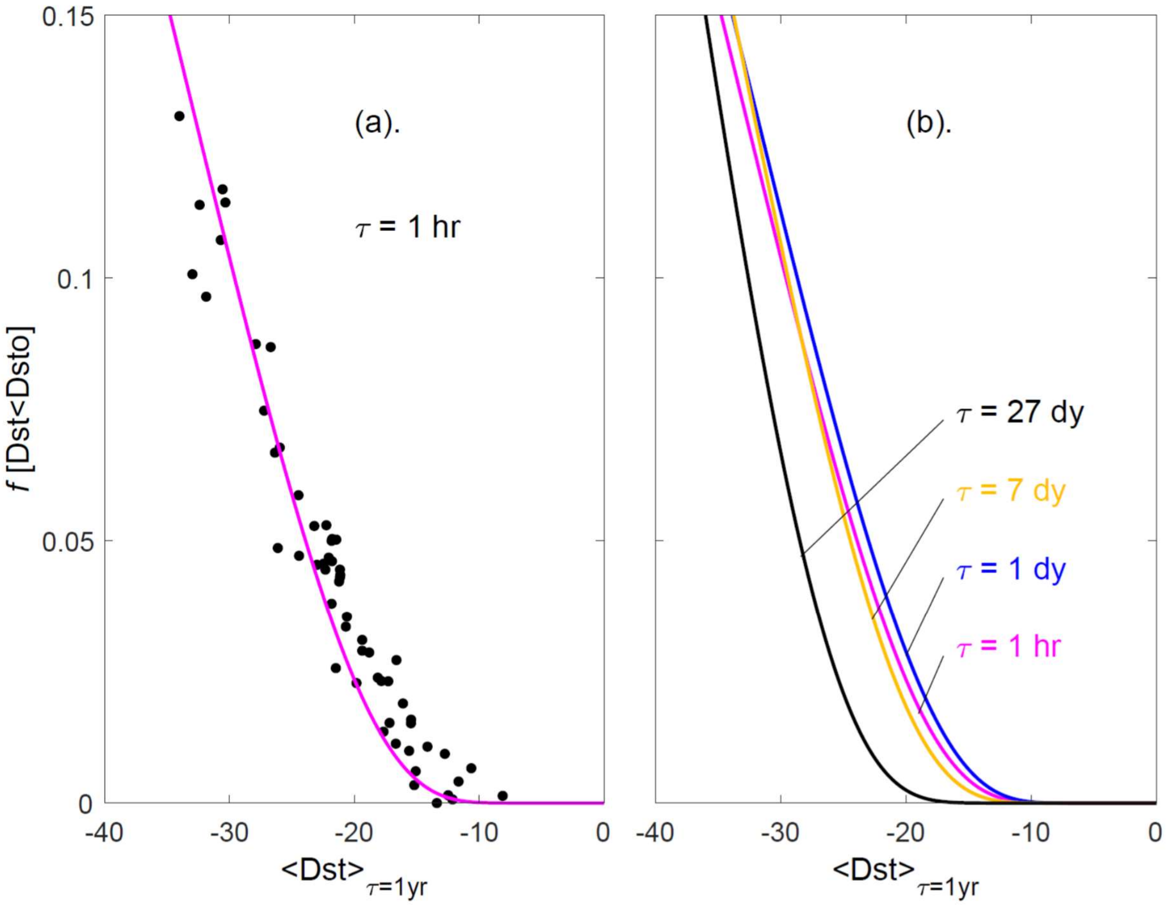

Figure 16. Same as Figure 12 for predictions by the model fit to the *Dst* distributions with τ shown in Figure 15. (a) The points show the observed probability that 1-hour values of *Dst* are in the top 5% of the overall distribution of *Dst* disturbance levels (for 1957-2016, 525960 samples), *f*[*Dst*<*Dst*o] (i.e. *Dst* is less than its 5-percentile of 1-hourly values, *Dst*o = –55.14 nT), as a function of the annual mean value of *Dst* values $<Dst>_{\tau=1yr}$. The mauve line is the model prediction for τ = 1hrs. (b). The family of model predictions of *f* [*Dst*<*Dst*o] as a function of $<Dst>_{\tau=1yr}$ for timescales τ of 1 hour (in mauve), 1day (in blue), 7 *days* (in orange) and 27 days (in black).

Appendix A. Probability distributions of *ap* and *Dst*

In the paper, we make use of two distribution forms, the Lognormal and the Weibull

(A-i). The equations of the Lognormal Distribution

For the lognormal distribution the two parameters that are usually used to specify the distribution are μ and σ. These are, respectively, the mean and standard deviation of the normal distribution in $\log_n(x)$ where x is the variable that is lognormally distributed. These are related to the mean m and variance v of x by

$$m = \exp(\mu+\sigma^2/2) \qquad \text{(A1)}$$

$$v = [\exp(\sigma^2-1)]\times\exp(2\mu+\sigma^2) \qquad \text{(A2)}$$

or conversely expressing μ and σ in terms of m and v we have

$$\mu = \log_n (m / (1 + v/m^2)^{1/2}) \qquad \text{(A3)}$$

$$\sigma^2 = \log_n (1 + v/m^2) \qquad \text{(A4)}$$

Hence specifying a lognormal distribution using μ and σ is precisely the same a specifying it using m and v . The advantage of using μ and σ is that the equation for the probability distribution of a lognormal is simpler:

$$f(x) = \{1/x\}\times\{1/(2\pi\sigma^2)^{1/2}\} \times\exp\{(-\log_n(x)-\mu)^2/ (2\sigma^2) \qquad \text{(A5)}$$

For any one combination of m and v, we compute μ and σ using equtions (A3) and (A4) and hence determine the full distribution using (A5).

(A-ii) The equations for a Weibull Distribution

For Weibull distribution (also called the Rosin Rammler distribution), the two parameters used to describe the distribution are a scale parameter λ and a shape parameter k. (Note that both λ and k are always positive).

The mean and variance of the distribution in x are again m and v, where

$$m = \lambda\Gamma(1+1/k) \qquad \text{(A6)}$$

$$v = \lambda^2 \{\Gamma(1+2/k) - (\Gamma(1+1/k))^2\} \qquad \text{(A7)}$$

Where Γ is a gamma function. The converse equations for λ and *k* cannot be derived analytically and we solve them iteratively by varying the shape parameter *k* until

$$\lambda = m / \{\Gamma(1+1/k)\} \qquad \text{(A8)}$$

and

$$\Gamma(1+2/k) = (v + m^2) / \lambda^2 \qquad \text{(A9)}$$

and then checking the full range of allowed *k* for a given *v* and *m* that the solution is unique.

The Weibull distribution is:

$$f(x) = (k/\lambda)\times(x/\lambda)^{k-1}\times\exp\{-(x/\lambda)^{k}\} \quad \text{for } x \geq 0$$

$$f(x) = 0 \quad \text{for } x < 0 \qquad \text{(A10)}$$

Hence, as for the logormal, the distribution is described by two parameters (μ and σ for a lognormal and *k* and λ for a Weibull) and in both cases specifying that pair is fully equivalent to specifying the mean and the variance. Note that in the paper we fit variables of the form $X/\langle X\rangle$ and so the mean value is $m = 1$ and the one fit variable is the variance *v*. The remainder of this Appendix gives the models used to generate the probability distribution functions, as a function of averaging timescale, τ, for the *ap* and *Dst* geomagnetic indices, shown in Figures 9 and 15, respectively.

(A-iii) Model for *ap*

The polynomial fit to the variation of the logarithm of the variance, *v*, with timescale τ for the *ap* index, shown by the black line in Figure 9b, gives

$$\log_{10}(v) = \beta = -0.0471\tau^6 + 0.1309\tau^5 + 0.0954\tau^4 - 0.3554\tau^3 - 0.1651\tau^2 - 0.2124\tau + 0.2048 \qquad \text{(A11)}$$

such that the model variance is

$$v(\tau) = 10^{\beta} \qquad \text{(A12)}$$

By normalizing the *ap* values by the annual mean $\langle ap\rangle_{\tau} / \langle ap\rangle_{\tau=1\mathrm{yr}}$, the annual distributions have a mean $m = 1$ at all τ

For *ap* the best fit is with the family of lognormal distributions.

$$\eta = 1 + (v/m^2) \qquad \text{(A13)}$$

$$\mu = \log(m/\eta^{0.5}) \qquad \text{(A14)}$$

$$\sigma = \log_n^{0.5}(\eta) \qquad \text{(A15)}$$

$$x = \langle ap \rangle_\tau / \langle ap \rangle_{\tau=1\mathrm{yr}} \qquad \text{(A16)}$$

$$a = (x\sigma(2\pi)^{0.5})^{-1} \qquad \text{(A17)}$$

$$b = \exp\{(-(\log(x) - \mu^2)/(2\sigma^2)\} \qquad \text{(A18)}$$

$$f(x,\tau) = ab,\ f(0,\tau) = 0 \qquad \text{(A19)}$$

The equations (A11) - (A19) allows the computation of the pdf f for a value of ap for an averaging timescale τ, $\langle ap \rangle_\tau$, if we know its annual mean, $\langle ap \rangle_{\tau=1\mathrm{yr}}$.

Comparison of Figures 5a and 9a of the main text demonstrate the fit of the family of distributions to the *ap* data.

(A-iv) Model for Dst

The polynomial fit to the variation of the logarithm of the variance, v, with timescale τ for the *Dst* index, shown by the black line in Figure 15b, gives:

$$\beta = -0.0158\tau^6 + 0.0353\tau^5 + 0.0462\tau^4 - 0.1283\tau^3 - 0.1387\tau^2 - 0.0318\tau - 0.1060 \qquad \text{(A20)}$$

such that the model variance is

$$v(\tau) = 10^\beta \qquad \text{(A21)}$$

The fraction of Dst' samples (with $Dst \leq 0$), as a function of timescale τ is given by the polynomial (the black line in Figure 14a)

$$f_{\mathrm{neg}} = -0.0003\tau^8 + 0.0004\tau^7 + 0.0035\tau^6 - 0.0039\tau^5 - 0.0161\tau^4 + 0.0052\tau^3 + 0.0461\tau^2 + 0.0578\tau + 0.8226 \qquad \text{(A22)}$$

(Note that such a high-order polynomial is needed to capture the observed variation with sufficient accuracy).

The polynomial fit to the ratio of the means of Dst' for intervals of length τ, $\langle Dst' \rangle_\tau$ (where Dst' is the subset of *Dst* values that are negative), and the annual mean of *Dst*, $\langle Dst \rangle_{1\mathrm{yr}}$ given by the black line in Figure 14b, is

$$R_m = \langle Dst' \rangle_\tau / \langle Dst \rangle_{1\mathrm{yr}} = 0.0003\tau^6 - 0.0024\tau^5 + 0.0033\tau^4 + 0.0145\tau^3 - 0.0215\tau^2 - 0.0770\tau - 1.1319 \qquad \text{(A23)}$$

For Dst', the best fit is with the family of Weibull distributions, the variance of which is

$$v(k) = \lambda^2 \{\Gamma(1+2/k) - (\Gamma(1+1/k))^2\} \qquad \text{(A24)}$$

where Γ is a gamma function. The best method is to find the factor *k* is by iteration to the value that gives

$$v_{\mathrm{m}}(\tau) = v(k) \qquad \text{(A25)}$$

Note that the mean of the distribution is, unlike for the *ap* case, not in general unity because of the exclusion of the positive *Dst* values. Rather, the mean is R_{m} given by equation (A23). This yields

$$\lambda = R_{\mathrm{m}}/\Gamma(1+1/k) \qquad \text{(A26)}$$

$$x = \langle Dst' \rangle_{\tau} / \langle Dst \rangle_{\tau=1\mathrm{yr}} \qquad \text{(A27)}$$

$$a = k/\lambda \qquad \text{(A28)}$$

$$b = (x/\lambda)^{k-1} \qquad \text{(A29)}$$

$$c = \exp(-(x/\lambda)^{k}) \qquad \text{(A30)}$$

$$f_{\mathrm{W}}(x,\tau) = f_{\mathrm{neg}}.a.b.c \quad \text{(always valid as } x \geq 0\text{)} \qquad \text{(A31)}$$

The normalising factor f_{neg} (given by equation (A22) for a given τ) is needed because the product of the terms *a*, *b* and *c* gives the pdf of *Dst'* , but they are only a fraction f_{neg} of the whole *Dst* sample. The equations (A10) - (A21) allows the computation of the p.d.f. *f* for a negative value of *Dst* for an averaging timescale τ, $\langle Dst' \rangle_{\tau}$, for an annual mean of *Dst*, $\langle Dst \rangle_{\tau=1\mathrm{yr}}$.

Comparison of Figures 13a and 15a demonstrate the fit of the family of distributions to the Dst data.

Appendix B. Relationship of daily means of *aa* and *ap* and correcting *ap*

Figure B-1 shows scatter plots of daily means of the *ap* index (by convention referred to as *Ap*) in 3-month intervals a function of the simultaneous daily mean of the *aa* index. This plot is restricted to data from between 1932 (the start of the *ap* index data) and 1956 (inclusive). The end date is because in 1957 there is a calibration error in *aa* introduced by the move of the northern hemisphere *aa* station from Abinger to Hartland. This has been corrected using the *ap* index by *Lockwood et al.* [2014] and *Matthes et al.* [2017] – hence it is not appropriate to use data for 1957 and after, either with or without that correction. There is considerable scatter about the trend in figure B-1, much of which is introduced by different annual responses of the two indices associated with the different geographic distribution of stations. Note there are also considerable diurnal differences also, but there are averaged out by taking daily means (which are *Ap* for *ap and Aa* for *aa).* The relationship between *Aa* and *Ap* depends on time-of-year (see Figure B-1) and the best-fit polynomials to the data for 4 fraction of year intervals, each covering a quarter of a year and centred on the times of the March equinox, June solstice, September equinox and December solstice are:

$$0.09 \le F \le 0.34 \qquad Ap^* = (7.241\times10^{-7})Aa^{*3} - (1.351\times10^{-3})Aa^{*2} + 1.108Aa^* - 8.410 \qquad \text{(B1)}$$

$$0.34 \le F \le 0.59 \qquad Ap^* = (8.959\times10^{-7})Aa^{*3} - (1.597\times10^{-3})Aa^{*2} + 1.182Aa^* - 9.236 \qquad \text{(B2)}$$

$$0.60 \le F \le 0.85 \qquad Ap^* = (7.131\times10^{-7})Aa^{*3} - (1.344\times10^{-3})Aa^{*2} + 1.127Aa^* - 8.539 \qquad \text{(B3)}$$

$$F \le 0.12 \text{ or } F \ge 0.8 \quad Ap^* = (6.621\times10^{-7})Aa^{*3} - (1.156\times10^{-3})Aa^{*2} + 0.907Aa^* - 4.969 \qquad \text{(B4)}$$

These polynomial fits and plus and minus their 2-sigma errors are shown in Figure B-1 (as solid and dashed lines, respectively). For the estimated *Aa** of the Carrington event [*Cliver and Svalgaard*, 2004], these fits yield *Ap** of 275±24, 277±44, 283±30 and 224±33 for the March equinox, June solstice, September equinox and December solstice data, respectively.

Our research into the response functions of geomagnetic indices (the collective response of the network of stations used to generate them and of the compilation algorithm used to combine the data from them) using the model of *Lockwood et al.* [2018d, e] has shown that the *am* geomagnetic index has a very flat, almost ideal, time-of-day/ time-of-year response. This is achieved because this index employs relatively uniform rings of mid-latitude stations in both hemispheres and uses weighted means to account for any spatial non-uniformity of the station network. On the other hand, the compilation of the *ap* index employs an irregular network of predominantly northern hemisphere (mainly European) stations and look-up tables to convert the observations from each into the value that would be seen at the reference Niemegk station before combining them by averaging. The look-up tables are specific to the station location and depend on time-of-day (UT), time-of-year (F) and the level of the activity. *Cliver and Svalgaard* [2004] recognized the value of the *am* index, compared to indices derived from less-ideal distributions of stations, and used it to correct for the false time-of-day variation in the *aa* index (and so created what they termed aa_m). However, they did not correct for the associated spurious time-of-year variation in *aa* [*Lockwood et al.*, 2018e] and then used the suggestion of *Allen* [1982] of 24-hour running means of aa_m (which they termed Aa_m*) which largely suppresses the false UT variation anyway. We here apply the same philosophy that *Cliver and Svalgaard* [2004] adopted, but use *am* to correct for any false time-of-year variation in *ap*. We do this because the am index data only extends back to 1959 whereas the ap index is available from 1932 onward.

We have generated a corrected *ap* index, ap_C, which allows for effects as a function of the fraction of each year (F) and the *ap* level using the formula

$$ap_\mathrm{C}(F) = ap(F) \times \underline{C}_\mathrm{ap}(F,ap) \qquad \text{(B5)}$$

where the correction factor is given by

$$\begin{aligned} C_\mathrm{ap}(F,ap) &= (\langle am(F,ap)\rangle_\mathrm{bin} / \langle am\rangle_\mathrm{all}) / (\langle ap(F,ap)\rangle_\mathrm{bin} / \langle ap\rangle_\mathrm{all}) \\ &= (\langle am(F,ap)\rangle_\mathrm{bin} / \langle ap(F,ap)\rangle_\mathrm{bin}) \times (\langle ap\rangle_\mathrm{all} / \langle am\rangle_\mathrm{all}) \qquad \text{(B6)} \end{aligned}$$

The subscript “all” refers to the averaging of all co-incident *ap* and *am* data for 1959-2017 (inclusive) and the subscript “bin” refers to the averaging of data in a given F and *ap* bin during the same interval. Multiplying by the ratio of the all-over means of *ap* and *am* means that we correct for the variation with F but do not change the average levels of *ap*. In practice, the data were divided into 40 percentiles of the overall *ap* distribution, giving 6282 samples in each *ap* bin, the values of $C_\mathrm{ap}(F,ap)$ were then fitted with a 6th order polynomial in F. The derived correction factor $C_\mathrm{ap}(F,ap)$ is shown as a function of F (x axis) and $\log_{10}(Ap)$ (y axis) in figure B-2. Note that we are not concerned with any limitations in the UT dependence of the response of *ap* because we use averages over 24-hour intervals, as discussed below. This correction is only approximate because the network of stations used to generate the *ap* index has changed several times since 1932. However, we do not find any detectable discontinuities in $C_{ap}(F,ap)$ at any of the changes since 1959 and so we use the assumption that effects of changes before this date also have negligible effect. The effect of the correction is not great (see Figure B-3) but is largest for the most active days. Many of these storm day values are hardly altered by the correction but those in northern hemisphere winter, in particular, are underestimated in *ap* and this is corrected in ap_C.

We follow the procedure of *Allen* [1982] to make 24-hour boxcar means of ap_C, Ap_C*. For the purposes of identifying and ranking storm days we take the largest value of the 8 such running-means in each calendar day $[Ap_C*]_{MAX}$. The 100 largest values of $[Ap_C*]_{MAX}$ since 1932 are given in rank order in Table S7 of the Supporting Information file. Although there are similarities, this list has a somewhat different ranking order to previous studies [e.g., *Nevanlinna et al.*, 2006; *Kappenman*, 2005; *Cliver and Svalgaard*, 2004], largely because of the allowance we make for the variation of the *ap* index response with time of year. Note that even quite small changes in the estimated magnitude of the storm day can have a very large effect on its position in the ranking order.

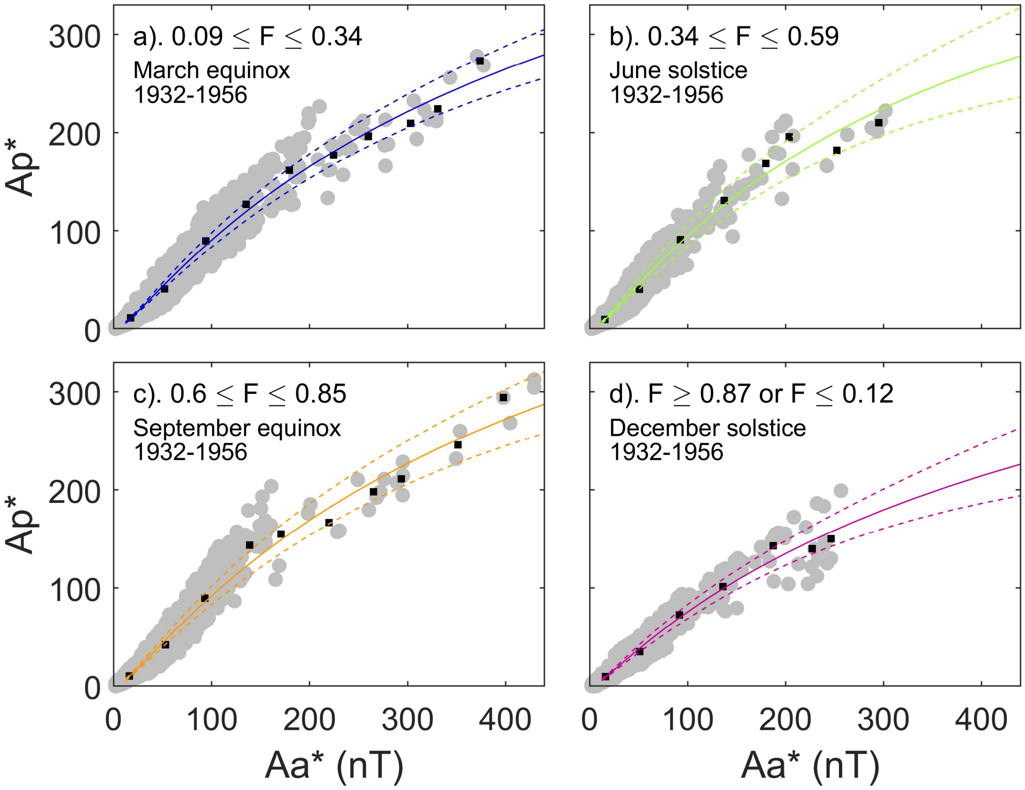

Figure B-1. Scatter plots of 24-hour means of the *ap* geomagnetic index, *Ap**, as a function of the corresponding means of the *aa* index, *Aa**, for 1932-1956 (inclusive) for 0.25-yr intervals around (a). March equinox; (b) June solstice; (c) September equinox and (d) December solstice. Black squares are means over aa bins 40 nT wide. The solid lines are third order polynomial fits and the dashed lines are plus and minus the best-fit 2-sigma error.

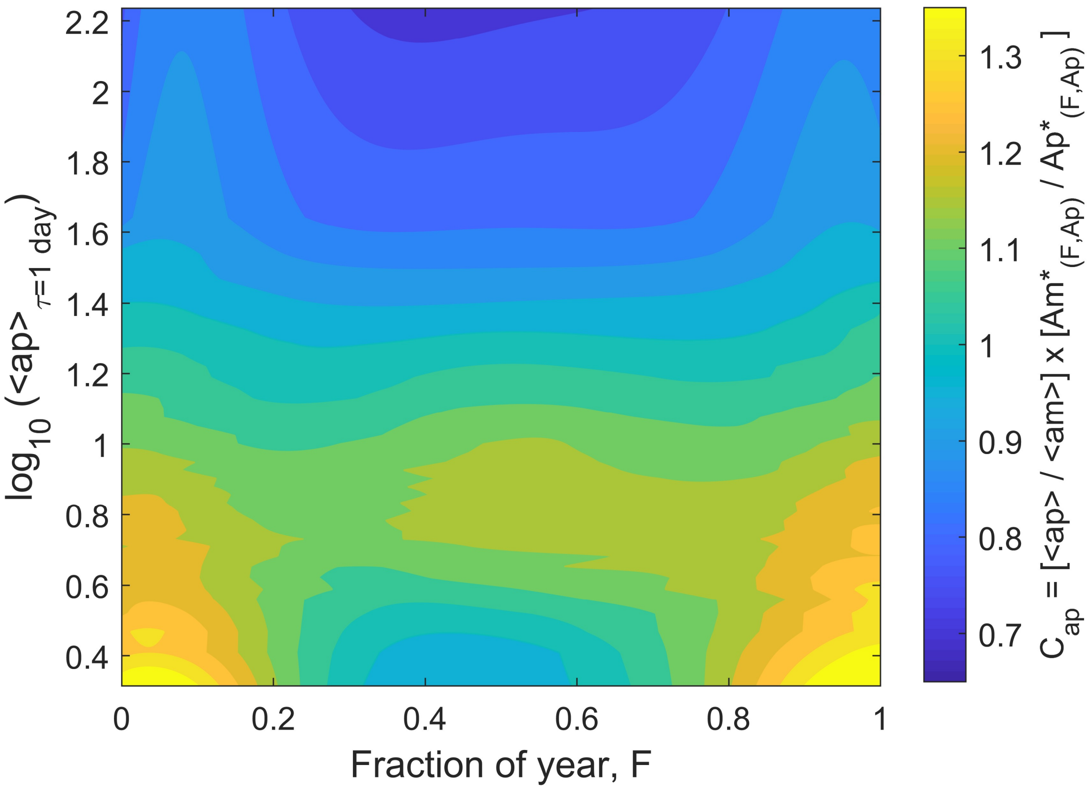

Figure B-2. The *Ap** correction factor $C_{ap} = (Am^*/Ap^*).(<ap>/<am>)$ as a function of the time of year, *F*, and the *ap* level (shown here on a logarithmic scale) derived from all the coincident *ap* and *am* index data (for 1959-2017, inclusive).

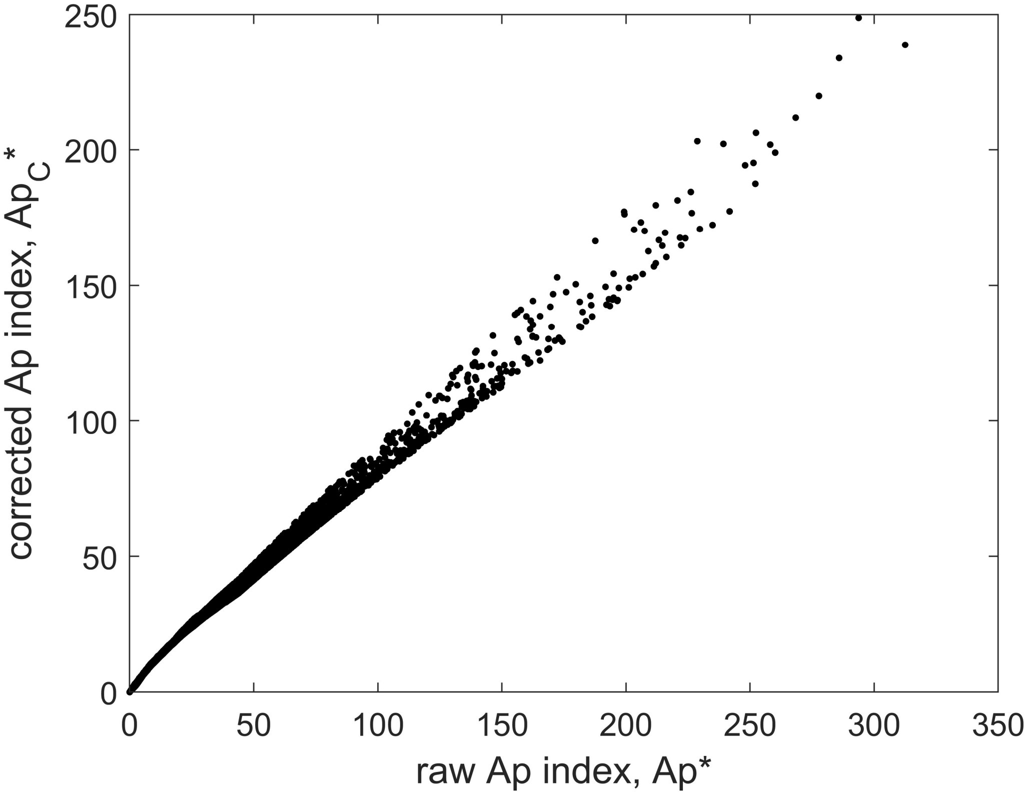

Figure B-3. The effect of correcting 24-hour means of the *ap* index for its dependence on time of year, *F*: a scatter plot of Ap_C* (8-point running means of the corrected $ap_C = ap.C_{ap}$) as a function of the corresponding running means of the original *ap* values, *Ap**. The plot is for all *ap* index data to date (1932-2017, inclusive).