File: Lockwood_SW_S1.doc
Submitted to Space Weather
Date: 21 November 2021

Solar wind - magnetosphere coupling functions: pitfalls, limitations and applications

Mike Lockwood [1*]

[1] Department of Meteorology, University of Reading, Reading, U.K.

*** Correspondence:**
Corresponding Author
m.lockwood@reading.ac.uk

Keywords: Solar wind, magnetosphere, coupling functions, performance metrics, activity distributions.

Abstract

Solar wind-magnetosphere coupling functions have been in use for almost 50 years. In that time, a very large number of formulations have been proposed. As they become increasingly subsumed into systems analysis and machine-learning studies of the magnetosphere, it is timely to establish best practice in their derivation and their limitations. This paper carried out a number of studies to establish some key points. Particular attention is paid to the best metric used to evaluate their performance and how it depends on the application for which the coupling function is intended.

Plain Language Summary

Coupling functions are mathematical combinations of variables observed in the solar wind, just before it impacts near-Earth space. They are used to predict the effect that the solar wind will have (or, for retrospective studies, will have had) on the space-weather environment of the Earth. This paper reports some studies aimed at improving their performance, how to best test their performance for a given application, and which define best practice in their derivation and application.

1 Introduction

Coupling functions are widely-used constructs in space physics designed to predict, or to retrospectively analyze, the effect of given set of solar wind conditions incident upon the near-Earth space environment, the magnetosphere. They do not try to allow for every physical mechanism involved explicitly, rather they attempt to capture and amalgamate the key drivers and explain a large fraction of the variance of a terrestrial space weather index or indicator. Correlations between interplanetary parameters and terrestrial disturbance indices became possible after the first spacecraft to visit interplanetary space had acquired sufficient data (e.g., *Arnoldy*, 1971) and the concept of combining parameters into a coupling function that allows for the different influences on terrestrial space-weather disturbance was first introduced in the PhD thesis of *Perreault* (1974). This led to the much-used "epsilon factor" coupling function, ε (*Perreault & Akasofu*,1978). Unfortunately, there was an error in the theoretical basis for ε (*Lockwood*, 2019) which causes it to perform significantly less well than other coupling functions on all timescales (*Lockwood & Finch*, 2007). A large number of alternative formulations have been proposed since (see reviews by *McPherron, et al*., 2015 and *Lockwood & McWilliams*, 2022). Some of these coupling functions are based on theory, others are empirical fits to observations. In reality, most are a mixture of both approaches, with theory guiding the selection of parameters, and the mathematical formulation used to combine them, for empirical coupling functions, whereas theoretically-derived coupling functions often use coefficients, branching ratios or exponents that are taken from observations. Coupling functions have also been derived and/or tested using global numerical MHD simulations of the magnetosphere (e.g., *Wang et al.*, 2014).

For all coupling functions, correlation with terrestrial space weather disturbance index has traditionally been used as the metric by which their merit and performance is evaluated. In the past not much attention was paid to the effects of this choice of performance metric, nor the effects of averaging timescale, nor the fact that different parts, features and indices of the coupled magnetosphere-ionosphere-thermosphere system respond differently to a given set of conditions. In addition, when building a space weather climatology, we will need to know the form of the occurrence probability distributions of indicators of space weather phenomena to predict probabilities of certain conditions, events and integrated effects (*Lockwood et al.*, 2019b, 2020) and little attention has been given to matching the distributions of a proposed coupling functions to those of the space weather indicator that they are designed to predict. Studies of coupling between the solar wind and the magnetosphere are now starting to apply systems analysis and machine learning techniques (e.g., *McGranaghan et al*., 2017; *Camporeale*, 2019; *Borovsky and Osmane*, 2019; *Stephens et al.*, 2020; see also collection of papers edited by *Camporeale et al*., 2018). This makes it very timely to take a detailed look at coupling functions, and the pitfalls inherent in their use, so that any mistakes and limitations are not carried forward and built into these new techniques.

A limitation of correlation studies that had not received much attention, at least until recently, is "overfitting" (*Chicco*, 2017). This is a recognized pitfall when signal-to-noise ratio in data is low, as is often the case in disciplines such as climate science (*Knutti et al.*, 2006) or population growth and ecology (*Knape & de Valpine*, 2011). Overfitting occurs when a fit has too many degrees of freedom and it can start to fit to the noise in the training data which, by definition, is not the same as the noise in the test or operational data. As a result, the fit has reduced predictive accuracy and power. Overfitting is a particular problem for the generation of coupling functions because there are a great many sources of noise, not all of which have been recognized and some of which we cannot do much about, particularly considering the need to have large datasets to cover all potential regions of solar wind/magnetosphere parameter space.

In correlative studies of solar wind-magnetosphere coupling some major sources of noise are:

• Measurement errors and limitations in the interplanetary observations

• Measurement errors and limitations in the observations of the terrestrial space weather indicator

• Propagation errors. The lag between the interplanetary observations and the terrestrial response can generally be accommodated by locking at the variation of the performance metric with applied lag between the interplanetary and terrestrial data (for example using a lag correlogram, the correlation coefficient as a function of lag). More of a problem is spatial structure and/or temporal evolution in the solar wind and/or non-radial solar wind flow. All of these can cause the solar wind/interplanetary magnetic field (IMF) detected by the spacecraft to be different to that incident upon the magnetosphere. In this context, we must remember that many upstream monitor satellites are in large halo orbits around the L1 Lagrange point and so are further from the Sun-Earth line than satellites in geocentric orbits around the Earth.

• Effects of the bow shock and magnetosheath. In particular, the orientation of the shocked IMF in the magnetosheath at the dayside magnetopause is a critical factor in the coupling of energy, mass, and momentum into the magnetosphere and this may not always be simply related to the IMF orientation in the undisturbed solar wind. This difference is very likely to be highly dependent on the averaging timescale, τ.

• Systematic errors introduced by Earth's orbital characteristics. These include seasonal effects on the global conductivity distribution of the ionosphere, dipole tilt effects at the dayside magnetopause, on the tail lobes, on instantaneous antisunward flux transfer (and hence transpolar voltage) (Lockwood et al., 2020d) and the cross-tail current sheet and, in theory, even annual changes in the Sun-Earth distance.

• Data gaps. These are often ignored on the grounds that their effects average out. That is not entirely the case because they are source of noise in correlation studies. In particular, they facilitate overfitting. *Lockwood et al.* (2019a) demonstrated errors (both random and systematic) introduced into optimum coupling functions by introducing synthetic data gaps into near-continuous data.

• Short data series. If the training data do not adequately cover the range of possible values the applicability of the coupling function will be compromised. Larger datasets also give greater statistical significance to fits, allow higher time resolution studies and give lower uncertainties.

• Time history and pre-conditioning. The fundamental idea inherent in the derivation of most coupling functions is that there is a given terrestrial response to a given set of upstream conditions. In practice we known that, for example, the response of the tail in generating substorms to a given set of conditions is different after a prolonged period of northward IMF (that leaves a low open magnetospheric flux) compared to that following period of southward IMF that has generated a large open flux. There are several other pre-conditioning mechanism that have been proposed, which are discussed in section 13 of this paper. Preconditioning effects become a greater factor at lower averaging timescales, τ.

Several of these noise sources raise the issue of averaging timescale (τ) used for both the interplanetary and the terrestrial data. Correlations increase dramatically with averaging timescale, so that whereas correlation coefficients of r of 0.7 (i.e., explaining just r^2 = 0.49 of the variance of the terrestrial indicator) is a good achievement for τ = 1 min., values of r of 0.98 (i.e., explaining r^2 = 0.96 of the variance of the terrestrial indicator) are readily achieved for τ = 1 year. There are a number of reasons for this. The greater numbers of samples means that the effects of random transient fluctuations are reduced and statistical significance increased. In addition, observation and propagation errors are averaged out and systematic orbital errors are reduced (and even averaged out completely for τ that is an integer number of full years). In addition, short-term preconditioning effects are averaged out.

The purpose of this article is not to evaluate and compare the many individual coupling functions that have been proposed. In general, they are similar in that the differences in their performance are relatively minor and often not statistically significant; furthermore, a coupling function that generates the largest r^2 for one terrestrial indictor is often not optimum for another or for a different averaging timescale. Rather, this paper looks at general principles, limitations, and pitfalls.

2 Terrestrial disturbance indicators and indices

This paper aims to exploit the large dataset of interplanetary observations made since 1995, because it has data gaps that are both much fewer in number and much shorter in duration than before this date (*Lockwood et al*., 2019a). This requires comparison with terrestrial space weather indicators that have been available almost continuously throughout that interval and that are homogeneous, in that they have not changed to any significant extent in their accuracy, resolution, region of coverage, method of construction, or dynamic range. These studies also require indicators that are, where possible, global so that results are not specific to a restricted region and ideally have no seasonal effects. This does not leave a great many possibilities.

For global geomagnetic indices there are the *kp* (and corresponding *ap*) and *am* indices, both derived from the range of variation in the horizontal field component in 3-hour intervals, as detected by mid-latitude stations. Of these *kp* (and hence *ap*) are not suitable, partly because their construction involves mapping the observations back to what would have been observed by the reference Niemegk station using look-up tables before averaging. This imprints the characteristics of the Niemegk site and location on the index and so, although they are measured by a network of stations across the globe, *ap* and *kp* do not have a global response. In addition, the network of stations used to generate them is clustered and not uniform (and predominantly northern hemisphere) and has changed several times during the interval of interest. In contrast, the compilation of the *am* index has remained relatively homogeneous and employs two rings of nearly equi-spaced stations at mid-latitudes, one in each hemisphere (*Mayaud*, 1980). It also deploys weighting functions to minimize the effects of the relatively small inhomogeneities in the rings (in particular, the large longitudinal gap in the southern hemisphere ring caused by the Pacific ocean). This makes the response of the *am* index highly constant as a function of both Universal Time (*UT*) and time-of-year, whereas *kp* and *ap* have strong *UT* and time-of-year variations which would be sources of noise in correlation studies (*Lockwood et al*., 2019d). Hence *am* is ideal for coupling function studies, other than its major limitation that it is only 3-hourly in resolution.

For higher time resolution geomagnetic indices, we have 1-minute values of the auroral electrojet indices, *AU*, *AL* and *AE* (*Davis & Sugiura*, 1966). These are generated by a ring of 12 auroral stations in the northern hemisphere. Southern hemisphere equivalents have been generated for limited intervals but large longitudinal gaps between stations, caused by oceans, give them a strong *UT* variation (*Maclennan et al*., 1991; *Weygand and Zesta*, 2008). Of particular importance is *AL* which becomes increasingly negative as the nightside auroral electrojet intensifies, making it a sensitive monitor of the substorm current wedge. A limitation of *AL* is that when terrestrial activity is high the auroral oval moves equatorward of the stations, so very large activity is underestimated. This is overcome by the SuperMAG *SML* index, constructed in the same way as *AL* but using all available northern hemisphere mid-latitude stations (typically 100 in number) (*Newell & Gjerloev*, 2011). The resulting advantages of *SML* over *AL* have been demonstrated and discussed by *Bergin et al*. (2020). We here have carried out studies using both *SML* and *AL* and results are often not significantly different and in most cases we only show the results for *SML*. The major limitation of both indices is the fact that they are for the northern hemisphere only and this gives them a seasonal variation that is a noise factor in correlation studies, which is only averaged out by averaging timescales that are an integer number of whole years. Note that, in theory, both *AL* and *SML* can have positive values; however, in the years 1996-2020 (inclusive) used here, only 53 out of 13150017 valid 1-minute *SML* samples were positive and so $-SML$ essentially behaves in the same way as a coupling function, i.e., having a minimum value of zero and increasing with the level of activity.

For studies of the ring current the *Dst* index is widely used, compiled from a ring of four near-equatorial stations. This small number of stations gives *Dst* a marked UT variation for which there are first-order corrections (*Takalo and Mursula*, 2001) but makes it is reliable only at hourly resolution. Alternatives are *SYM-H* which uses 11 low-latitude stations (9 in the northern hemisphere, 2 in the southern) and is available at 1-minute resolution) and the SuperMAG *SMR* index which uses all available stations (typically 100 in number) at magnetic latitudes between $-50°$ and $+50°$ with a magnetic latitude correction factor and is available at 1-minute resolution (*Newell & Gjerloev*, 2012). The problem with all these indices for coupling function studies is that they respond not only to the ring current but also the magnetopause currents and the tail currents so there are negative values caused by ring current enhancements and positive ones (of smaller magnitude) caused by compressions that enhance the magnetopause currents and bring them closer to the observing stations (*Burton et al*., 1975). Coupling functions on the other had have a baselevel of zero and only increasingly positive values as activity is enhanced. We here use the SuperMAG *SMR* index, but to make it suitable for coupling function studies we apply the correction that *Burton et al* (1975) applied to *Dst* to remove the effects of the magnetopause currents and so give an index, *Dst**. *Dst** is dominated by the ring current effect, with predominantly negative values that grow increasingly negative as activity is enhanced. *Burton et al*. (1975) derived $Dst^{*}=Dst-b(P_{sw})^{1/2}-c$, where P_{sw} is the solar wind dynamic pressure. Estimates of the optimum coefficients b and c vary slightly, and we employ the frequently used values by *O'Brien and McPherron* (2000) of b = 0.76 (for P_{sw} in nPa) and c = 11 nT. Hourly means of *SMR* correlate very highly with *Dst* (r=0.92) with a linear regression $Dst=1.031\times SMR-3.911$ nT which yields a correspondingly modified of *SMR* with a first order correction for magnetopause and tail currents, $SMR^{*}=7.04\times SMR-10.7$ nT, which is what we employ here.

A recent survey of data from the SuperDARN radars over the past 25 years has yielded a dataset of hourly means of the transpolar voltage, Φ_{PC} (*Lockwood and McWilliams*, 2021). However, unlike the above geomagnetic indices, it cannot be used as a continuous data series. The reason is that the "map-potential" method used to derive Φ_{PC} is a data assimilation technique employing a model of the ionospheric convection pattern, driven by the IMF orientation in the upstream solar wind. Tests against values from satellite over-passes show that an average number of radar echoes for the thirty 2-minute pre-integrations in each hour must exceed 255 for the influence of the model in the Φ_{PC} data to be reduced to an undetectable level and this condition leaves 65133 usable hourly-mean Φ_{PC} values, about one third of the total obtained over 25 years. Despite not being a continuous record and despite the fact that it is only of hourly time resolution, these data are included in the present study because magnetic flux transport (i.e., voltage) is a known to be the key and fundamental part of the coupling of solar wind mass momentum and energy into the magnetosphere

3 Compiling a coupling function

When compiling a coupling function for a given averaging timescale, very important principle that has sometimes been overlooked is that parameters should be combined at the highest resolution available and then averaged. Large errors can result if the data are averaged and then combined, particularly if they vary considerably during the averaging intervals. This general principle can be understood conceptually if, for example, we consider coupling functions that are aimed at quantifying the power input into the magnetosphere, P_{α}: over the averaging period τ, we want the total power input in that time, which by definition of the mean is

$$\int_0^{\tau} P_\alpha \, dt \;=\; \tau \times < P_\alpha >_\tau \qquad (1)$$

Similarly, if we use a coupling function aimed at quantifying the dayside reconnection voltage, Φ_D we want the total magnetic flux opened in the period τ, which is the integral of Φ_D over the interval, equal to $\tau \times <\Phi_D>_\tau$. A common functional form used by a great many proposed coupling functions (see Table 1 of *Lockwood and McWilliams*, 2022) is

$$< C_f >_\tau = < B_\perp^a \, \rho_{sw}^b V_{sw}^c \, sin^d(\theta/2) >_\tau \qquad (2)$$

Where $B_\perp$ is the transverse component of the IMF, perpendicular to the Sun-Earth line, V_{SW} is the solar wind speed, ρ_{SW} is the solar wind mass density ($\rho_{SW} = m_{SW}N_{SW}$, where N_{SW} is the number density and m_{SW} is the mean ion mass); and θ is the clock angle of the IMF in the Geocentric Solar Magnetospheric (GSM) frame of reference, defined as $\theta = tan^{-1}(|B_Y|/B_z)$. It is important to note the difference between this definition and $\theta' = tan^{-1}(B_Y/B_z)$. The angles θ′ and θ increases together from 0 to π, but as θ′ increases further from π to 2π, θ decreases from π back down to 0. Thus whereas θ′ has a discontinuous change from 2π down to zero at purely northward IMF, but there is no such discontinuous change in θ. This means that the extreme problems that arise for θ′ when averages straddle the discontinuity do not arise with θ. Adopting θ also means that there is no difference between IMF $B_Y > 0$ and $B_Y < 0$ as far as the coupling functions are concerned. This may not be adequate because power input into the magnetosphere and/or magnetopause reconnection voltage and/or tail loading and unloading could all, potentially, be asymmetric with respect to the polarity of B_Y. However, it is vital we that avoid the 2π-to-zero discontinuity so we must use the definition of θ and would have to amply a separate term employ to allow for any effects of the polarity of B_Y.

From above, we want the combine-then-average value over an interval of duration τ , $<C_f>_\tau$. In contrast, an average-then-combine procedure would yield

$$[C_f^*]_\tau \;=\; [B_\perp]_\tau{}^a . < \rho_{sw} >_\tau{}^b . < V_{sw} >_\tau{}^c . sin^d([\theta]_\tau/2) \qquad (3)$$

Which uses the inappropriate formulae

$$[\theta]_\tau \;=\; tan^{-1}(| < B_Y >_\tau | / < B_z >_\tau) \qquad (4)$$

and

$$[B_\perp]_\tau = (< B_Z >_\tau^2 + < B_Y >_\tau^2)^{1/2} \qquad (5)$$

Figure 1a of *Lockwood and McWilliams* (2022) demonstrates that $<C_f>_\tau$, from equation (2), and $[C_f^*]_\tau$, from equations (3-5), are not only different, they are often very poorly correlated. This means that, given that $<C_f>_\tau$ is what we require, an average-then-combine procedure can introduce considerable noise into the correlation and hence considerable error into the derived coupling function.

Lockwood and McWilliams (2022) show that the major errors in $[C_f^*]_\tau$, compared to the required value of $<C_f>_\tau$, largely arise from the $sin^d([\theta]_\tau/2)$ term and using $[\theta]_\tau$ from Equation (4) rather than computing θ at the highest available time resolution, then combining all the parameters to compute C_f and only then averaging. There is a similar but smaller problem with using equation (5) to compute $B_\perp = (B_Z^2 + B_Y^2)1/2$ and *McPherron et al* (2013) and *Lockwood and McWilliams* (2022) recommend it is computed at the highest available time resolution before inclusion in C_f.

There is, however, another problem that arises because, in general, there is a difference between "Hölder means" (also called "power means") $[<x^p>_\tau]^{1/p}$ of a general variable x and the corresponding arithmetic means $<x>_\tau$ and hence between $<x^p>_\tau$ and $<x>_\tau^p$. Only in the special case that the exponent $p = 1$ or that x is constant over the interval τ does and $<x>_\tau^p$ exactly equal $<x^p>_\tau$. However, *Lockwood and McWilliams* (2022) show that p is close enough to unity and the variability over the intervals of interest to make $<x>_\tau^p \approx <x^p>_\tau$ a valid approximation for x of $B_\perp$, ρ_{sw} and V_{sw}, but not for $sin^d(\theta/2)$, because of the greater variability of $sin(\theta/2)$, a problem made worse if large d is used (values of d up to 9 have been proposed in the literature). This makes it valid to use

$$[C'_f]_\tau = < B_\perp >_\tau^a . < \rho_{sw} >_\tau^b . < V_{sw} >_\tau^c . < sin^d(\theta/2) >_\tau \qquad (6)$$

This is helpful because there is a problem in using equation (2) with an iterative procedure to evaluate optimum values of the exponents. The average has to be re-computed at the start of each round of the iteration, which takes computer time when dealing with large numbers of samples (25 years' data at 1-mimnute resolution is over 13 million samples); but, more importantly, is likely to cause the iteration to fail to converge to within the required accuracy, especially when noting that the intervals that have to be excluded, and be treated as a data gap, because they do not meet a set error requirement also changes with the exponent, as is discussed below.

The procedure adopted here to implement Equation (6) is that used by *Lockwood and McWilliams* (2022). Specifically, we use a fixed d that is varied between 1 and 7.5 in steps of 0.1. For each d the values of $<sin^{d}(\theta/2>_{\tau}$ are precalculated. Equation (6) is then used with the Nelder-Mead simplex search method (*Nelder and Mead*, 1965; *Lagarias et al.*, 1998) to find the a, b, and c that maximize the performance metric of choice for each d. (*Lockwood and McWilliams* use correlation coefficient for this but other performance metrics could be used). A version of the test proposed by *Vasyluinas et al.* (1982) is the deployed to determine the value of d (which sets the required values of a, b, and c) that gives linearity of $<C_f>_{\tau}$ with $<T>_{\tau}$, where T is the optimally lagger terrestrial index that we wish to predict. This procedure is outlined in Section 9 of this paper and here we just note that the polynomial fitting used can be weighted to give linearity over the whole range of T or over a selected smaller range of T, based on the desired application of the coupling function.

Data gaps are often neglected in solar wind-magnetosphere coupling studies on the pretext that their effects average out. In reality, they add noise to correlation studies and facilitate overfitting.. *Lockwood et al.* (2019a) studied the effect on coupling functions by introducing synthetic data gap into near-continuous data (with distributions of durations drawn from pre-1995 data when data gaps were both more common and longer). Above, it was pointed out that the difference between $<x^{p}>$ and $<x>^{p}$ increases for a non-unity power p with the variability of x within the averaging period and, as explained below, that variability is also a key factor in determining what should be defined as a data gap in averaged data. We can study the variability of interplanetary parameters by looking at their autocorrelation functions (a.c.f.s) as a function of lag time. These are presented in Figure 1, which is an extended and expanded version of Figure 1a of *Lockwood et al.* (2019a) (see also survey by *Maggiolo et al.*, 2017). It can be seen that the solar wind speed, V_{SW} has the highest persistence of all the interplanetary parameters because its a.c.f. (in blue) declines least rapidly with lag time. V_{SW} also shows the strongest peaks associated with solar rotation effects with a significant peak near 27 days and clear harmonics at 54 days and 81 days. The most variable parameter, with the lowest persistence, is the IMF orientation factor $sin(\theta/2)$, the a.c.f. for which is shown in orange.

As well as the other interplanetary parameters, (the transverse IMF $B_{\perp}$ in mauve, the solar wind mass density ρ_{SW} in green, the solar wind speed V_{SW} in blue, and the IMF orientation factor $sin(\theta/2)$ in orange, Figure 1 shows the a.c.f.s for three example coupling functions. The black line is for the same example coupling function as used by *Lockwood and McWilliams* (2022) namely the estimate of the energy input to the magnetosphere estimate by *Vasyluinas et al.* (1982), P_{α}, for a coupling exponent α = 1/3 (the one free fit parameter in P_α), which yields a = 2/3, b = 1/3 and c = 5/3, for which the optimum IMF orientation term exponent is found to be d = 4 (see Section 10). In addition, the dashed black line shows the same coupling function but with d = 2 to highlight the effect of varying d. A third coupling function is presented by the cyan line: this is the empirical coupling function C_{BEA} of *Boyle et al*. (1997) which was designed to predict transpolar voltages and uses additive terms:

$$C_{BEA} = 10^{-4}\, V_{sw}^2 + 11.7\, B\, sin^3(\theta/2) \tag{7}$$

where V_{sw} is in km s^{-1} and B is in nT. Notice this requires use of an empirical "branching ratio" (10^{-4}/11.7) which is derived from a best-fit to available data and the effects of getting into parameter space beyond its applicability would be considerable as one or other of the two terms could then dominate. This coupling function C_{BEA} does not fit the commonly used formulation given in Equation (2).

It is interesting how the a.c.f.s of the interplanetary parameters influence that of these example coupling functions. For the higher d, P_α behaves rather like $sin(\theta/2)$, whereas for the lower d, the greater persistence of the other parameters reduces this tendency. Despite using a higher d of 3 than the dashed line, C_{BEA} has less variability (more persistence) because of the additive term in V_{SW}^2.#

These a.c.f.s have an important implication for how we should handle data gaps. This is investigated directly in Figure 2, which shows the root mean square (r.m.s.) percentage errors found by synthetically inserting data gaps into continuous 1-minute data and computing the error they cause in the mean as done by *Lockwood et al.* (2019) in their Figure 1b. For P_α, for example, data from the years 1996-2020 (inclusive) yield N = 5153517 10-minute intervals in which all ten 1-minute data integrations are available (in all the parameters required to compute P_α) and N= 589,293 1 hour intervals in which all 60 1-minute data integrations are available. For each of these intervals a fraction of the one minute variables were taken out at random and the mean computed for the remaining fraction f of samples and the percentage error of that compared to the known mean for all samples computed. This was repeated 10 times for each interval and the r.m.s. error, ε, for the 10N estimates for that f computed. Data gaps were introduced in such a way as to match the distribution of data gap durations that exists in the full data series. The blue lines in Figure 2a shows that the high persistence of V_{SW} means that just one of the 10 one-minute samples gives a 10-minute mean that has only a very small error (ε< 0.5%). Figure 2a shows that errors are larger for the other solar wind parameters which have lower persistence. To get ε below, for example, 2% (the lowest horizontal gray line) requires at least 60% of ρ_{SW} samples in the 10.miute interval, 90% of $B_\perp$ and C_{BEA} samples and all 10 samples for $sin(\theta/2)$ and both the P_α estimates. The very small error introduced by V_{SW} means that the errors for $B_\perp$ and C_{BEA} are almost identical. The effect on the two P_α coupling functions show that errors caused by the low persistence of the IMF orientation factor are considerably increased by a larger exponent *d*.

Figure 2b shows that errors are smaller by a factor of about 2 for the hourly means. To get the error below 2% requires at least 11% of the 60 samples for ρ_{SW}, 51% for $B_\perp$, and 75% for $sin(\theta/2)$. For the coupling functions C_{BEA} requires 62%, and P_α requires more than 91% for $d = 2$ and more than 97.5% for $d = 4$. This effect of the persistence on the required number samples in an average value has been understood and exploited in the past: in particular, telemetry requirements have been minimized by reducing the sampling rate of high-persistence parameters such as V_{SW}. However, it has not been used in coupling function studies. In particular , the exponent *d* has often been treated as a free fit parameter using the $sin^d(\theta/2)$ IMF orientation factor formulation.

The effect of increasing the exponent *d* on the errors in coupling functions due to data gaps has not been considered before. *Lockwood et al.* (2019a) showed that interpolation to fill data gaps was the worst policy for dealing with them, and just ignoring their existence was actually preferrable. Best practice is to define an acceptable limit to the error ε of the average values over the averaging interval τ and then only use samples that, from graphs like those in Figure 2, meet the corresponding requirement for fractional availability in the interval, *f*. Correlations should then be done using piecewise removal of the terrestrial data series before it is averaged at the time of the data gaps, allowing for the propagation lag. The analysis presented here has demonstrated that increasing *d* also increases the error in a coupling function introduced by data gaps. This additional noise makes overfitting more likely when *d* is large. Good practice to avoid this would be to ensure that all coupling function averages (for any *d*) are made from enough high resolution-samples to meet the set required accuracy for the highest *d* that you with to try to use. This would ensure that the number and accuracy of the averaged data used would be independent of the *d* employed, which means the merit of a given value of *d* can be tested without introducing data accuracy issues.

4 The effects of the location of the upstream monitor and the limits to predictability

As noted in section 1, one potential source of noise in correlation studies, and hence error in the derived coupling function, is the propagation of the solar wind from the monitoring spacecraft to the vicinity of the Earth. One aspect of this is the required propagation lag from the spacecraft to the dayside bow shock, where the solar wind- magnetosphere interaction sequence begins. To some extent, this lag can be allowed for by varying the time lag and then using the optimum lag δt that generates peak performance metric (if that metric is correlation, then this is the peak of a lag correlogram). The difficulty is that the optimum lag changes with solar wind speed (and to a lesser extent direction) and the IMF orientation. This means the duration of the intervals over which the correlations are taken is a factor: if this is interval is too long the variations in the true lag will introduce noise, but if the intervals is too short the correlation and its significance is reduced because the number of samples is reduced.

A potentially more significant problem is that the spatial structure in interplanetary space and/or non-radial solar wind flows and/or evolution of the conditions in propagation mean that the conditions that impinge on the magnetosphere are different from those that were observed by the upstream monitor. The most continuous data series from upstream interplanetary spacecraft comes from spacecraft in halo orbits around the L1 Lagrange point. In particular, the Advanced Composition Explorer (ACE), the Global Geoscience International Physics Laboratory (known As "Wind") and the Deep Space Climate Observatory (DSCOVR) have made observations from such orbits since 1996, 2004 and 2015, respectively. The study by *Crooker et al.* (1982) showed that distance $R_{ZY} = (Z^2+Y^2)^{1/2}$ of an L1 spacecraft (at coordinates X, Y and Z in the Geocentric Solar Ecliptic, GSE, frame of reference) from the Sun-Earth line (the X axis) had a key influence on the correlation with the corresponding data in the magnetosheath. The ACE halo orbit keeps its R_{ZY} below about $40R_E$ (where $1R_E$= 6370 km is a mean Earth radius) and for DSCOVR R_{ZY} is below about $50R_E$. Wind is in a larger Halo orbit with R_{ZY} up to about $100R_E$. The X coordinates of these spacecraft vary between 194 R_E and 264 R_E. *Walsh et al.* (2019) show that the differences between conditions in the magnetosheath and as observed from an L1 orbit increase with the R_{ZY} value of the L1 monitor.

Figure 3 studies the effect of the location of the L1 monitor on coupling function performance by looking at correlations of an example coupling function with the *AL* and *SML* indices. The top panels are for hourly means, the middle panels for 10 minute means and the bottom panels are for the basic 1-miniute integrations of the data. The coupling function used is the optimum fit of C_f (equation 2) to *SML* (which was almost identical to that for *AL*) found by *Lockwood and McWilliams* (2021) with a = 0.662, b = 0.061, c = 1.746; and d = 5.20. A basket of other coupling functions were used and the results differed only in small details and the behavior discussed here was the same in all cases. The left-hand column compares the distributions of correlation coefficient for C_f and *SML* (in blue) with those for C_f and *AL* (in red) for all 2-day intervals available in the 1996-2020 (inclusive) dataset. The distributions for *AL* and *SML* are almost identical at all three averaging timescales, τ. The correlations increase with τ, with the means and modes of the distribution both increasing. The other distributions in the left-hand panels will be discussed in the next section. The middle column of panels is for *AL* and the right hand column for *SML*: both consider the location of the L1 craft. The light grey areas are the overall distribution shown in the corresponding panel of the left hand column and the colored lines are subsets of the data sorted by the distance from the Sun-Earth line, R_{YZ}. The data are divided into five quantile ranges of R_{YZ} (i.e., each containing 20% of the data). For both *SML* and *AL*, the distributions are close to that for all cases for the four data subsets with $R_{YZ} < 81R_E$. In these cases, shown by the green, orange, blue and cyan lines, there are no systematic changes with R_{YZ}. However, there is a systematic change for the fifth quantile range $R_{YZ} \geq 81R_E$ (the mauve lines) for which r values are consistently lower. Interestingly the effect is different for *AL* and *SML*, with *AL* showing significantly more lower values of r at all three averaging timescales, whereas for *SML*, although the correlations significantly lower for $R_{YZ} \geq 81R_E$ at τ = 1 min., the effect, although still present, is much smaller for the larger τ values. It is expected that the effect of large R_{YZ} would become greater at lower τ as the averaging smooths out spatial structure in interplanetary space. We conclude that although the correlations do start to degrade at $R_{YZ} \geq 81R_E$ there is no evidence that at lower R_{YZ} the distance for the L1 monitor from the Sun-Earth line is introducing significant noise into the overall correlations.

As mentioned above, correlation coefficients are not the only metric by which a coupling function should be derived and evaluated, indeed it is sometimes not the best metric to use. A very important application of coupling functions is in predicting large space weather events, and correlation coefficient can be dominated by the core of the distribution of space weather indicators, rather than the large-event tail of that distribution. Figure 4 employs a two-dimensional normalized histogram format that, hereafter, will be referred to as a “data-density plot”: this is used in preference to a scatter plot, in which information would be lost as many data points would be overplotted because they are so numerous. The fraction of all data points $n/\Sigma n$ falling in small bins is color-coded on a logarithmic scale: in each panel of Figure 4 the bins of width 0.05 along both axes. The low end of the color scale used is chosen to be just below the "one count level" ($\log_{10}(1/\Sigma n)$, i.e., corresponding to $n = 1$) to ensure that outlier data, with just one sample in a bin, show up as a blue pixel. Overlaid on the data density plots in Figure 4 are quantile-quantile (q-q) plots. The latter are a test of how alike two distributions (of general parameters x and y) are and the points of a q-q plot line up along the $x = y$ diagonal if the distributions of x and y are identical. The form of deviations from the diagonal can be used to infer in what way the distributions differ. Figure 4 is for the *SML* index (normalized by dividing by its overall mean value) along the x axis, and the similarly normalized coupling function C_f /< C_f > in the y axis where C_f is the same as was used in Figure 3. The q-q plots use 1000 quantiles, 0.1% apart, shown by the white dots connected by the thin red line. This means that the top/rightmost white dot in each panel is for the 99.9 percentile of the distributions. In each case a lag correlogram is taken and the plot is for the optimum lag δt which gives the peak correlation, r. Both δt and r are given in each panel. Note that δt is the lag after the predicted arrival rime of the measured C_f at the nose of the bow shock . The top row is for τ = 1hr, the middle for τ = 10min., and the bottom for τ = 1min. The left-hand column is for all data, the middle column for the 20% of L1 data from closest to the Sun-Earth line ($R_{YZ} < 28R_E$) and the right-hand plots for the 20% of L1 data from furthest away from the Sun-Earth line ($R_{YZ} \geq 81R_E$). The core of the data density plots give and overall indication of the level of agreement (also given by the correlation coefficient r) but the outliers (in blue) give an idea of the scatter for the largest events. For the all-data column on the left, the q-q plot for τ = 1hr is very close to the ideal diagonal for all quantiles up to the 99.5 percentile. There is some small systematic deviation at the very lowest values, with the q-q plots below the line at the very lowest values and slightly above the above that. This is more pronounced for τ = 10min and τ = 1min and will be discussed later. However, above the 99.5 percentile (*SML*/<*SML*> above 6) we see a slight but increasing deviation toward the large C_f /< C_f >, which means the C_f distribution is bvery slightly "heavy-tailed" (also called "thick tailed" or "fat tailed"), compared to that for *SML*. In other words, this C_f tends to predict slightly too many of the very large *SML* events. The middle plot of the top row is again for hourly means but only for data taken close to the Sun-Earth line. Here the tail of the q-q plot remains close to the ideal line all the way to the 99.8 percentile and only start to show slight signs of a fat-tail C_f distribution at the 99.9 percentile. On the other hand, for the data taken furthest from the Sun-Earth line (the top right plot) the fat-tail of C_f is more pronounced and is for data above the 99.2 percentile. Hence there is a tendency for all these data to predict too many large events, but the data density plots scatter around this trend is large.

This behavior is essentially the same for τ = 10 min and τ = 1 min. and is particularly pronounced for the latter and the deviation seen for the large R_{YZ} data subset extends to close to the 50% percentile (the median, which is quite close to the mean *SML*), so the occurrent of all above-average events is overestimated. This shows the fat-tailed nature of the tail of the C_f distribution is worst at the larger R_{YZ} values for the L1 satellite location, which is consistent with peak geoeffective solar wind conditions (very large C_f) passing over the spacecraft but then missing the Earth (which probably still receives large C_f, but not as large as seen by the spacecraft) and the *SML* enhancement is not as large as we would predict from the L1 data. The tendency can also be seen from the data density plot at large C_f. Note that there are far fewer examples of the opposite happening, i.e., that the spacecraft fails to see the largest C_f that hits the Earth. This asymmetry is the cause of the deviation of the q-q plot. Hence this is more than a matter of spatial structure in the solar wind and random chance as that would enhance C_f at Earth as much and as often as reduce it.

Hence the q-q plots and the data density plots at large values (meaning typically 5 times the mean and above) indicate that increased distance from the Sun-Earth line of an L1 monitor does somewhat reduce our ability to predict or quantify the largest events. Note that this effect would not have been identified from the correlation coefficients alone.

5 The effect of spacecraft location on correlations with auroral activity indices

Walsh et al. (2019) make the point that coupling across the magnetopause depends on the properties of the near-magnetopause magnetosheath rather than those in interplanetary space and that the two differ because the solar wind and IMF are process on crossing the bow shock and passing through the magnetosheath. We here investigate this, and the effect of spatial structure in the undisturbed solar wind, using data from the THEMIS-B spacecraft. THEMIS stands for *Time History of Events and Macroscale Interactions during Substorms* and for the time interval studied here (2011-2018, inclusive), the THEMIS-B spacecraft was in geocentric orbits and between about 55 R_E and 65R_E from Earth which resulted in it being in the undisturbed solar wind for approximately 70% of the time, in the shocked solar wind of the magnetosheath about 15% of the time and inside the magnetosphere for the remaining 15%. Correlations between L1 craft and THEMIS-B data are here found in this section using the same procedure as in the previous section for L1 data, by taking the peaks of the lag correlograms for 2-day segments of data.

Because of the near-circular geocentric orbit of THEMIS-B, the undisturbed solar wind data are at distances R_{YZ} from the Sun-Earth line of between zero and 55R_E (towards both the dawn and dusk flank of the magnetosphere). A study of correlation coefficients with *SML* data revealed no consistent changes with R_{YZ} (not shown here). The THEMIS-B magnetosheath data are from between X of $-25R_E$ and $-50R_E$ (i.e., down-tail) and at similar ranges of Y values (in GSE), both positive and negative (i.e., on either flank) and at Z that precesses between $-5R_E$ and $+5R_E$. Hence these are not promising locations from which to be calculating coupling functions, when one really wants to know the sheath conditions near the nose of the magnetosphere. However, although there will be differences between the conditions in the magnetosheath near the nose and at THEMIS-B, one does at least know that THEMIS is sampling solar wind that has impacted Earth's bow shock.

The orange and green lines in the left-hand panels of Figure 3 show the correlations with the *SML* index of the example coupling function from THEMIS-B data when it was in the undisturbed solar wind (orange line) and in the magnetosheath (green line). The same coupling function is used as in the other panels of Figure 3. The orange line should be compared with the thin dashed blue line that shows the corresponding distribution of correlations for the L1 data taken at the same time as the solar wind THEMIS-B data (allowing for the optimum propagation lag). Similarly, the green line should be compared with the thin dot-dash blue line that shows the corresponding distribution of correlations for the L1 data taken at the same time as the sheath THEMIS-B data (again allowing for the optimum propagation lag).

There are a number of points to note from these comparisons that are seen at all three averaging times. Firstly, the distributions of correlations using THEMIS-B data are quite similar to the simultaneous distribution for the L1 data, especially for when THEMIS-B is in the undisturbed solar wind. Hence, it initially appears that the propagation from L1 to THEMIS-B is making only small differences to the correlations. However, it transpires that this is only true for the overall average performance demonstrated by the distribution of r values: it is not true for the individual correlations taken over 2-day intervals. For some of the 2-day intervals, the correlations for L1 data and THEMIS-B data with *SML* are essentially identical whereas in others they can differ greatly: the r.m.s. difference between the two for τ = 1min, τ = 10 min and τ = 1 hr. was 0.15, 0.15 and 0.17 when THEMIS-B was in the undisturbed solar wind and 0.28, 0.28 and 0.32 when THEMIS-B was in the sheath. Secondly, the distributions for the THEMIS data show a marked tendency to lower values than seen for the study of all L1 data (the solid blue lines). However, because the subsets of the L1 data at the same times also show the same tendency, we know that this is not predominantly a propagation effect associated with the locations of the spacecraft. Rather, detailed inspection shows that these lower correlations are caused by lower average levels of solar and space weather activity during the THEMIS-B interval (2011-2018) than during the full L1 dataset (1996-2020). Thus, coupling functions perform better when space weather activity is high. This will be demonstrated and discussed again later. However, there are differences between the performance for the L1 and Near-Earth data. These are generally small for THEMIS-B in the solar wind and the most significant is for $\tau = 1$ hr. when THEMIS in the magnetosheath, an observing location that generated a significantly better distribution of correlations than the L1 data. Thirdly, the correlations for THEMIS-B in the magnetosheath are similar to those for THEMIS-B in the undisturbed solar wind, but again there are differences for $\tau = 1$ hr. when better agreement is found with *SML* when THEMIS-B is in the sheath than in the solar wind.

At all three locations (near L1, near-Earth undisturbed solar wind and down-tail magnetosheath) averaging over 10 minutes makes only marginal improvements to correlations with *SML*, whereas averaging over an hour makes significant improvements. This is expected, given that *SML* is enhanced during substorm expansion phases which is the response of the magnetosphere to the accumulation of open flux in the geomagnetic tail during the prior growth phase (*McPherron*, 1970; *Milan et al.*, 2009) which usually lasts between about 15 and about 90 minutes (*Partamies et al.*, 2013). *Li et al*. (2013) show that for very high open flux production rates, growth phases can last less than 10 min and for very low rates more than 90 min.; from their distributions 0.1% of growth phases last less than 10 min. and whereas 64% last more than one hour but only 30% last more than 90 min. In addition, *Li et al*. show that the substorm expansion phases that follow growth phases lasting longer than an hour are considerably weaker (quantified by maximum auroral power). Hence $\tau > 1$ hr. should give higher correlation coefficients with *SML* by integrating the coupling function over the strong growth phases.

6 Comparisons of L1 data and near-Earth interplanetary and magnetosheath data indices

To understand better the correlation differences caused by spatial structure in the solar wind, Figures 5 and 6 compare directly the data observed at L1 with that observed by the THEMIS-B spacecraft during the 2011-2018 period. Figure 5 is for 10min. averages. The top row shows the distribution of correlation coefficients r between L1 data and the corresponding data recorded by THEMIS-B when in the undisturbed solar wind. The rows are for different parameters, from left to right: the IMF B, the solar wind number density N_{sw}, the solar wind speed V_{sw}, the IMF orientation factor $sin(\theta/2)$, and the example coupling function used in Figures 1 and 2, namely the *Vasyluinas et al.* (1982) estimate of the energy input into the magnetosphere P_α with a coupling exponent of $\alpha = 1/3$ and $d = 4$. The mean ion mass m_{sw} was assumed not to change between L1 and THEMIS and in computing P_α for the THEMIS-B data. In each plot, the grey area is for all data and the blue and mauve distributions look at the distribution for below two, relatively low, global space weather activity levels. These are determined using the planetary 3-hourly *am* index, averaged over the 2-day period over which the corresponding correlation was taken. The mauve histogram is for $am < 40$ nT and the blue line is for $am < 20$ nT. The behavior that can be seen in all the distributions is that the lower correlations are occurring at low activity levels, with most values below 0.5 occurring at $am < 20$ nT and almost all at $am < 40$ nT. This effect is emphasized by the bottom row of panels in Figure 5. These panels show the scatter plot of the r values as a function of average *am* for the 2-day interval they are computed over. It can be seen that lowest r values occur at low *am*. The 20nT and 40nT thresholds used in the upper panels are also shown. Superposed in cyan on the scatter plots are the mean values in 6 quantile ranges of *am*. They show that, on average, the correlation increases with larger *am* for B, N_{SW} and V_{SW}, but only very slightly for the IMF orientation factor $sin(\theta/2)$ and P_α. The rise in correlations with *am* implies that there is less small-scale structure in the interplanetary medium when activity is high, consistent with the fact that high activity is driven by large-scale coherent interplanetary structures such as Coronal Mass Ejections (CMEs) and Corotating Interaction Regions (CIRs).

The loss of correlation between L1 and THEMIS-B can be caused by spatial structure but also could reflect random instrumental errors in both or either of the measurements made by the spacecraft. Note that systematic calibration errors (in gain or offset) would not degrade the correlation between the two data series. The similarity of the correlation distributions for the different parameters strongly suggests that they have a common cause, such as spatial structure in the solar wind, rather than error in the various instruments that observed the parameters.

The middle panels are the same as the upper panels but are for THEMIS-B in the magnetosheath. The correlations with L1 data for B, N_{SW} and V_{SW} are all lowered and in these cases the global activity level quantified by *am* appears to have little effect. The reduction in r varies with the location of THEMIS-B in the sheath and is caused by the processing of the plasma and field on crossing the bow shock and passing through the magnetosheath to the point of observation. However, the distribution of correlations with the L1 data for $sin(\theta/2)$ are very similar for THEMIS-B in the magnetosheath and THEMIS-B in the undisturbed solar wind. Theoretically, the clock angle in the undisturbed solar wind would be conserved into the magnetosheath, for some simplified situations but this is not generally the case and, in practice, although clock angle θ is conserved to some degree, changes are introduced and can be quite large (*Coleman*, 2005; *Crooker et al.*, 1985; *Walsh et al.*, 2019; *Zhang et al.*, 2019). Thus, it is somewhat surprising that the distributions of r for $sin(\theta/2)$ for THEMIS inside and outside the bow shock are so similar. Even more surprising, given the changes to the correlations for B, N_{SW} and V_{SW} caused by the processing in passing through the bow shock and magnetosheath, is that the coupling function P_α shows a similar distribution of correlations for THEMIS-B inside and outside the bow shock. This must reflect the dominant role in P_α (and all coupling functions) of the clock angle and the IMF orientation term.

Part (e) of Figure 5 gives an insight into where agreement between the coupling function P_α and the *SML* index is likely to be lost for this averaging timescale of 10 min. The mean correlation with the L1 value of P_α falls from unity at L1 to 0.71 in the near-Earth undisturbed solar wind which falls further to 0.66 in the magnetosheath and to 0.53 with *SML* The distributions of r reflect this fall in mean correlation and the mode values falling from unity to 0.86 in the near-Earth solar wind to 0.66 after crossing the bow shock and to 0.59 in the auroral electrojet. The end-to-end correlation caused by this interaction chain can only be as good as that of the weakest link and the similarity of the distributions of r with *SML* for L1, near-Earth and sheath observations (discussed in the previous section) implies that although some correlation is clearly lost because of spatial structure in interplanetary space and by the processing by traversal of the bow shock and sheath, the biggest uncertainty remains in accounting for the driving mechanisms of the auroral currents by the magnetosheath flow.

Figure 6 is the same as Figure 5 for hourly means. The most obvious difference is that correlations are all greatly enhanced. The same analysis of the loss of agreement can be applied. The mean correlation falls to 0.84 in near-Earth undisturbed solar wind which falls further to 0.75 after crossing the bow shock and to 0.63 in the auroral electrojet. The mode values of the distributions of r fall to 0.94 in the near-Earth solar wind to 0.82 after crossing the bow shock and to 0.75 in the auroral electrojet.

7 A quick survey of coupling function performance as a function of timescale

Given the extremely large number coupling functions proposed since epsilon was published in 1978, it is not possible to survey the performance of all the proposed formulations. Nor would it be a useful comparison. *Lockwood and McWilliams* (2022) make the point that a coupling function's performance depends upon both the averaging timescale τ used and which terrestrial space-weather indicator it was designed to predict. In this section, we review the performance of a small basket of proposed coupling functions against the terrestrial indices discussed in section 2 but, importantly, as a function of timescale.

Figure 7 is a "postage stamp" presentation of lag correlograms for averaging timescales τ which increase from 1 minute to 1 year from left to right with different rows being for different parameters. The top 6 rows (a-f) are interplanetary parameters, the next 5 rows (g-k) are for different coupling function combinations of interplanetary parameters and the bottom 4 rows (l-o) are for the four selected terrestrial disturbance indexes. In each row the lag correlograms are given for the parameter in question with: (mauve line) $-SML$; (green line) $-SMR^*$; (blue line) Φ_{PC} (available for $\tau \geq 1$ hr. only); and (orange line) *am* (available for $\tau \geq 3$ hrs. only). In all case, a positive lag corresponds to the parameter, defined by the row number, being lagged. The lags covered are chosen for teach τ to be large enough to define the width of the main peak but small enough to make any lags between the peaks detectable. In all cases the correlations are for all valid data taken between 1995 and 2019 (inclusive).

Before discussing the correlograms between interplanetary parameters and the terrestrial activity indices, we first discuss the relationships between those terrestrial indices, analyzed in the bottom 4 rows of Figure 7. In these rows, one of the four variations plotted is therefore an a.c.f. rather than a cross-correlogram. Row (l) is for the *am* index (for $\tau \geq$ 3hrs). It can be seen that $-SML$ is highly correlated with *am* which is known because both these indices are dominated by the auroral electrojet of the substorm current wedge (see supplementary information to *Lockwood et al*., 2019a). The $-SMR^*$ index shows a more persistent but delayed response consistent with the longer integration times of solar wind forcing that correlate best with enhanced ring current (*Lockwood et al*., 2016). The transpolar voltage Φ_{PC} correlates less well with *am* than the geomagnetic indices and tends to lead *am* at peak correlation. This can be seen most clearly in row (m) which is for Φ_{PC} (for $\tau \geq$ 3hrs) and which shows the responses of all the geomagnetic indices are slightly after the peak in Φ_{PC}. This is expected because of the duration of substorm growth phases in which Φ_{PC} is enhanced by reconnection in the dayside magnetopause but the geomagnetic indices not yet strongly enhanced. But note also that there is a second component because Φ_{PC} is also enhanced by reconnection in the cross-tail current sheet in the subsequent substorm expansion phases (*Lockwood and McWilliams* , 2021a) as predicted by the expanding-contracting polar cap (ECPC) model of the excitation of ionospheric convection (*Cowley and Lockwood*, 1992). This gives an in-phase element to the relationship between Φ_{PC} and the $-SML$ and *am* indices. Rows (n) and (o) confirm the delayed response of $-SMR$ with respect to the other indices discussed above. For $\tau = 27$ days the correlograms cover −2.5 BR to +2.5 BR, where BR is a Bartels solar rotation period (27 days) and the harmonic peaks seen in the a.c.f.s of interplanetary parameters in Figure 1 are a factor and make the main peak less pronounced. For the annual timescales shown in the right-hand panels of rows the same behavior is seen in all 4 of these rows which is associated with the nature of the solar cycle, as discussed below.

The top row (a) of Figure 7 is for the magnitude of the IMF, *B*. Moving from left to right we see that peak correlations with $-SML$ increase from a modest 0.5 for $\tau = 1$ min. to 0.95 for $\tau = 1$ year. Peak correlations with $-SMR^*$ for $\tau < 1$ day exceed those for $-SML$ but lag behind them reflecting the time for the ring current to be enhanced. All parameters show the rise in peak correlation with τ. This rise has four main causes. Firstly, random noise (such as measurement errors) are increasingly averaged out within the averaging period. Secondly, systematic noise (such as seasonal and dipole tilt effects) are averaged out, but only completely for $\tau = 1$ year. Thirdly the ranges of variability in both *B* and the terrestrial indices are reduced by the averaging. And lastly, factors which have an influence at low τ tend towards constant values under the central limit theorem (*Fischer*, 2010): as will be shown in the next section, the most important example of this is the IMF orientation factor. The lag correlograms for *am* and $-SML$ are somewhat asymmetric, with higher correlation tending to linger after the peak, slightly more than the persistence seen in the rise up to the peak. This asymmetry is considerably more pronounced for $-SMR^*$. This was also noted in the correlation study for $\tau = 5$ min. by *Maggiolo et al.* (2017). For τ up to about 1 day the correlograms are similar for all four terrestrial indices ($-SML$, $-SMR^*$, Φ_{PC} and *am*). However, the asymmetry is seen even in the annual means for the geomagnetic indices $-SML$, $-SMR^*$ and *am* but not in the transpolar voltage Φ_{PC}. This difference between the behavior is interesting as it implies there is some magnetospheric “memory” (i.e., preconditioning by prior years) in action for geomagnetic indices that is not present in Φ_{PC}. This could indicate that there is a small solar cycle variation in the difference between dayside and nightside reconnection rates and hence in open solar flux which means prior years. This also will be discussed below in relation to the coupling functions.

The second row in Figure 7 is for the solar wind mass density, ρ_{sw}. Peak correlations are much lower in this case than for IMF *B* and the lag correlograms are highly asymmetric, with small, zero or even negative correlations before the peak and larger ones after it. The peaks for ρ_{sw} lag behind the peaks for *B*. There is no significant effect of ρ_{sw} on annual timescales. The results of *Lockwood and McWilliams* (2021a) show that transpolar voltage is increased slightly by enhanced solar wind dynamic pressure (and hence ρ_{sw}) and *Lockwood et al.* (2019b) show that geomagnetic activity is also enhanced by enhanced solar wind dynamic pressure. Modelling shows that this second effect is associated with the squeezing of the tail and its effect in increasing the energy stored in the tail and the total current flowing in the cross-tail sheet *Lockwood et al.* (2019c), a factor which increases the amplitude of the "equinoctial" (a.k.a., "McIntosh") pattern in geomagnetic activity that is associated with the dipole tilt (*Lockwood et al.*, 2019c). This is consistent with the lack of an effect of ρ_{sw} in annual means, for which dipole tilt effects are averaged out. The enhanced correlations are seen for timescales up to about a day, which is what we expect for the squeezing effect in the tail and the persistence of the solar wind dynamic pressure.

The third row of Figure 7 is for the solar wind speed, V_{sw}. As for *B*, peak correlations grow with τ. Correlograms are again asymmetric, but in the opposite sense to ρ_{sw}, with larger values at negative lags that fall sharply for positive lags. These negative lags (meaning that correlation is seen with solar wind speed that was observed after the terrestrial activity) does not question causality and was also noted in the correlation study for τ = 5 min. by *Maggiolo et al.* (2017), who correctly interpret it as being due to the geoeffectiveness of enhanced solar wind density and field (and the potential for enhanced southward out-of-ecliptic IMF) in CIR and CME fronts ahead of fast solar wind, caused by the enhanced solar wind interacting with slow solar wind ahead of it. That a similar asymmetry is present for annual means in the far right can be understood because solar wind effects are greater in the declining phase of the solar cycle when low-latitude extensions to coronal holes form giving more fast streams: hence the declining phase activity will correlate better with annual means for the prior year than for the next year.

The next three rows, (d), (e) and (f), look at the variation of three IMF orientation factors of the form $sin^d(\theta/2)$, being for d = 2, d = 4 and d = 6 The behavior is very similar in all three cases, with peak correlations growing with τ up to about 1 day, after which they decline again. Note that no significant correlations (at the 2-σ level) were obtained at τ = 1 year because at this timescale $sin^d(\theta/2)$ (for all d) is essentially constant, as will be discussed later. This is also the reason why peak correlations for

$sin^d(\theta/2)$ decline for $\tau > 1$ day and one of the most important reasons why correlations with B and V_{sw} frow with increased τ. *Lockwood and McWilliams* (2021) review proposed values for d and note that they range from 1 to 9, but most lie in the range between 2 and 6 studied here.

The next five rows are for examples of coupling functions that combine the factors studied in the previous rows and have been chosen to cover a range of d from 2 to 6. Row (g) is for the *Vasyluinas et al.* (1982) power input into the magnetosphere estimate P_α for $\alpha = 1/3$ and $d = 4$; row (h) is for the *Boyle et al.* (1997) transpolar voltage prediction, C_{BEA} (which uses an additive term with $d = 3$); row (i) is for the empirical "Nearly Universal" coupling function of *Newell et al.* (2007), C_U (for which $d = 2.67$); row (j) is for the theory-based coupling function of *Borovsky and Birn* (2014) (for which $d = 2$); and row (k) is for the empirical coupling function of *Temerin and Li* (2006), C_{TL} (for which $d = 6$). Although there are some small differences, Figure 7 makes the point that they are actually rather similar in their performance is we used correlation as a metric. Correlations rise with τ from typically 0.65 for $\tau = 1$ min to up to 0.99 for $\tau = 1$ year. Incidentally, this is not to say that the form of the coupling function does not matter; for example, the best performance of a coupling function found in this survey was that derived by *McPherron et al*. (2015) for predicting the $-SML$ index, giving a correlation of 0.688 at $\tau = 1$ min, rising to 0.973 for $\tau = 1$ year: this coupling function (with $a = 0.70\pm0.01$, $b = 0.096\pm0.009$, $c = 1.92\pm0.04$ and $d = 3.67\pm0.04$) was derived empirically using best practice and the *AL* index (which is very similar to *SML*) for $\tau = 1$ hour. The worst-performing was the epsilon factor, ε, which yielded correlations that varied from 0.47 to 0.62 for the same range of τ.

The correlations for coupling functions are dominated by the effect of the IMF orientation factors up to about $\tau = 1$ day (removing the other factors was found to cause only relatively minor loss of correlation) but for $\tau = 1$ year the IMF orientation factor makes no difference and no loss of correlation at all is incurred if it is omitted at this timescale. Rather than looking at the relatively small differences between the various coupling functions, we here look at their common behavior. A feature seen for all the better coupling functions is that after the peak, the correlation falls faster for Φ_{PC} than for all the three geomagnetic indices. This is consistent with the effect of some open flux, transported into the tail by enhanced Φ_{PC} remaining there and continuing to drive and enhanced level of geomagnetic activity until the enhanced open flux has decayed away (*Lockwood and McWilliams*, 2021). The annual correlations are also asymmetric, with transpolar voltage weakly showing the sort of solar cycle effect noted for solar wind speed, whereas the geomagnetic parameters weakly showing the long-term memory effect seen for the IMF. As noted above, this is also seen in the correlograms for τ = 1 year between the various terrestrial indices in rows (l)-(o).

8 Occurrence distributions of coupling functions and the data that they predict

An aspect of coupling functions that has not attracted much attention in the past is what they predict for the occurrence distribution of a given terrestrial parameter. Much of this has because interest has focused on large events, in other words on the large-event tail of the distribution and not on its "core" around the mode value. The reason for this is that much of space weather science is concerned with the major disturbance events. However, there is another aspect the space weather associated with the integrated effects of activity. Examples include: integrated lifetime radiation doses for spacecraft electronics and for astronauts; and integrated GIC induced current effects on power grid transformers and on pipeline corrosion. However, it is not just these "lifetime dose" issues that mean we should also consider the full distributions, they are very likely to be important for understanding preconditioning effects in the magnetosphere, in which the response of the magnetosphere in a large event also depends on the accumulated activity level of prior intervals. Preconditioning is discussed further in Section 13. If we are to learn how to allow for and predict preconditioning effects, we need to look at the whole distribution of the coupling function and how well it matched that of the parameter it is attempting to predict. For these reasons a full space weather climatology should look at the full distributions of parameters, and not just the large event tails. These distributions depend critically on averaging timescale (*Lockwood et al*., 2019a; b; c).

Figure 8 presents a postage stamp plot of occurrence distributions for the same parameters, averaging timescales and data interval as Figure 7. The grey histograms give the number of samples n, normalized by its peak value $n/n_{max,}$, in bins that are 0.01wide. These are histograms of the normalized parameter value $x/<x>$, where $<x>$ is the mean over all available samples. The vertical mauve lines are at the distribution mean, $x/<x> = 1$. The n values are normalized by n_{max} rather than Σn because the latter can requires some very large y-axis scales when the distribution tends to a delta function.

The distributions for the IMF, B and solar wind mass density ρ_{sw} at $\tau < 1$ day are close to lognormal and do not change much in form with τ. However, for larger τ, the central limit theorem begins to have a large effect and distributions narrow and become more Gaussian as they evolve towards the delta function that would be obtained for τ equal to the whole 25-year dataset. The distribution for solar wind speed V_{sw} is different because, unlike B and ρ_{sw}, V_{sw} never falls anywhere close to zero and has a baselevel value of around 350 km s^{-1}. This makes the distribution of V_{sw} /< V_{sw} > narrower: it too narrows under the central limit theorem as τ is increased.

Lockwood et al. (2019b) and *Lockwood and McWilliams* (2022) explain the details of how the strange distributions of $sin(\theta /2)$, and hence of $sin^d(\theta /2)$ shown in Figure 8, arise for low τ. However, it should be noted that such a distribution will arise for most IMF orientation factors that allow for the "half-wave rectifier" aspect of IMF orientation control of coupling between the solar wind and the magnetosphere. For example, because the distribution of the southward component of the IMF in GSM coordinates, B_Z, is symmetric about zero, use of a half-wave rectified southward ($B_S = -B_Z$ for $B_Z < 0$ and $B_S = 0$ for $B_Z \geq 0$), yields a distribution in which 50% of the samples are in a delta function at $B_S = 0$. Because of the large variability in θ these distributions of $sin^d(\theta /2)$ evolve quickly with increased τ, in general from the strange distributions at high time resolution to a lognormal, and then to a gaussian that then thins to a delta function. However, the evolution depends strongly on the value of d: for example, for $d = 2$ the distributions remains symmetric at all τ and the lognormal phase is not seen. In general, the larger the value of d, the greater the delta function spike at zero and the larger is the τ value needed to remove it and obtain a lognormal form. Hence the value of d used has great effects on the distribution of the coupling function . Note that for τ = 1 yr., the larger variability of θ means that the distributions of $sin^d(\theta/2)$ are reduced to delta functions for all. This is the reason why highly successful coupling functions on annual timescales, giving correlations of about 0.98, do not contain an IMF orientation terms. It is also why no significant correlations could be found for τ = 1 yr. in rows (d), (e) and (f) of Figure 7.

The coupling function distributions in rows (g)-(k) show that for τ ≤ 1 day they are highly dependent of the form of the IMF orientation term used and, in particular, the value of d. The strange shape to the distributions only evolves at larger τ into the lognormal and quasi-Gaussian distributions seen for the terrestrial indices. For the *Borovsky and Birn* (2014) coupling function with $d = 2$, this is achieved by τ = 3 hrs., whereas for the *Temerin and Li* (2006) coupling function with $d = 6$ this is only achieved with τ ≥ 1 day. At timescales of an hour and less the $sin^d(\theta/2)$ formulations do not work well in terms of matching the core and low-activity end of the distribution, but the additive formulation of the *Boyle et al.* (1997) coupling function C_{BEA} means that it provides a much better match at low τ. However, close inspection shows that the *Boyle et al.* (1997) formulation has a light high-activity tail compared to the terrestrial geomagnetic indices, whereas $sin^d(\theta/2)$ formulations (sometimes with large d) can fit the high activity tail rather better. Thus, none of the formulations available at the present time are suitable for quantifying both the core of the distribution and the large-event tail. We note that the *Boyle et al.* (1997) formula was designed to predict transpolar voltages, the distribution for which do not show as heavy large-event tail as the geomagnetic indices, particularly at larger τ.

9 Correlation coefficients as a metric of performance

Previous sections have used linear correlation coefficients as a metric to assess the performance of various coupling functions. This is indeed the metric that has been used in almost all coupling function studies. In this section, we look at the implications of the adoption of this metric and ask if it is always appropriate. *Vasyliunas et al* (1982) make the important point that correlation coefficients do not guarantee linearity of the coupling function over the range of activity level you are most interested in. For example, if you are interested in a very large and extreme event tail, correlation coefficient could be set by the large number of samples in the core of the distribution and that might not be the best fit to the small number of tail samples that you are interested in. More subtly, *Lockwood and McWilliams* (2022) demonstrate that even in the core of the distribution, peak linear correlation coefficient between samples does not necessarily guarantee you linearity and linearity between a coupling function and the terrestrial indicator that it aims to predict is what one requires.

Figures 9, 10 and 11 of *Lockwood and McWilliams* (2022) present their implementation of the test for linearity suggested by *Vasyliunas et al* (1982) (for Φ_{PC}, *am* and *SML*, respectively). This ensures that $F(\theta)$ is of the correct form for the proposed G to give a linear coupling function. Note that the polynomial fit used by *Lockwood and McWilliams* (2022) could be weighted to ensure that the linearity is over the range of the terrestrial index that is of greatest interest, although in the implementation by *Lockwood and McWilliams* (2022) equal weighting was given to equal width averaging bins that covered the whole range of $F(\theta)$. The full procedure for the derivation of the optimum value of d and its 1σ, 2σ, and 3σ uncertainties, is described in the paper by *Lockwood and McWilliams* (2022). At each d, the Nelder-Mead simplex search yields the values of exponents a, b, and c and hence their optimum values (and their uncertainties) are also defined.

The correlation obtained for Φ_{PC}, for example, is r = 0.858, an extremely high value for τ = 1 hr. which means that r^2 = 73.6% of the variance in hourly transpolar voltages is explained. It is the fit given by the linearity condition. However, it is not the fit that gives the highest possible value of r:

that was obtained for d = 2.20 and was r = 0.8646 (explaining 74.8% of the variance). What is happening here can be seen in Figure 8 of *Lockwood and McWilliams* (2022): Part a of that figure is for d = 1.1 (too low); Part b is for d = 2.2 (and yields the highest correlation); Part c is for d = 6.5 (that is too high); and Part d is for d = 2.5 (that yields linearity). These plots show that maximizing r also minimizes the rms deviation of observation and coupling function fit, Δ_{rms} and so minimum Δ_{rms} does not give linearity either.

There is an important point to note about Figure 8c of *Lockwood and McWilliams* (2022) in which the d value is too large, in relation to identifying and quantifying transpolar voltage "saturation" effects (e.g., *Hairston et al.*, 2005; *Shepherd*, 2007). These effects are when the transpolar voltage does not increase as much with solar wind forcing at higher forcing levels than at lower ones and may even lead to a leveling off such that there is a maximum voltage that can be achieved. Global MHD simulations (for example, *Kubota*, 2017) can reproduce such effects and so do indicate it is a real phenomenon. There is no reason to doubt that saturation does occur to some extent in Φ_{PC}, but the nature of the mechanism(s) is still a matter of debate and it is not at all clear that this saturation effect in Φ_{PC} causes saturation to anything like the same degree in geomagnetic indices, if at all (*Borovsky*, 2021b). Some geomagnetic indices do not show any saturation effect, others do but to varying degrees. In general, any geomagnetic index saturation is less pronounced than that in Φ_{PC}, a fact that is reflected in the distributions shown in Figure 8 in that the large event tail is fatter and longer for geomagnetic indices than for Φ_{PC}. This urges caution when quantifying a saturation effect because it demonstrates that similar behavior can be brought about by an inadequate coupling function. A similar point was recently made by *Borovsky* (2021b).

10 Fitting the bulk of the distribution and the large event tail

Figure 9 returns to the point about fitting the core of distribution, versus fitting the large-event tail for the fits to Φ_{PC} described above. The gray area bounded by the black line in Figure 9a is the distribution of the 65133 valid hourly transpolar voltage values in the survey of 25 years of SuperDARN by *Lockwood and McWilliams* (2021). The yellow and cyan lines break this distribution into two subsets, for IMF $B_Z \geq 0$ (in GSM) in cyan and $B_Z < 0$ in yellow. The two sub-set distributions cross at $\Phi_{PC}/<\Phi_{PC}> = 0.85$, below which the distribution is increasingly dominated by northward IMF and above which is increasingly dominated by southward IMF: at Φ_{PC} below about

$0.3<\Phi_{PC}>$ the distribution is almost exclusively due to northward IMF whereas for above about $2<\Phi_{PC}>$ it is almost exclusively due to southward IMF.

Figure 9b repeats the observed distribution (as a black line) and compares it to the mauve line, which is the distribution for the optimum (linear) fit C_f (given by equation 2 with d = 2.50) discussed in the previous section. Also shown for comparison, the blue line is C_f for the optimum fit to the simultaneous $-SML$ data (given by Equation 2 with d = 5.20) and the green line is the predictions of the *Boyle et al* (1997) formula (Equation 7), C_{BEA}. It can be seen that C_{BEA} matches the distribution exceptionally well, whereas C_f with d = 2.50 only matches well for Φ_{PC} above about $2.5<\Phi_{PC}>$ and is very poor in the northward-IMF dominated part of the distribution below $0.85<\Phi_{PC}>$. This problem increases with d and is very severe for C_f with d = 5.20).

To understand the implications of these fits, the lower panels of Figure 9 are the same data as in the upper panels, but shown as cumulative distributions functions (c.d.f.s) rather than probability distribution functions (p.d.f.s). Figure 9c makes the point that about a third of the total flux transport across the polar cap takes place when the IMF is northward and about 2/3 when it is southward. This is a prediction of the ECPC convection model, as discussed by *Lockwood and McWilliams* (2021). Figure 9d shows the predictions for integrated flux transport by the various coupling function fits. The observed distribution shows that 36% of the total flux transport is at below-average values of Φ_{PC} and, not surprisingly given Figure 9b, this is very well matched by C_{BEA} which predicts 38%, but not so well matched by C_f which gives 28% with d = 2.50 and just 22% with d = 5.20. Hence using C_f with even the optimum d underestimates the total flux transport in quiet times, but because in Figure 9f the mauve line reaches close to unity at very large activity levels, it correspondingly overestimates the flux transport at high activity levels.

Figure 10 is the same as Figure 9 but for the observed distribution of $-SML$ values for τ = 1 hr. The p.d.f.s and c.d.f.s are shown for in the right hand panels for C_f with 4 different d values, including (in green) the optimum linear fit value of d = 5.20, derived using the same procedure as was used for Φ_{PC} in Section 10. The other lines are for (in mauve) d = 2.0 (too low); (in blue) d = 3.82 (gives peak correlation, r); and (in orange) d = 7.5 (too high). As noted earlier, the observed distribution for $-SML$ is much more "heavy tailed" than for Φ_{PC} and this makes the poor fit for the northward-IMF-dominated part of the curve less pronounced although still present. The higher d of the fit means the behavior for $-SML$ is much more like a "half-wave rectified" behavior than it is for Φ_{PC}. In terms of the integrated value, the problem of missing the contribution of very quiet times is not as marked as it is for Φ_{PC} , but it is still present and is indeed inherent in the $sin^d(\theta/2)$ formulation. Figures 10b and 10d show that the higher *d* value or 5.2, on the other hand, can match the large-event tail of the distribution well.

Figure 11 looks at how well individual cases, as well the *SML* distribution, are matched by these coupling functions. The four panels are for the same four values of *d* that were used in Figure 10. Each shows a data density plot of normalized *SML* against normalized C_f, overlaid with a q-q plot, as used in Figure 4. The colour scale is chosen so that a single hourly average in a counting bin (pixel) of size 0.03×0.03 will show up in blue and so, at the extremes, the data density plot takes on the information of a scatter plot for individual samples. It can be seen that although the fit is good at low, average and moderately high values, the scatter is increasingly large at high and extreme values. The q-q plots in Figure 11c show how the optimum linear fit, despite giving a lower correlation than in Figure 11b, is matching the observed distribution very well, the agreement being almost perfect up to the 99 percentile and the predicted tail is just marginally heavier than that for the observations for the largest percent. For the peak correlation, shown in Figure 11b, the deviation from the ideal is also small but the predicted tail is detectably thin for $SML/\langle SML \rangle$ above about 2.5. Again, peak correlation is not giving the best match to the very large events, but we note that the scatter is high and so the accuracy of individual hourly predictions is not high, despite the overall correlation being high.

Note that all the q-q plots in Figure 11 shows a deviation from linearity at low values. This is a problem inherent with the $sin^d(\theta/2)$ IMF orientation factor that becomes more pronounced as *d* is enhanced. This means using this formulation with a large *d* to fit the large event tail causes problems for fitting the core and small value end of the distribution. To avoid this, we could use different coupling functions to model the core and the large event tail of the distribution, but a better solution would be to derive a better IMF orientation factor that can accommodate both.

11 Variations over parameter space

Thus far, we have been looking at coupling functions derived and evaluated over all usable observations taken over a 25-year interval. But this does not tell us how a coupling function performs in different parts of parameter space. To illustrate this point, Figure 12 looks at the performance of two different coupling functions, both designed to predict Φ_{PC}, over IMF (*B*) and solar wind speed

(V_{sw}) parameter space. Data are divided into 40 inter-quantile ranges of both V_{sw} and B. Figure 12a gives the fraction of available samples in each V_{sw} -B bin (on a logarithmic scale). Note the bin widths change because they are defined by the quantiles for that parameter. Figure 12b givens the mean value of Φ_{PC} in the same bins. The other two panels of Figure 12 show the fraction of the variance r^2 of the transpolar voltage Φ_{PC} that is explained by coupling functions (c) C_{BEA} and (d) P_α for d = 2.5, where r is the correlation coefficient. The peak of the lag correlograms r is found and r^2 plotted as a function of V_{sw} (along the x-axis) and B (along the y axis). Figure 12a shows that samples are rarest for high V_{sw} and low B and low V_{sw} and high B, whereas they are most common is low B and low V_{sw}. When taking correlations between the coupling functions only those with significance exceeding the 2-σ level are retained and those failing that test are in the parameter space where sample numbers are low.

Figures 12c and 12d are for the *Boyle et al.* (1997) and *Vasyluinas et al.* (1982) coupling functions, C_{BEA} and P_α respectively, where P_α is for α = 1/3 and d = 2.5. The overall correlations are r are 0.733 and 0.784 respectively (r^2 of 54% and 61%). The variation over parameter space is very similar for the two coupling functions and also considerable, with r^2 varying between 0.25 and 0.75. There is a marked trend with highest correlations for low V_{sw} and high B and lowest correlations for high V_{sw} and low B. It is clear that a high correlation can hide a considerable variation in performance over parameter space.

12 Preconditioning

The fact that observed correlations are as high as they are (although the level varies considerably with averaging timescale, τ) places limits on the importance of factors that have been omitted, which thus far we have largely regarded as a source of noise. This section looks at preconditioning by the pre-existing state of the magnetosphere-ionosphere-thermosphere system or other variations that can alter the response to a given set of interplanetary conditions, which is (in the author's view) the most interesting and the most challenging scientifically.

There are two ways in which preconditioning can come about. The first concerns the orbital and characteristics of Earth which cause an annual variation in the Earth-Sun distance, and seasonal and Universal Time (UT) effects associated with the dipole tilt (see review by *Lockwood et al.*, 2020a). There have been attempts to allow for dipole tilt effects in coupling functions (*Svalgaard*, 1977; *Murayama et al*., 1980, *Li et al*., 2007, *Luo et al.*, 2013) with terms that allow for the fraction of the calendar year, F, and UT. In addition, such effects have been included in the filters used in the linear prediction filter technique (*McPherron et al*., 2013). However, the *F-UT* effects are not independent of solar variations. For example, ionospheric conductivity effects will also depend on the flux of EUV and X-ray ionizing radiations (for which F10.7 or sunspot number are often used as proxy indices) and *Lockwood et al.* (2020b; c; d) have shown that the amplitude of the *F-UT* pattern in geomagnetic activity (the "equinoctial", a.k.a. "McIntosh" pattern) is linearly proportional to the solar wind dynamic pressure. There are a number of theories as to how this dipole tilt arises (see *Lockwood et al.,* 2020a) and each has implications for how an *F-UT* dependence should be introduced. Note also that seasonal effects mean that this allowance will be different for truly global indices such as *am* and for northern-hemisphere-only indices such as *AL* and *SML*. Hence although some allowance for these effects could be achieved used by using *F* and *UT* with average values for solar-terrestrial variables, full allowance is likely to require the inclusion of more free fit parameters which increases the potential for, and probability of, overfitting.

The second form of pre-conditioning relates to the pre-existing activity level of the magnetosphere-ionosphere-thermosphere system and hence on the prior history of the solar wind driving it. There are a number of proposed mechanisms. The storage-release system that yields the substorm cycle shows that the response of the magnetosphere depends on the pre-existing flux of open magnetospheric field. A method to allow for this using an extremely large number coupled equations was proposed by *Luo et al* (2013). Another way of dealing with this non-linearity is by using neural networks (e.g., *Gleisner and Lundstedt*, 1999). One more widely-used technique to allow for the non-linearity of response caused by this type of preconditioning is the local linear prediction filter technique (*Vassiliadis et al*., 1995; *Vassiliadis, 2006)*, in which moving average filters are continually calculated as the system evolves and these are used to compute the output of the system. The filter used is derived or selected according to the state of the system.

The design of the best filter to use, or the best set of coupled equations would, in general, depend on the physical preconditioning mechanism(s) that are active and many have been proposed. These are numerous. They include: mass loading of the near-Earth tail with ionospheric O^+ ions from the cleft ion fountain (*Yu and Ridley*, 2013); the formation of thin tail current sheets (*Pulkkinen and Wiltberge*r, 2000); the development of a cold dense plasma sheet (*Lavraud et al*., 2006); and mass loading of the dayside magnetopause reconnection region (*Walsh and Zou*, 2021).

The best way to include the effects of the above mechanisms into coupling functions is far from clear, although system science studies could potentially provide answers. However, some other proposed preconditioning effects may be easy to include because they involve other terrestrial indices that can be predicted using a purpose-designed coupling function. An example would be the proposed effect on the reconnection rate in the cross-tail current sheet of enhanced ring current (*Milan et al*. 2008; 2009; *Milan* 2009) for which predictions of the *Dst*, *SYM-H* or *SMR* indices, based on the prior history of a relevant coupling function, could be used to modify the predicted response in another index (for example Φ_{PC} or *SML*) for a given values of its optimum coupling function. The magnetosphere sometimes responds to continued solar wind forcing (over a period of tens of minutes) by generating a substorm, or a string of substorms and sometimes with a steady convection event (e.g., *Kissinger et al*, 2012; *Lockwood et al.*, 2009; *Milan et al.*, 2021). It is known that the response of the auroral electrojet indices depends on the current *Dst* value (*Gleisner and Lundstedt*, 1999; *O'Brien et al.*, 2002; *Juusola et al*., 2013). This evidence points to using a preconditioning factor based on *Dst*, or other ring current index, may be viable. This raises an interesting point about timescales, as *Lockwood et al* (2016) have shown that *Dst* correlates best with the integrated solar wind forcing over a prolonged (~12 hr.) prior period. Hence the precondition term may well require a different averaging timescale than the main coupling function.

13 Concluding remarks

This paper has taken a general and detailed look at solar wind-magnetosphere coupling functions. These have been used for almost 50 years now, but an in-depth review is now timely because systems analysis techniques are increasingly being applied to the magnetosphere (see review by *Borovsky and Valdivia*, 2018). For example, *Borovsky and Osmane* (2019) introduced methodology using a state-vector-reduction technique and canonical correlation analysis which treats the magnetosphere as an example of a multivariable system driven by multiple inputs that identifies independent modes of reaction of the magnetospheric system to its drivers. Techniques such as these are likely to offer solutions to many of the limitations of traditional coupling function-terrestrial observation correlation analysis, particularly in the limitations of preconditioning and the effects of the pre-existing state of the magnetosphere. In addition, application of machine learning techniques should avoid common problems such as overfitting (e.g., *Camporeale*, 2019; *Baumann and McCloskey*, 2021). However, other limitations and sources of noise may be unwittingly carried forward into these techniques. Hence it is timely to step back review them.

Testing the predictive and analysis uses of coupling functions also raises another set of complications, with a variety of performance metrics available for consideration (*Liemohn et al*., 2018). The most appropriate one (or ones) for the application in question should be deployed, especially in the context of forecasting (*Owens*, 2018). The derivation and testing of coupling functions has, in the past, been almost entirely based on correlation analysis and it clearly has an important role into the future, but this paper has highlighted that it is not always the most appropriate metric to be using, and metrics more appropriate to the specific application are likely to be needed.

Acknowledgements. The work presented in this paper was supported by STFC consolidated grant number ST/M000885/1 and by the SWIGS NERC Directed Highlight Topic Grant number NE/P016928/1/. The author acknowledges the use of data from the SuperDARN project. SuperDARN is a collection of radars funded by national scientific funding agencies of Australia, Canada, China, France, Italy, Japan, Norway, South Africa, United Kingdom and the United States of America. In addition he is grateful to the staff of the Space Physics Data Facility, NASA/Goddard Space Flight Center, who prepared and made available the OMNI2 dataset used: these interplanetary data were downloaded from http://omniweb.gsfc.nasa.gov/ow.html; and of the World Data Center for Geomagnetism, Kyoto who generate and make available the AL index from http://wdc.kugi.kyoto-u.ac.jp/aeasy/index.html and of L'École et Observatoire des Sciences de la Terre (EOST), a joint of the University of Strasbourg and the French National Center for Scientific Research (CNRS) and the International Service of Geomagnetic Indices (ISGI) for making the am index data available from http://isgi.unistra.fr/data_download.php. He is also grateful to the many groups who built and operated the instruments that have monitored near-Earth interplanetary space, particularly on the spacecraft ACE, Wind and THEMIS-B, and to the SuperMAG project for the SML index and acknowledges the following projects and PIs: Intermagnet; USGS, Jeffrey J. Love; CARISMA, PI Ian Mann; CANMOS; The S-RAMP Database, PI K. Yumoto and K. Shiokawa; The SPIDR database; AARI, PI Oleg Troshichev; The MACCS program, PI M. Engebretson, Geomagnetism Unit of the Geological Survey of Canada; GIMA; MEASURE, UCLA IGPP and Florida Institute of Technology; SAMBA, PI Eftyhia Zesta; Chain, PI K. Yumoto; SAMNET, PI Farideh Honary; The institutes who maintain the IMAGE magnetometer array, PI Eija Tanskanen; PENGUIN; AUTUMN, PI Martin Connors; DTU Space, PI Rico Behlke; South Pole and McMurdo Magnetometer, PI’s Louis J. Lanzarotti and Alan T. Weatherwax; ICESTAR; RAPIDMAG; PENGUIn; British Antarctic Survey; McMac, PI Peter Chi; BGS, PI Susan Macmillan; Pushkov Institute of Terrestrial Magnetism, Ionosphere and Radio Wave Propagation (IZMIRAN); GFZ, PI Juergen Matzka; MFGI, PI B. Heilig; IGFPAS, PI J. Reda; University of L’Aquila, PI M. Vellante; BCMT, V. Lesur and A. Chambodut; Data obtained in cooperation with Geoscience Australia, PI Marina Costelloe; and the SuperMAG, PI Jesper W. Gjerloev. The author also thanks Joe Borovsky for insightful conversations about this work and, with Simon Wing, for organizing, the on-line Workshop “Solar Wind - Magnetosphere Interaction” in August/September 2021 at which this work was first presented.

References

Arnoldy, R. (1971), Signature in the interplanetary medium for substorms, *J. Geophys. Res*., **76** (22), 5189–5201, doi:10.1029/JA076i022p05189.

Baumann, C. and A.E. McCloskey (2021) Timing of the solar wind propagation delay between L1 and Earth based on machine learning, *J. Space Weather Space Clim.*, **11** (2021) 41, doi: 10.1051/swsc/2021026

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

Borovsky, J. E. (2013), Physical improvements to the solar wind reconnection control function for the Earth's magnetosphere, *J. Geophys. Res. Space Physics*, **118**, 2113– 2121, doi:10.1002/jgra.50110.

Borovsky J.E. (2021a) Is Our Understanding of Solar-Wind/Magnetosphere Coupling Satisfactory?, *Front. Astron. Space Sci.*, **8**, 634073, doi: 10.3389/fspas.2021.634073

Borovsky J.E. (2021b) On the Saturation (or Not) of Geomagnetic Indices, *Front. Astron. Space Sci.* **8**, 740811, doi: 10.3389/fspas.2021.740811

Borovsky, J. E., and Birn, J. (2014) The solar wind electric field does not control the dayside reconnection rate, *J. Geophys. Res. Space Physic*s, **119**, 751– 760, doi: 10.1002/2013JA019193.

Borovsky, J.E. & J.A. Valdivia (2018) The Earth's Magnetosphere: A Systems Science Overview and Assessment, *Surv Geophys*., **39** (5), 817–859, doi: 10.1007/s10712-018-9487-x

Borovsky, J. E. and Osmane, A. (2019) Compacting the description of a time-dependent multivariable system and its multivariable driver by reducing the state vectors to aggregate scalars: the Earth's solar-wind-driven magnetosphere, *Nonlin. Processes Geophys*., **26**, 429–443. doi: 10.5194/npg-26-429-2019.

Boyle, C. B., Reiff, P. H., and Hairston, M. R. (1997), Empirical polar cap potentials, *J. Geophys. Res.*, **102** (A1), 111– 125, doi:10.1029/96JA01742.

Burton, R. K., McPherron, R., & Russell, C. (1975). An empirical relationship between interplanetary conditions and Dst. *Journal of Geophysical Research*, **80** (31), 4204–4214. https://doi.org/10.1029/JA080i031p04204

Camporeale, E. (2019). The challenge of machine learning in Space Weather: Nowcasting and forecasting. *Space Weather*, **17**, 1166– 1207. doi: 10.1029/2018SW002061

Camporeale, E., S. Wing and J.R. Johnson (eds.) (2018) Machine Learning Techniques for Space Weather, Elsevier, ISBN 978-0-12-811788-0, doi:10.1016/C2016-0-01976-9

Chicco, D. (2017) Ten quick tips for machine learning in computational biology, *BioData Mining*, **10** (1), Article # 35, doi: 10.1186/s13040-017-0155-3

Coleman, I. J. (2005). A multi-spacecraft survey of magnetic field line draping in the dayside magnetosheath. *Annales Geophys.*, 23, 885–900, doi: 10.5194/angeo-23-885-2005

Cowley, S. W. H., and Lockwood, M. (1992), Excitation and decay of solar-wind driven flows in the magnetosphere-ionosphere system, *Annales Geophys*., **10**, 103-115.

Crooker, N. U., Luhmann, J. G., Russell, C. T., Smith, E. J., Spreiter, J. R., & Stahara, S. S. (1985).Magnetic field draping against the dayside magnetopause. *Journal of Geophysical Research*, **90**, 3505–3510, doi: 10.1029/JA090iA04p03505

Crooker, N. U., Siscoe, G. L., Russell, C. T., & Smith, E. J. (1982). Factors controlling degree of correlation between ISEE 1 and ISEE 3 interplanetary magnetic field measurements. *Journal of Geophysical Research*, **87** (A4), 2224–2230. doi: 10.1029/JA087iA04p02224

Davis, T. N., and Sugiura, M. (1966), Auroral electrojet activity index AE and its universal time variations, *J. Geophys. Res*., **71** (3), 785–801, doi:10.1029/JZ071i003p00785.

Hairston, M.R., Drake, K. A., and Skoug, R. (2005), Saturation of the ionospheric polar cap potential during the October–November 2003 superstorms, *J. Geophys. Res*., **110**, A09S26, doi:10.1029/2004JA010864.

Kissinger, J., McPherron, R. L., Hsu, T.-S., and Angelopoulos, V. (2012), Diversion of plasma due to high pressure in the inner magnetosphere during steady magnetospheric convection, *J. Geophys. Res*., **117**, A05206, doi:10.1029/2012JA017579.

Knutti, R., Meehl, G.A., Allen, M.R. & Stainforth, D.A. (2006). Constraining climate sensitivity from the seasonal cycle in surface temperature. *Journal of Climate*, **19** (17), 4224 4233. doi : 10.1175/JCLI3865.1

Knape, J., & de Valpine, P. (2011) Effects of weather and climate on the dynamics of animal population time series. *Proc. Royal Society B: Biological Sciences*, **278** (1708), 985-992, doi: 10.1098/rspb.2010.1333

Kubota, Y., Nagatsuma, T., Den, M., Tanaka, T., and Fujita, S. (2017), Polar cap potential saturation during the Bastille Day storm event using global MHD simulation, *J. Geophys. Res. Space Physics*, **122**, 4398– 4409, doi: 10.1002/2016JA023851

Finch, I.D., and M. Lockwood (2007) Solar wind-magnetosphere coupling functions on timescales of 1 day to 1 year, *Annales Geophys*., **25**, 495-506, doi: 10.5194/angeo-25-495-2007

Fischer H. (2010) A History of the Central Limit Theorem. Sources and Studies in the History of Mathematics and Physical Sciences. Springer, New York, NY. ISBN 978-0-387-87856-0. doi: 10.1007/978-0-387-87857-7

Gleisner, H., and H. Lundstedt (1999), Ring current influence on auroral electrojet predictions, *Annales Geophys*., **17** (10), 1268–1275. doi: 10.1007/s00585-999-1268-x

Juusola, L., N. Partamies, and E. Tanskanen (2013), Effect of the ring current on preconditioning the magnetosphere for steady magnetospheric convection, *Geophys. Res. Lett.*, **40**, 1917–1921, doi:10.1002/grl.50405

Lagarias, J. C., J. A. Reeds, M. H. Wright, and P. E. Wright. (1998) Convergence Properties of the Nelder-Mead Simplex Method in Low Dimensions. *SIAM Journal of Optimization*. **9**, (1) 112–147, doi: 10.1137/s1052623496303470

Lavraud, B., M. F. Thomsen, J. E. Borovsky, M. H. Denton, and T. I. Pulkkinen (2006), Magnetosphere preconditioning under northward IMF: Evidence from the study of coronal mass ejection and corotating interaction region geoeffectiveness, *J. Geophys. Res.*, **111**, A09208, doi:10.1029/2005JA011566.

Liemohn, M. W., McCollough, J. P., Jordanova, V. K., Ngwira, C. M., Morley, S. K., Cid, C., et al. (2018). Model evaluation guidelines for geomagnetic index predictions. *Space Weather*, **16**, 2079–2102. doi: 10.1029/2018SW002067

Li, H., C.Wang, and Z. Peng (2013), Solar wind impacts on growth phase duration and substorm intensity: A statistical approach, *J. Geophys. Res. Space Physics*, **118**, 4270–4278, doi: 10.1002/jgra.50399.

Li, X. L., K. S. Oh, and M. Temerin (2007), Prediction of the AL index using solar wind parameters, *J. Geophys. Res*., **112**, A06224, doi: 10.1029/2006JA011918

Lockwood, M. (2019) Does adding solar wind Poynting flux improve the optimum solar wind-magnetosphere coupling function? *J. Geophys. Res. (Space Physics)*, **124** (7), 5498-5515 doi: 10.1029/2019JA026639

Lockwood, M. & K.A. McWilliams (2021) A survey of 25 years' transpolar voltage data from the SuperDARN radar network and the Expanding-Contracting Polar Cap model, *J. Geophys. Res*., **126**, e2021JA029554 , doi: 10.1029/2021JA029554 (2021)

Lockwood, M. & K.A. McWilliams (2022) On optimum solar wind – magnetosphere coupling functions for transpolar voltage and planetary geomagnetic activity, *J. Geophys. Res*., in press, doi: 10.1029/2021JA029946

Lockwood, M., M.R. Hairston, I.D. Finch, and A.P. Rouillard (2009) Transpolar voltage and polar cap flux during the substorm cycle and steady convection events, *J. Geophys. Res*., **114**, A01210, doi:10.1029/2008JA013697

Lockwood, 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., S. Bentley, M.J. Owens, L.A. Barnard, C.J. Scott, C.E. Watt, and O. Allanson (2019a) The development of a space climatology: 1. Solar-wind magnetosphere coupling as a function of timescale and the effect of data gaps, *Space Weather*, **17**, 133-156. doi: 10.1029/2018SW00185

Lockwood, M., S. Bentley, M.J. Owens, L.A. Barnard, C.J. Scott, C.E. Watt, O. Allanson and M.P. Freeman (2019b) The development of a space climatology: 2. The distribution of power input into the magnetosphere on a 3-hourly timescale , *Space Weather*, **17**, 157-179. doi: 10.1029/2018SW002016

Lockwood, M., S. Bentley, M.J. Owens, L.A. Barnard, C.J. Scott, C.E. Watt, O. Allanson and M.P. Freeman (2019c) The development of a space climatology: 3. The evolution of distributions of space weather parameters with timescale, *Space Weather*, **17**, 180-209. doi: 10.1029/2018SW002017

Lockwood, M., Chambodut , A., Finch, I. D., Barnard, L. A., Owens, M.J. and Haines, C. (2019d) Time-of-day / time-of-year response functions of planetary geomagnetic indices, *J. Space Weather Space Clim.*, **9**, A20, doi: 10.1051/swsc/2019017

Lockwood, M., M.J. Owens, L.A. Barnard, C. Haines, C.JV Scott, K.A. McWilliams, and J.C. Coxon (2020a) Semi-annual, annual and Universal Time variations in the magnetosphere and in geomagnetic activity: 1. Geomagnetic data, *J. Space Weather Space Clim*., **10**, 23, doi: 10.1051/swsc/2020023

Lockwood, M., K.A. McWilliams, M.J. Owens, L.A. Barnard, C.E. Watt, C.J. Scott, A. McNeill and J.C. Coxon (2020b) Semi-annual, annual and Universal Time variations in the magnetosphere and in geomagnetic activity: 2. Response to solar wind power input and relationships with solar wind dynamic pressure and magnetospheric flux transport, *J. Space Weather Space Clim*., **10**, 30, doi: 10.1051/swsc/2020033

Lockwood, M.J. Owens, L.A. Barnard, C.E. Watt, C.J. Scott, J.C. Coxon and K.A. McWilliams (2020c) Semi-annual, annual and Universal Time variations in the magnetosphere and in geomagnetic activity: 3. Modelling, *J. Space Weather and Space Climate*, **10**, 61 doi: 10.1051/swsc/2020062

Lockwood, C.A. Haines, L.A. Barnard, J. Owens, C.J. Scott, A. Chambodut, and K.A. McWilliams (2021) Semi-annual, annual and Universal Time variations in the magnetosphere and in geomagnetic activity: 4. Polar Cap motions and origins of the Universal Time effect, *J. Space Weather and Space Climate*, **11**, 15, doi: 10.1051/swsc/2020077

Luo, B. X., X. L. Li, M. Temerin, and S. Q. Liu (2013), Prediction of the AU, AL, and AE indices using solar wind parameters, *J. Geophys. Res. Space Physics*, **118**, 7683–7694, doi: 10.1002/2013JA019188.

Maclennan, C. G., Lanzerotti, L. J., Akasofu, S.-I., Zaitzev, A. N., Wilkinson, P. J., Wolfe, A., and Popov, V. (1991), Comparison of “Electrojet” Indices from the northern and southern hemispheres, *J. Geophys. R*es., **96** (A1), 267– 274, doi:10.1029/90JA01366.

Maggiolo, R., Hamrin, M., De Keyser, J., Pitkänen, T., Cessateur, G., Gunell, H., & Maes, L. (2017). The delayed time response of geomagnetic activity to the solar wind. Journal of Geophysical Research: Space Physics, 122, 11,109–11,127, doi: 10.1002/2016JA023793

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

McGranaghan, R. M., Bhatt, A., Matsuo, T., Mannucci, A. J., Semeter, J. L., & Datta-Barua, S. (2017). Ushering in a new frontier in geospace through data science. *Journal of Geophysical Research: Space Physics*, **122**, 12,586–12,590. doi: 10.1002/2017JA024835

McPherron, R. L. (1970), Growth phase of magnetospheric substorms, *J. Geophys. Res.*, **75** (28), 5592– 5599, doi:10.1029/JA075i028p05592.

McPherron, R. L., D. N. Baker, T. I. Pulkkinen, T. S. Hsu, J. Kissinger, and X. Chu (2013), Changes in solar wind-magnetosphere coupling with solar cycle, season, and time relative to stream interfaces, *J. Atmos. Sol. Terr. Phys.*, **99**, 1–13, doi:10.1016/j.jastp.2012.09.003

McPherron, R. L., T.-S. Hsu, and X. Chu (2015), An optimum solar wind coupling function for the AL index, *J. Geophys. Res. Space Physics*, **120**, 2494–2515, doi: 10.1002/2014JA020619

Milan, S. E. (2009), Both solar wind-magnetosphere coupling and ring current intensity control of the size of the auroral oval, *Geophys. Res. Lett*., **36**, L18101, doi:10.1029/2009GL039997

Milan, S. E., P. D. Boakes, and B. Hubert (2008), Response of the expanding/contracting polar cap to weak and strong solar wind driving: Implications for substorm onset, *J. Geophys. Res*., **113**, A09215, doi:10.1029/2008JA013340

Milan, S.E., J. Hutchinson, P. D. Boakes, and B. Hubert (2009a) Influences on the radius of the auroral oval, *Annales Geophys*., **27** (7), 2913 – 2924 doi: 10.5194/angeo-27-2913-2009

Milan, S. E., Grocott, A., Forsyth, C., Imber, S. M., Boakes, P. D., and Hubert, B. (2009b) A superposed epoch analysis of auroral evolution during substorm growth, onset and recovery: open magnetic flux control of substorm intensity, *Annales Geophys*., **27** (2), 659-668, doi: 10.5194/angeo-27-659-2009

Milan, S. E., Carter, J. A., Sangha, H., Bower, G. E., & Anderson, B. J. (2021). Magnetospheric flux throughput in the Dungey cycle: Identification of convection state during 2010. *Journal of Geophysical Research: Space Physics*, **126**, e2020JA028437. doi: 10.1029/2020JA028437

Murayama, T., T. Aoki, H. Nakai, and N. Hakamada (1980), Empirical formula to relate the auroral electrojet intensity with interplanetary parameters, *Planet. Space Sci.*, **28**, 803–813, doi: 10.1016/0032-0633(80)90078-1

Nelder, JA & R. Mead (1965) A Simplex Method for Function Minimization, *The Computer Journal*, **7**, (4), 308–313. doi: 10.1093/comjnl/7.4.308 and Errata, *The Computer Journal*, 8 (1) Issue 1, 27, doi: org/10.1093/comjnl/8.1.27

Newell, P. T., & J. W. Gjerloev (2011), Evaluation of SuperMAG auroral electrojet indices as indicators of substorms and auroral power, *J. Geophys. Res.*, **116**, A12211, doi:10.1029/2011JA016779.

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

Newell, P. T., Sotirelis, T., Liou, K., Meng, C.-I., and Rich, F. J. (2007), A nearly universal solar wind-magnetosphere coupling function inferred from 10 magnetospheric state variables, *J. Geophys. Res*., **112**, A01206, doi:10.1029/2006JA012015.

Newell, P. T., Sotirelis, T., Liou, K., and Rich, F. J. (2008), Pairs of solar wind-magnetosphere coupling functions: Combining a merging term with a viscous term works best, *J. Geophys. Res*., **113**, A04218, doi:10.1029/2007JA012825.

O'Brien, T. P., and McPherron, R. L. (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., S. M. Thompson, and R. L. McPherron (2002), Steady magnetospheric convection: Statistical signatures in the solar wind and AE, *Geophys. Res. Lett.,* **29** (7), 1130-1133, doi:10.1029/2001GL014641.

Owens, M. J. (2018). Time-window approaches to space-weather forecast metrics: A solar wind case study. Space Weather, 16, 1847– 1861, doi: 10.1029/2018SW002059

Partamies, N., Juusola, L., Tanskanen, E., Kauristie, K. (2013) Statistical properties of substorms during different storm and solar cycle phases, *Annales Geophys*., **31** (2), 349-358, doi: 10.5194/angeo-31-349-2013

Perreault, P.D. (1974). On the relationship between interplanetary magnetic fields and magnetospheric storms and substorms, PhD thesis, Univ. of Alaska Fairbanks, Ann Arbor. (Order No. 7910269). Available from ProQuest One Academic. (302740827). Retrieved from https://www.proquest.com/dissertations-theses/on-relationship-between-interplanetary-magnetic/docview/302740827/se-2?accountid=13460

Perreault, P.D., & Akasofu, S. I. (1978). A study of geomagnetic storms. *Geophysical Journal International*, **54**, (3), 547-573. doi: 10.1111/j.1365-246X.1978.tb05494.x

Pulkkinen, T.I. and M. Wiltberger (2000) Thin current sheet evolution as seen in observations, empirical models and MHD simulations, *Geophysical Research Letters*, **27**, 1363-1366 doi: 10.1029/1999GL003726

Shepherd, S.G. (2007) Polar cap potential saturation: Observations, theory, and modeling, *J. Atmos. Sol.-Terr. Phys*., **69** (3), 234-248, doi : 10.1016/j.jastp.2006.07.022

Stephens, G. K., Bingham, S. T., Sitnov, M. I., Gkioulidou, M., Merkin, V. G., Korth, H., et al. (2020). Storm time plasma pressure inferred from multi-mission measurements and its validation using Van Allen Probes particle data. *Space Weather,* **18**, e2020SW002583. doi: 10.1029/2020SW002583

Svalgaard, L. (1977), Geomagnetic activity: Dependence on solar wind parameters, in *A Monograph from Skylab Solar Workshop 1, Coronal Holes and High Speed Wind Streams*, edited by J. B. Zirker, pp. 371–441, Colo. Assoc. Univ. Press, Boulder.

Takalo, J., & K. Mursula (2001), A model for the diurnal universal time variation of the Dst index, *J. Geophys. Res*., **106** (A6), 10905– 10913, doi:10.1029/2000JA000231.

Temerin, M., and X. Li (2006), Dst model for 1995– 2002, *J. Geophys. Res*., **111**, A04221, doi: 10.1029/2005JA011257

Vassiliadis, D. (2006), Systems theory for geospace plasma dynamics, *Rev. Geophys*., **44**, RG2002, doi:10.1029/2004RG000161.

Vassiliadis, D., A. Klimas, D. Baker, and D. Roberts (1995), A description of the solar wind-magnetosphere coupling based on nonlinear filters, *J. Geophys. Res*., **100** (A3), 3495–3512, doi: 10.1029/94JA02725.

Vasyliunas, V. M., Kan, J. R., Siscoe, G. L., & Akasofu, S.-I. (1982). Scaling relations governing magnetospheric energy transfer, *Planetary and Space Science*, **30** (4), 359–365. doi: 10.1016/0032-0633(82)90041-1

Walsh, B. M., & Zou, Y. (2021). The role of magnetospheric plasma in solar wind-magnetosphere coupling: A review. *Journal of Atmospheric and Solar-Terrestrial Physics*, **219**, 105644. doi: 10.1016/j.jastp.2021.105644

Walsh, B. M., Bhakyapaibul, T., & Zou, Y. (2019). Quantifying the uncertainty of using solar wind measurements for geospace inputs. *Journal of Geophysical Research: Space Physics*, **124**, 3291–3302, doi: 10.1029/2019JA026507

Wang, C., Han, J. P., Li, H., Peng, Z., and Richardson, J. D. (2014), Solar wind-magnetosphere energy coupling function fitting: Results from a global MHD simulation, *J. Geophys. Res. Space Physics*, **119**, 6199– 6212, doi:10.**1002**/2014JA019834.

Weimer, D. R., D. M. Ober, N. C. Maynard, M. R. Collier, D. J. McComas, N. F. Ness, C. W. Smith, and J. Watermann (2003) Predicting interplanetary magnetic field (IMF) propagation delay times using the minimum variance technique, *J. Geophys. Res.*, **108** (A1), 1026, doi: 10.1029/2002JA009405.

Weygand, J.M. & E. Zesta (2008), Comparison of auroral electrojet indices in the Northern and Southern Hemispheres, *J. Geophys. Res*.,**113**, A08202, doi:10.1029/2008JA013055.

Yu, Y., and Ridley, A. J. (2013), Exploring the influence of ionospheric O+ outflow on magnetospheric dynamics: dependence on the source location, *J. Geophys. Res. Space Physics*, **118**, 1711– 1722, doi:10.1029

Zhang, H., S. Fu, Z. Pu, J. Lu, J. Zhong, C. Zhu, W. Wan and L. Liu (2019) Statistics on the magnetosheath properties related to magnetopause magnetic reconnection, *Astrophys. J*., **880** (2), 122 (37p.p.), doi 10.3847/1538-4357/ab290e

Figures

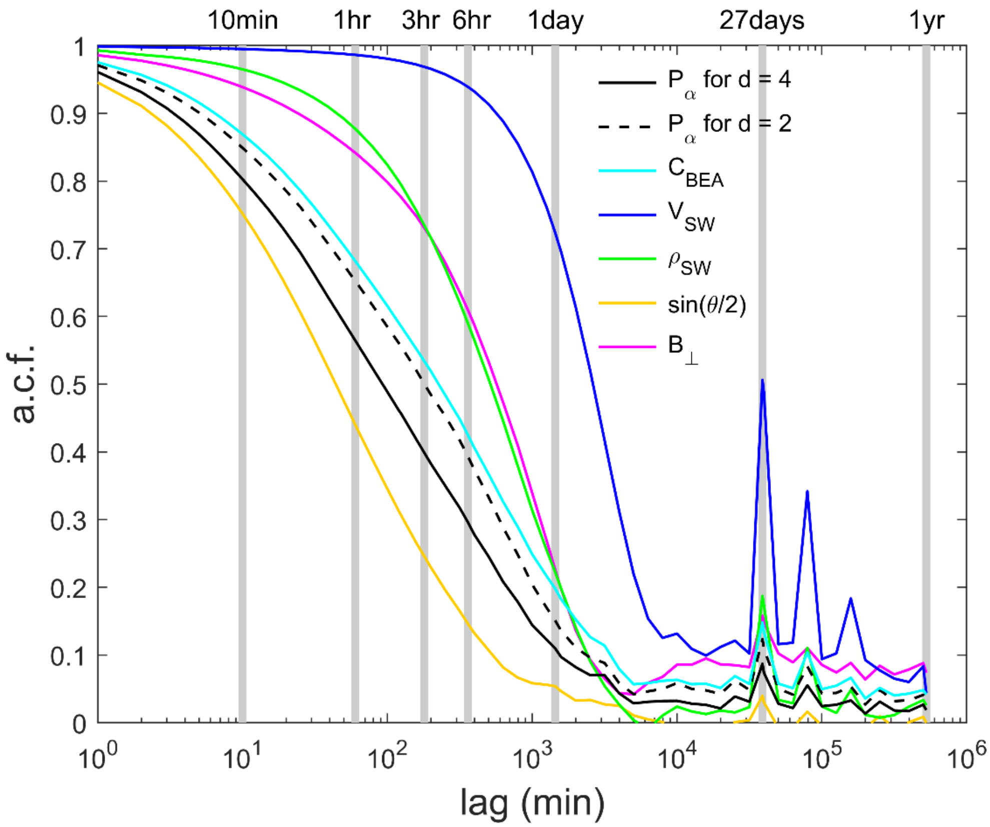


Figure 1. (a) Autocorrelation functions (a.c.f.s) as a function of lag time (on a log scale) for 1-min samples of: (blue) solar wind speed, V_{SW}; (green) solar wind mass density, ρ_{SW}; (mauve) the transverse component of the interplanetary magnetic field (IMF), $B_\perp$; (orange) the IMF orientation factor *sin*(θ/2), where θ is the IMF clock angle in GSM coordinates; (black) the estimated power input to the magnetosphere, P_α for a coupling exponent of α = 1/3 and *d* of 4; (black dashed) P_α for a coupling exponent of α = 1/3 and *d* of 2; and (cyan) the coupling function of *Boyle et al.* (1997) (see Equation 7 of text). The vertical gray lines marks times of 10 minutes, 1hour, 2 hours, 6 hours, 1 day, 27 days and 1 year.

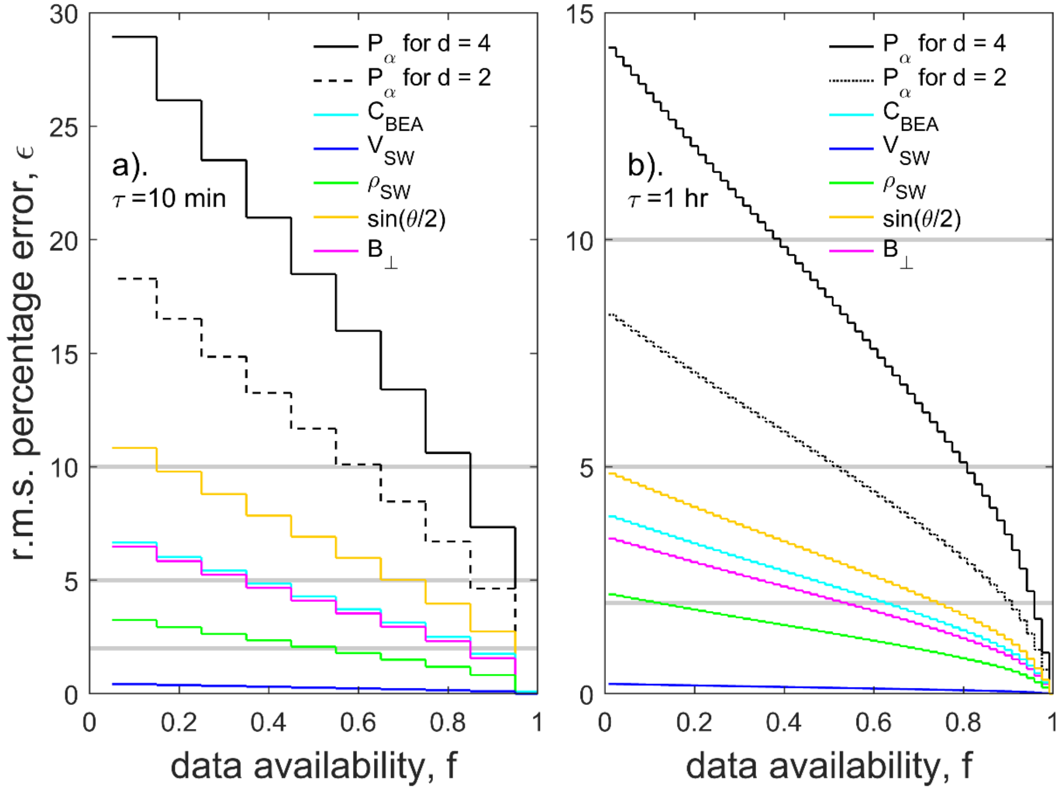


Figure 2. The root mean square (r.m.s.) percentage errors, ε, in (a) 10-minute and (b) 1-hour means of 1-minute integrated data as a function of data availability within the averaging interval, *f*, for the same parameters as in Figure 1, shown using the same color scheme. The horizontal gray lines in are uncertainties ε of 10%, 5% and 2%, in both panels and the graphs set threshold requirements for *f* to meet those levels of uncertainty. The data in (b) are based on all 589293 (boxcar) hourly means in the Omni data for 1996-2020 (inclusive) for which all 60 1-minute integrations of all parameters were available. The data in (a) are based on the 5153517 (boxcar) 10-minute means from the same interval for which all 10 1-minute integrations of all parameters were available. For each of these boxcar means, (1-*f*) samples were removed at random 10 times and the r.m.s. error thereby introduced computed by comparison with the value for all data (*f* = 1). (based on Figure 1b of *Lockwood et al.*, 2019).

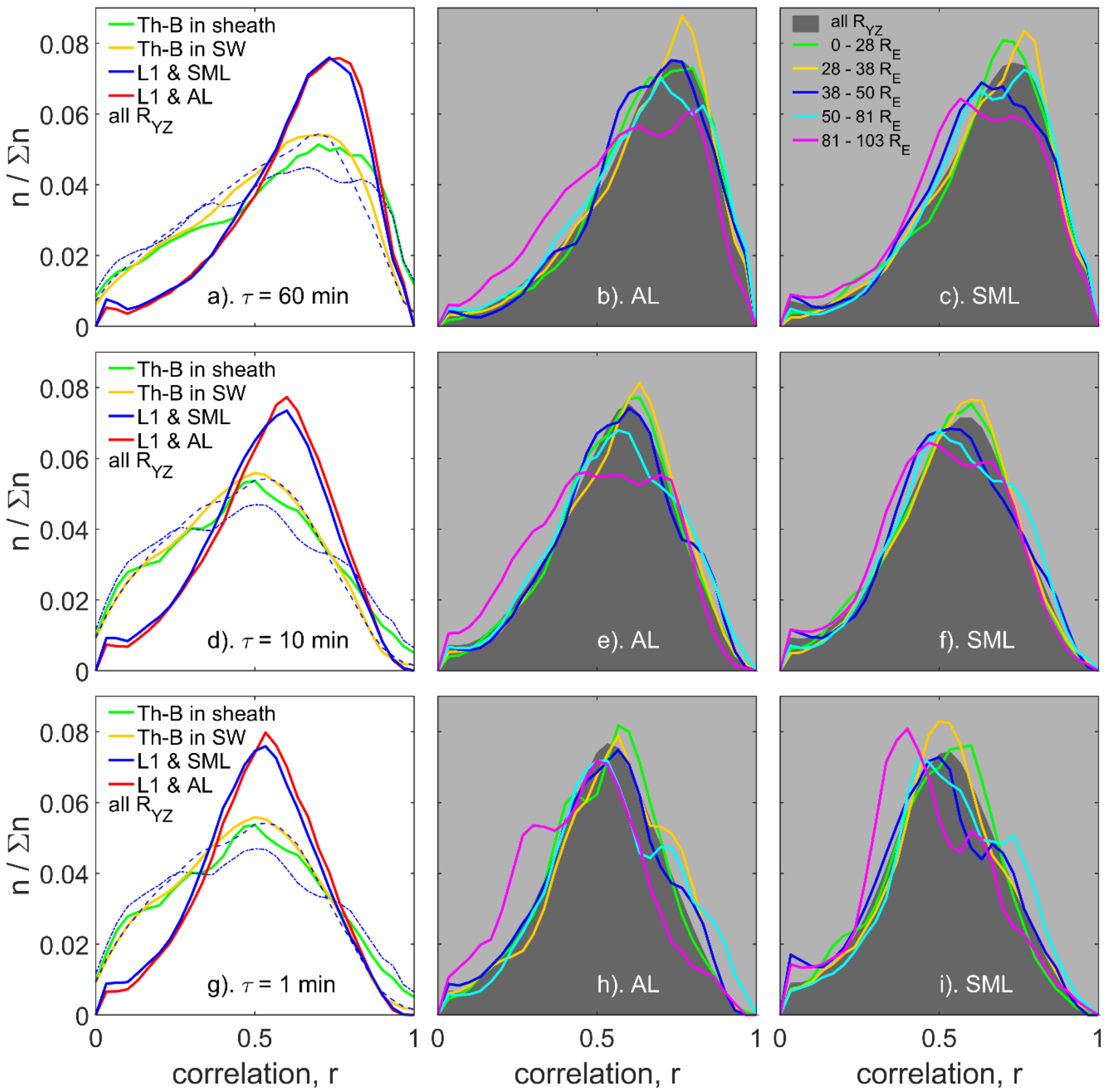


Figure 3. Distributions of correlation coefficients at optimum lags over 2-day intervals between an example coupling function from L1 satellite data (the Omni dataset) and the *AL* and *SML* geomagnetic indices. The example used is the optimum coupling function C_f with the best-fit coefficients for *SML* found by *Lockwood and McWilliams* (2021) giving $C_f = B^{0.662} (m_{sw}N_{sw})^{0.016} N_{sw}^{1.746} \sin^{5.2}(\theta/2)$. The data are for 1996-2018 (inclusive) for *AL* and 1996-2020 (inclusive) for *SML*. The bottom panels (g, h and i) are for 1-minute integrations of data ($\tau = 1$ min), the middle panels (d, e and f) for 10-minute running means of the 1-minute data ($\tau = 10$ min), and the top panels (a, b and c) for 1-hour running means of the 1-minute data ($\tau = 60$ min). For each 2-day interval, the correlations was evaluated at lags between the L1 data (propagated to the nose of the bow shock using the procedure of *Weimer et al.* (2003) and the geomagnetic index of between −80 min. and +120 min. and the peak value used. The left-hand plots compare the distributions for all data for *AL* (in red) and *SML* (in blue). These distributions are repeated by the light grey shaded areas for *AL* in

the middle panels and for *SML* in the right-hand panels: also shown by the colored lines in these panels are the variations for 5 quantile ranges (i.e., 20% of the data in each) of $R_{YZ} = (Y^2 + Z^2)^{1/2}$, the distance of the L1 craft from the Sun-Earth line (the *X* axis). The left-hand panels also show the distribution of correlations with SML for the same coupling function measured by THEMIS-B over the interval 2011-2018 when it was in the undisturbed solar wind (orange line) and in the magnetosheath (green line). The thin blue dashed line and dot-dash lines are the corresponding distributions for the L1 data and SML for the same times (allowing for the optimum propagation delay) as the THEMIS-B data from, respectively, the undisturbed solar wind and the magnetosheath.

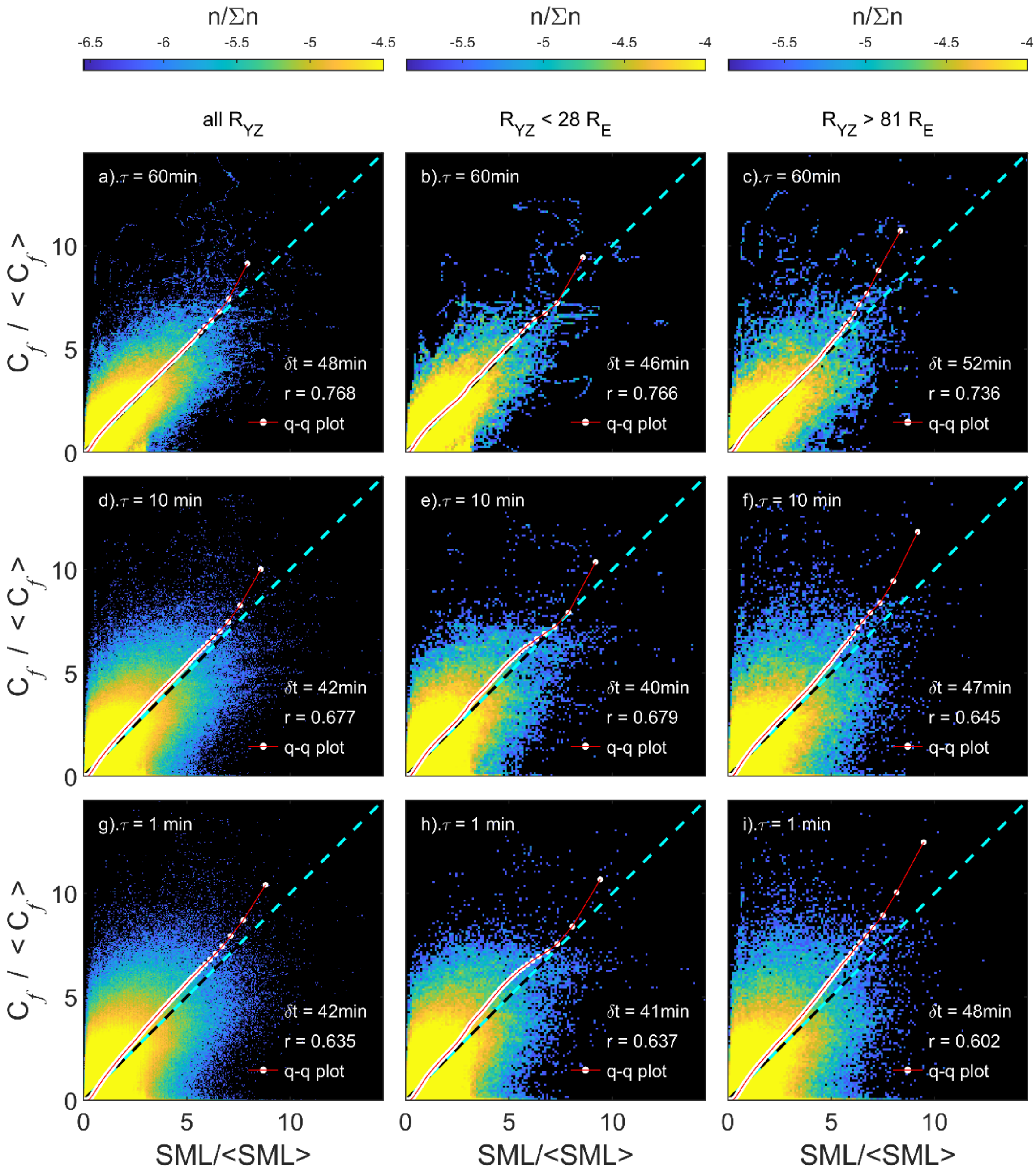


Figure 4. Data density plots overlaid with quantile-quantile (q-q) plots to compare the distributions of the *SML* index (lagged by δt) and the example coupling function P_α used in Figure 1 of *Lockwood and McWilliams* (2022). The cyan dashed line in each panel is perfect agreement of lagged P_α and *SML*. The data are for 1996-2020 (inclusive) which yields Σn = 12,720,434 valid data pairs for τ = 1 min. As in Figure 1, the bottom panels (g, h and i) are for 1-minute integrated data (τ = 1 min), the middle panels (d, e and f) for 10-minute running means of the 1-minute data (τ = 10 min), and the top panels (a, b and c) for 1-hour running means of the 1-minute data (τ = 60 min). The left-hand plots are for all data, irrespective of the location of the L1 monitor. The middle panels are for the 20% of

P_α samples closest to the Sun-Earth line ($R_{YZ} = (Y^2 + Z^2)^{1/2} < 28R_E$) and the right-hand panels are for the 20% of P_α samples furthest from the Sun-Earth line ($R_{YZ} > 81R_E$). The q-q plots use 1000 quantiles, 0.1% apart, shown by the white dots, the largest value being the 99.9% quantile of both the P_α and *SML* distributions. The underlying data density plots shows the fraction of samples $n/\Sigma n$, colored in bins that are 0.05 wide in both the $P_\alpha/<P_\alpha>$ and *SML*/<*SML*> in the left hand column (for the full dataset) and 0.10 wide in the other two columns (for which Σn is lower by a factor of 5 than for the full dataset). The lower end of the logarithmic color scales used is the one count level (i.e., for $n = 1$). The peak correlation r and the lag δt (between predicted arrival of the solar wind at the nose of the bow shock and the *SML* response) giving peak correlation is also given in each panel.

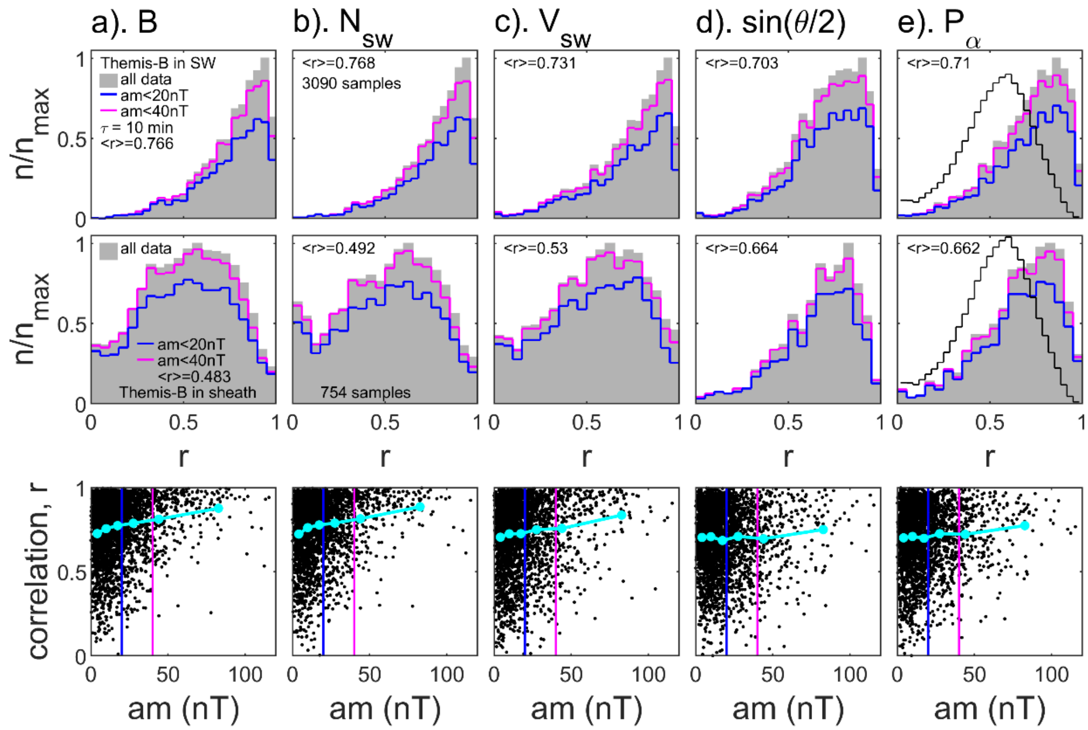


Figure 5. Peak correlation coefficients *r* between data from the near-Earth Themis-B spacecraft for the interval 2011-2019 (inclusive) and the lagged Omni dataset (from interplanetary spacecraft in halo orbits around the L1 point) for 8-hour intervals (which yields 48 fully-independent samples in each correlation). This plot is for 10-point running (boxcar) means of 1-minute integrations of the Omni data and corresponding means of the Themis-B data that have been linearly interpolated from the basic of 96-second integrations to the times of the Omni data. The upper panel gives the distribution of peak correlation coefficients, *r* for when Themis-B is in the undisturbed solar wind. The middle panels are the corresponding distributions for when Themis-B is in the magnetosheath. The lower panels are scatter plots of *r* for the Themis-B data from the undisturbed solar wind as a function of the *am* geomagnetic index (interpolated linearly from the 3-hourly index values). The L1 satellite data are lagged using the optimum lag which yields peak correlation. Columns are for (a) the Interplanetary Magnetic Field (IMF), *B*; (b) the solar wind number density, N_{sw}; (c) the solar wind speed, V_{sw}; (d) the IMF orientation factor, sin(θ/2), where θ is the IMF clock angle in Geocentric Solar Magnetospheric (GSM) coordinates; and (e) an example of a coupling function, the *Vasyluinas et al.* (1982) power input into the magnetosphere estimate, P_{α} for a coupling exponent $\alpha = 1/3$ and $d = 4$. Note that the mean ion mass m_{sw} for the Themis-B data are taken to be the same as from the Omni dataset. A correlation coefficient is assigned only if more than 90% of the possible 1-minute data pairs are available for all the 6 parameters, which yields 3090 correlations for each parameter for when Themis-B is in the undisturbed solar wind and 754 for when it is in the magnetosheath. In the upper plots, the histograms shaded grey are for all data, whereas the mauve and orange lines are histograms for *am* < 40 nT and *am* < 20 nT, respectively. The number of samples *n* in each bin is plotted as a ratio of is peak value. The mean value of *r* is also given. In the bottom panels, the *am* = 40 nT and *am* = 20 nT thresholds used in the upper panels are shown by mauve and blue vertical lines and the cyan dots are means of *r* and *am* in 6 equi-spaced quantile ranges of the *am* index. The

black lines in the upper two panels of part (e) give the distribution of r values between the optimally-lagged L1 values of P_α and the *SML* index for this τ of 10 min. This distribution has been normalized to the n/n_{max} y axis such that the area under the black line is the same as that of the grey area and gives a mean correlation between P_α and the *SML* of 0.531.

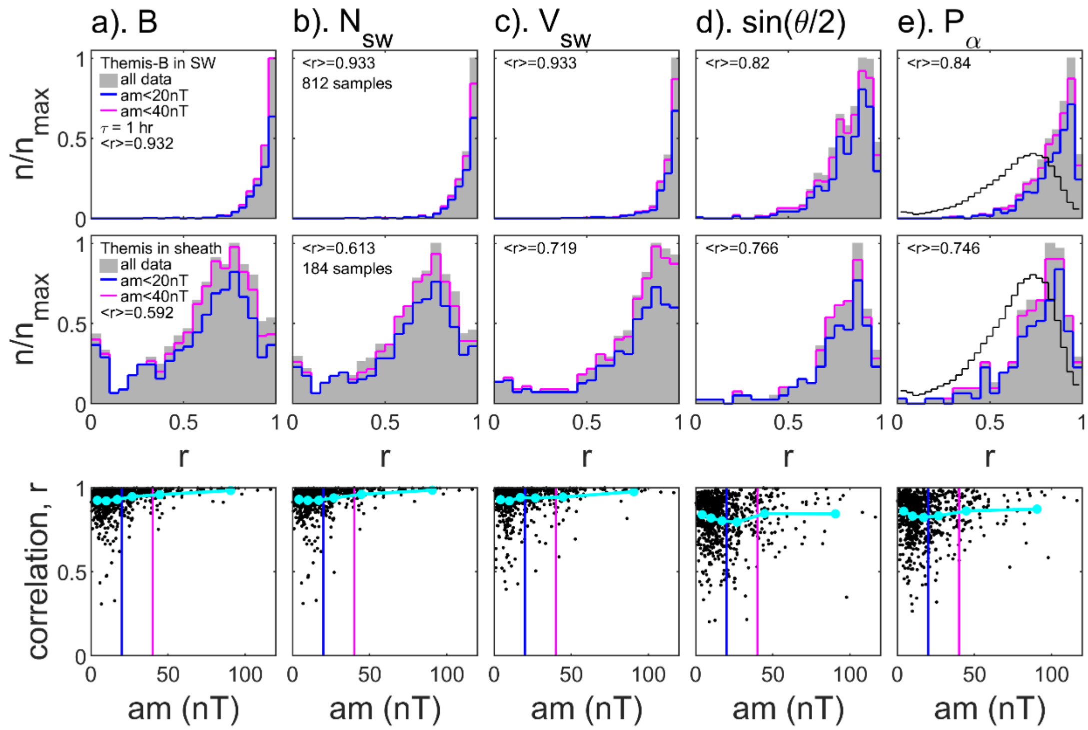


Figure 6. The same as Figure 5 for hourly means of the data from 2011-2019, inclusive. The correlations are made using 60-point running (boxcar) means of both the Omni and Themis-B data for 2-day data segments (which again yields 48 fully independent samples in each interval). Correlation coefficients *r* are only given if at least 90% the number of potential 1-minute data pairs are available in the 2-day interval for all 6 parameters which yields 812 correlations for each parameter for when Themis-B in in the undisturbed solar wind and 184 for when it is in the magnetosheath. The black lines in the upper two panels of part (e) is the normalized distribution of correlations between the optimally-lagged L1 values of P_α and the *SML* index for this τ of 1 hour which has a mean value of 0.632.

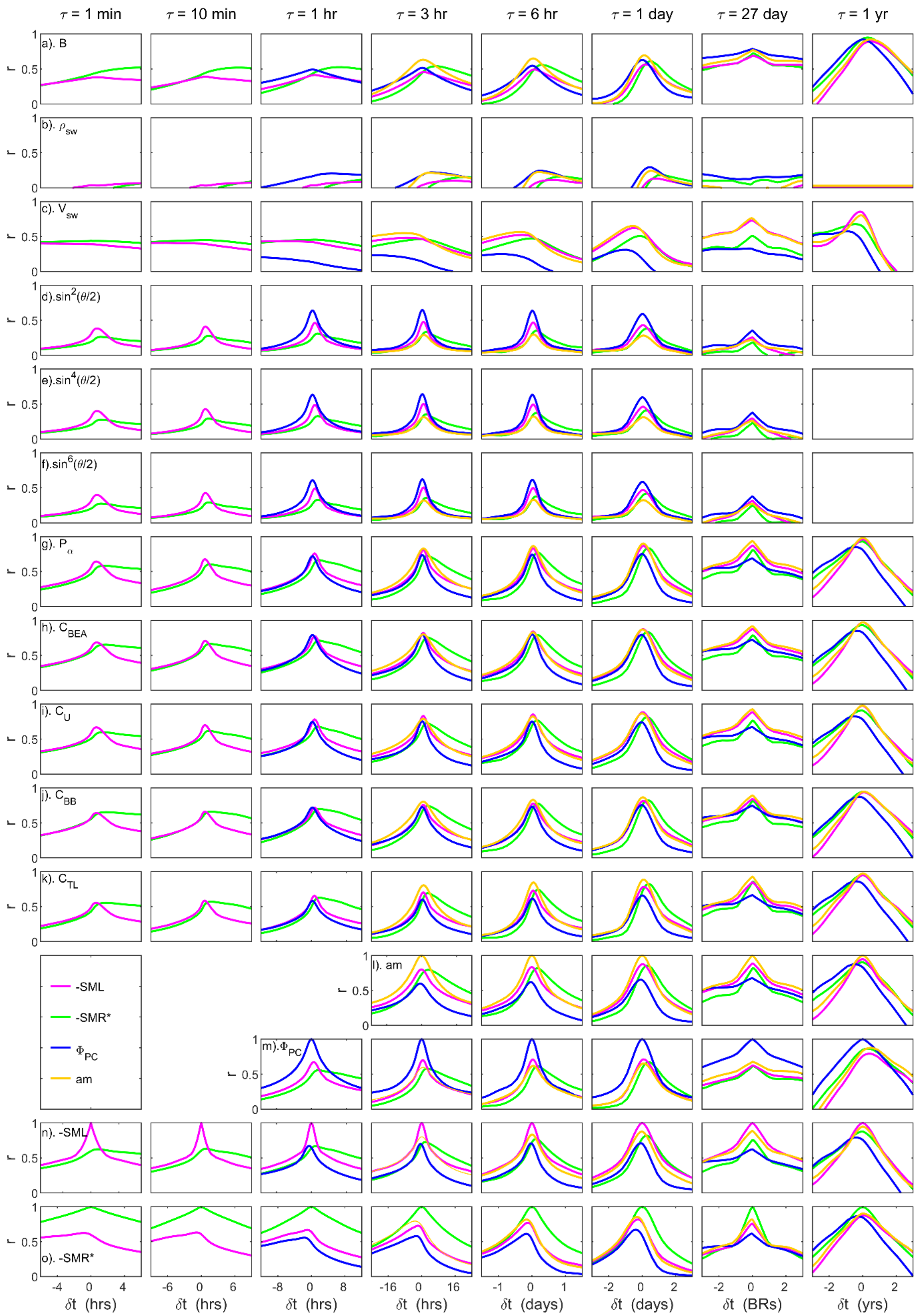
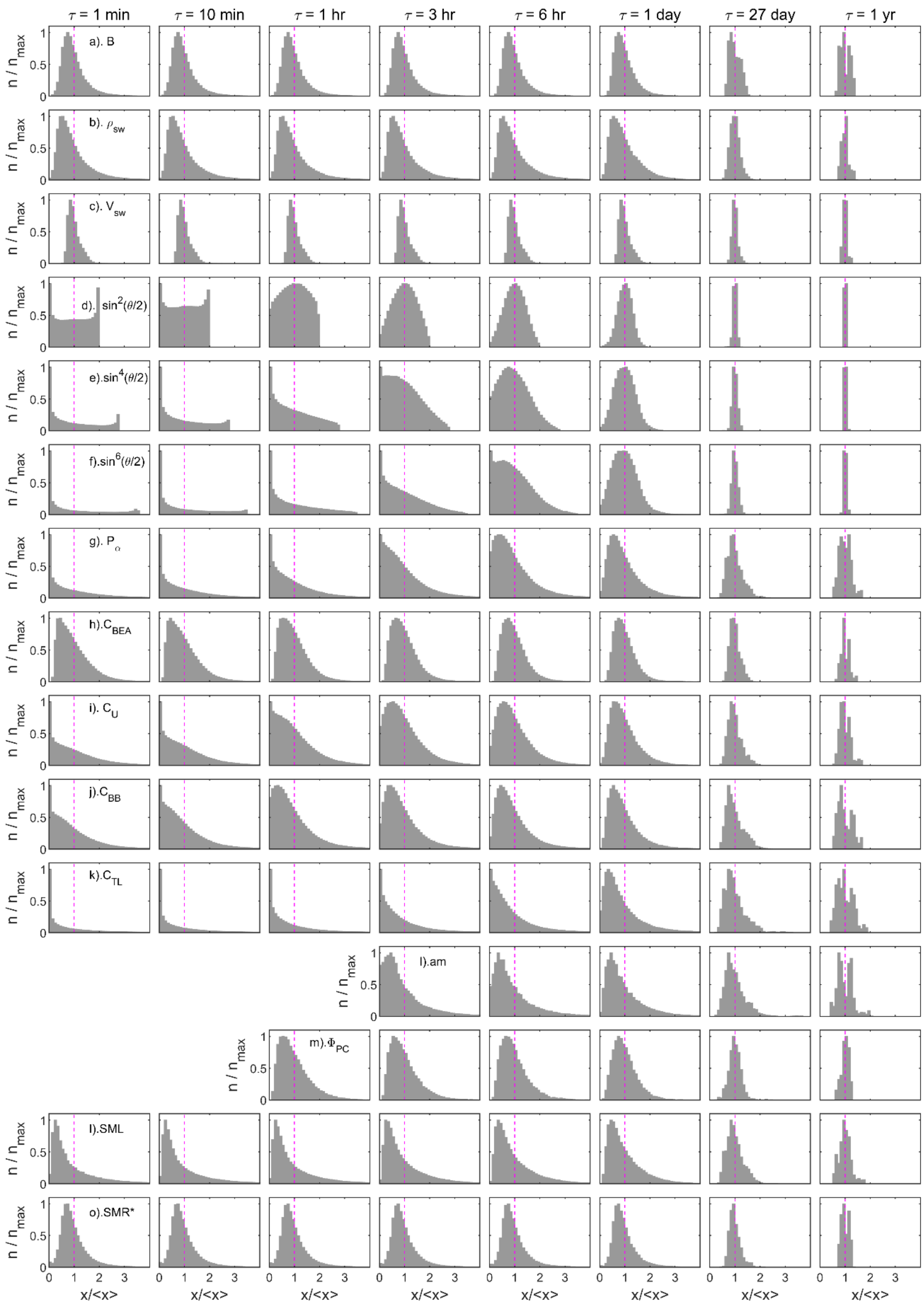


Figure 7. Lag correlograms for the various averaging timescales τ between interplanetary parameters, coupling functions and terrestrial space weather disturbance indices using data from

1996-2020. The columns (from left to right) are for τ = 1 min.; τ = 10min; τ = 1hr; τ = 3 hrs.; τ = 6hrs; τ = 1 day; τ = 27 days; and τ = 1 year. The top 6 rows (a-f) are interplanetary parameters, the next 5 rows (g-k) are for different coupling function combinations of interplanetary parameters and the bottom 4 rows (l-o) are for the selected terrestrial disturbance indexes. Specifically, rows are for: (a) IMF, *B*; (b) solar wind mass density, $_{SW}$; (c) solar wind speed, V_{SW}; (d) the IMF orientation factor $F(\theta) = \sin^2(\theta/2)$; (e) $F(\theta) = \sin^4(\theta/2)$; (f) $F(\theta) = \sin^6(\theta/2)$; (g) the *Vasyluinas et al.* (1982) power input into the magnetosphere estimate P_α for α = 1/3 and *d* = 4; (h) the *Boyle et al.* (1997) transpolar voltage prediction, C_{BEA}; (i) the empirical “Nearly Universal” coupling function of *Newell et al.* (2007), C_U (for which *d* = 2.67); (j) the theory based coupling function of *Borovsky and Birn* (2014) (for which *d* = 2); (k) the empirical coupling function of *Temerin and Li* (2006), C_{TL} (for which *d* = 6); (l) the *am* planetary geomagnetic index; (m) the transpolar voltage Φ_{PC} from the SuperDARN radar using the *Lockwood and McWilliams* (2021) dataset with mean number of echoes exceeding 255; (n) the SuperMAG *SML* auroral electrojet index (*Newell and Gjerloev*, 2011) and (o) the modified SuperMAG *SMR* ring current index, *SMR** (see Section 2 of text). Note that the Φ_{PC} data are hourly integrations and *am* is a range index derived from 3-hourly intervals. In each panel, the correlations if the parameter in question for that row with –*SML*, –*SMR**, Φ_{PC} ($\tau \geq$ 1 hr. only) and *am* ($\tau \geq$ 3 hrs. only) are shown by mauve, green, blue and orange lines, respectively. In all panels, a positive lag corresponds to the parameter, defined by the row number, is lagged. Note that the orange lines in row (l) are autocorrelation functions of *am*; the blue lines in row (m) are autocorrelation functions of Φ_{PC}; the mauve lines in (n) are autocorrelation functions of *SML*; and the green lines in (o) are autocorrelation functions of *SMR**. Correlations that do not meet the 2σ significance level are omitted.

1404 

1405 **Figure 8.** Distributions of the same parameters x and timescales τ as in Figure 7, also for data from
1406 1996-2020.. In each case, the histogram is of x (normalized to its overall mean value, i.e., $x/\langle x \rangle$)

and the vertical mauve dashed line is at $x/<x> = 1$. Bins are of width $<x>/10$ and histograms are normalized, with the number of samples in each bin n, being plotted as a ratio of its maximum value, n_{max}.

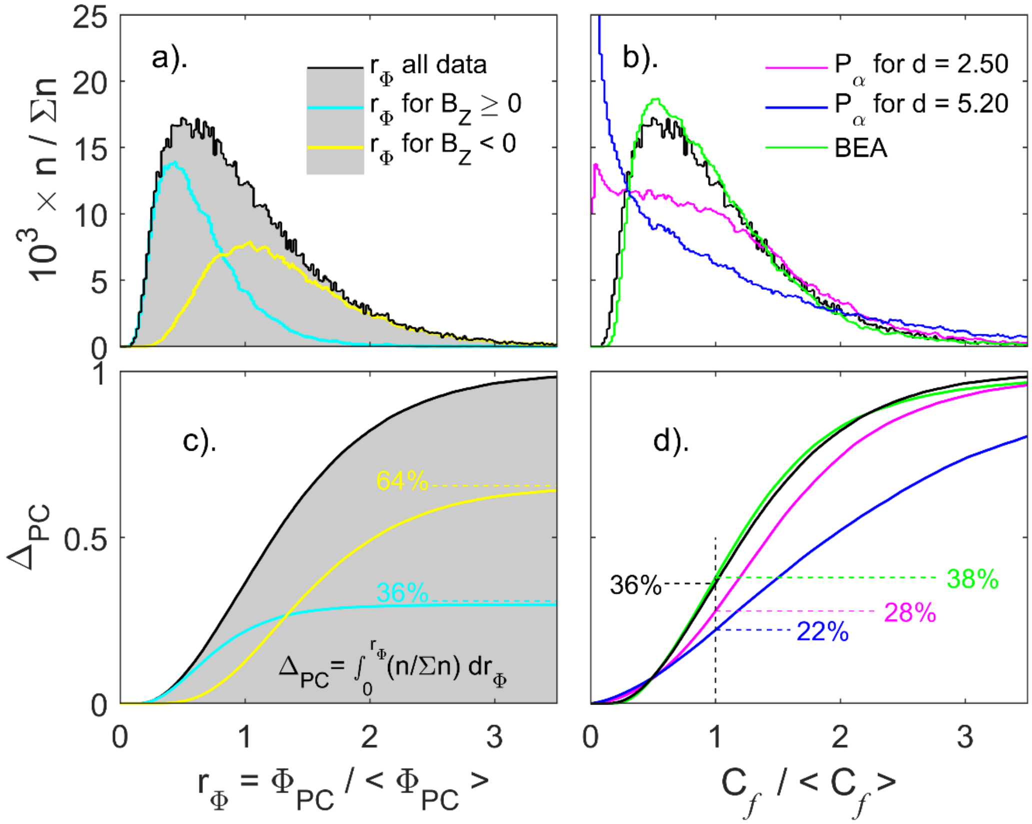


Figure 9. (Top) probability density functions (pdfs) and (bottom) cumulative distribution functions (c.d.f.) of (left) normalised observed transpolar voltage $\Phi_{PC}/<\Phi_{PC}>$ and (right) normalised hourly coulping functions, $C_f/< C_f>$. In the left hand plots tellow lines are for data when the lagged southward IMF is southward in the GSM frame ($B_z < 0$, using the optimum lag of 30 min. found by *Lockwood and McWilliams*, 2021) and the cyan line for when it is northward ($B_z \geq 0$)**.** In the right-hand plots, the green line is for the transpolar voltage predictor of *Boyle et al.* (1997); C_{BEA}, the mauve line the best empirical C_f fit to the transpolar voltage derived in Figure @@ of *Lockwood and McWilliams* (2022) and shown in Figure @@d (giving a= 0.642, b = 0.018, c = 0.552, d = 2.50). The blue line is the corresponding best fit to the *SML* geomagnetic index (with a= 0.662, b = 0.061, c = 1.746, d = 5.20: see Figure 10). The black lines in (b) and (d) are for the observations, i.e., they are the same as in (a) and (c), respectively. Part (c) shows that 64% of the total convection flux transport in the magnetosphere takes place during southward IMF and 36% takes place during northward IMF or IMF B_Z = 0. The vertical dashed line in part (d) is at $C_f/< C_f>$ = 1 and shows that below-average tranpolar voltage is responsible for 35% of the observed flux transport over the polar cap, which is very close to the 36% predicted by C_{BEA}. However, the poorer fits to the low values of the distribution mean that the empirical fits using the $F(\theta) = sin^d(\theta/2)$ formulation do not predict this number as well, C_f for d = 2.50 giving 28% and , C_f for d = 5.20 giving 22%.

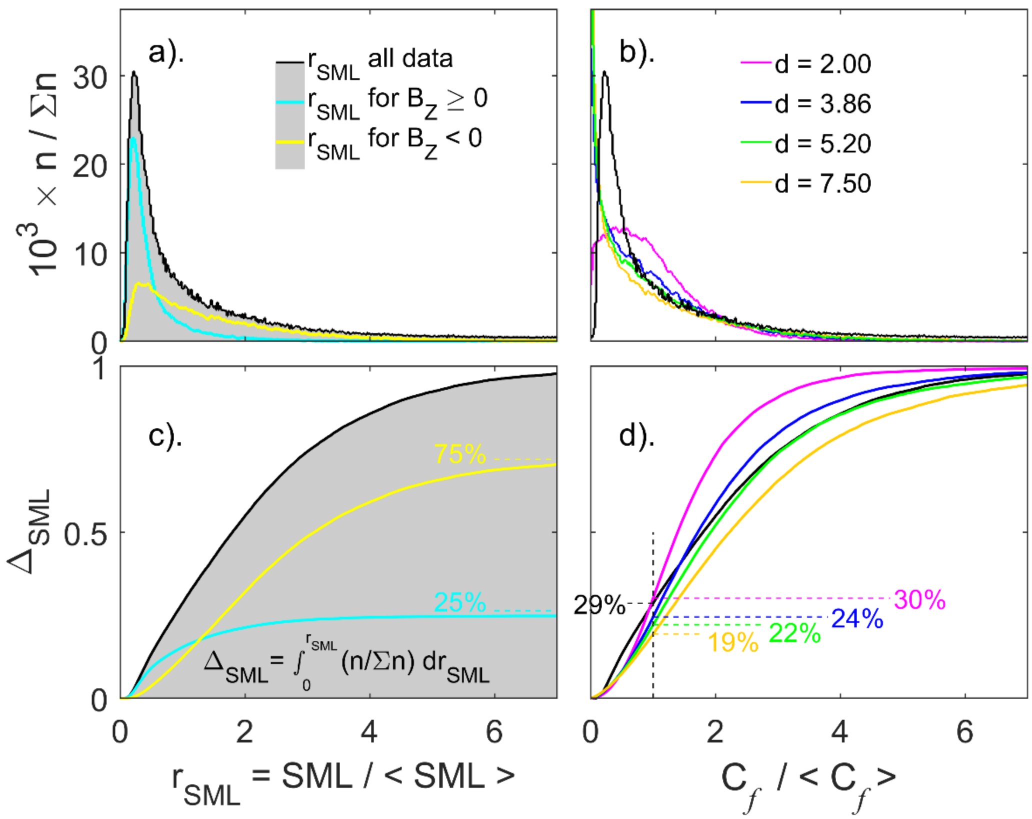


Figure 10. The same as Figure 9 for fits to the *SML* geomagnetic activity. Part (c) shows that 75% of the integrated activity in takes place during southward IMF and 25% takes place during northward IMF or IMF $B_Z = 0$. The vertical dashed line in part (d) shows that below average SML is responsible for 29% of the integrated activity which is very close to the 30% predicted by C_f for $d = 2.00$ (mauve line) but exceeds the 24% for $d = 3.86$ (which gives peak correlation between C_f and *SML*, *r* – the blue line), the 22% for the optimum d of 5.20 (which yields linearity between C_f and *SML* - green line) and the 19% for the excessive *d* of 7.5.

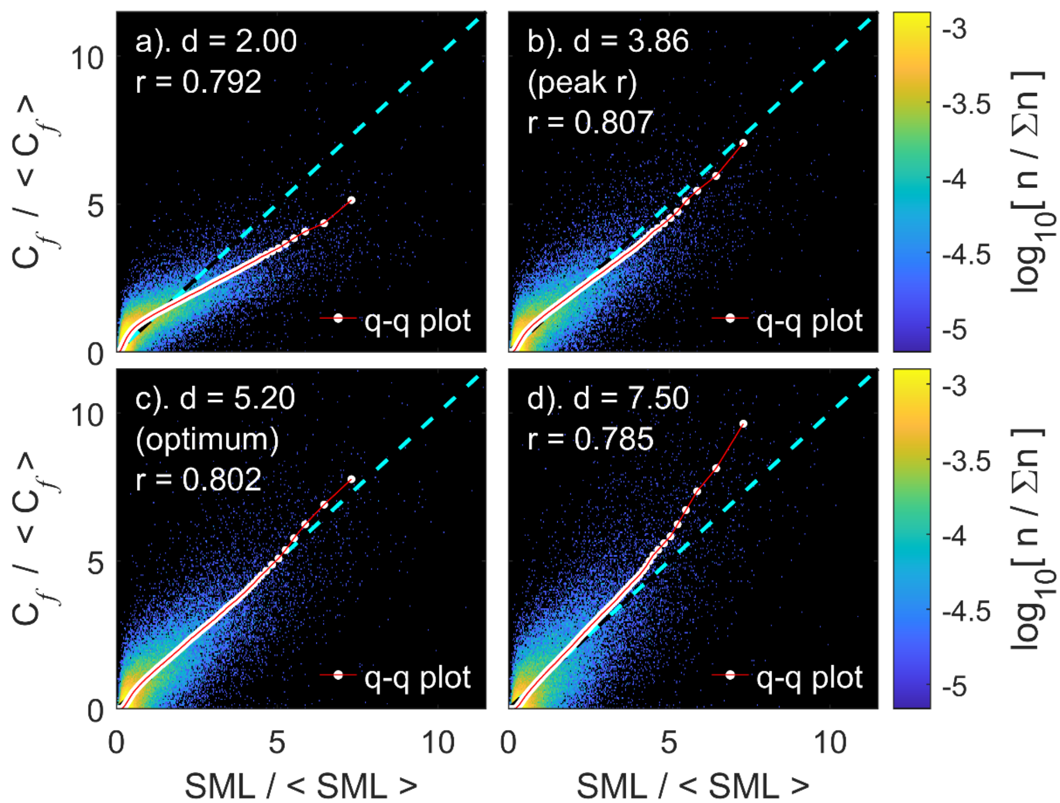


Figure 11. Data density and overlaid quantile-quantile (q-q) plots for the normalized hourly averaged *SML* geomagnetic index and the normalized empirical hourly averaged coupling function C_f for: (a) d = 2.00 (which gives the mauve lines in Figure 10); (b) d = 3.86 (which gives peak correlation between C_f and *SML* and the blue lines in Figure 10); (c). d = 5.20 (which yields linearity between C_f and *SML* and the green lines in Figure 10) and d = 7.5 (which gives the orange lines in Figure 10). 500 quantiles (white dots) are used at separation of 0.2% and the lower end of the colour scale used is below $log_{10}(1/\Sigma n)$ (i.e., the one count level, n = 1) to ensure that even single hourly samples show up as a blue pixel. Pixels (counting bins) are 0.03×0.03 in size. For these hourly data Σn = 61922 and so there are either 123 or 124 hourly values in each quantile.

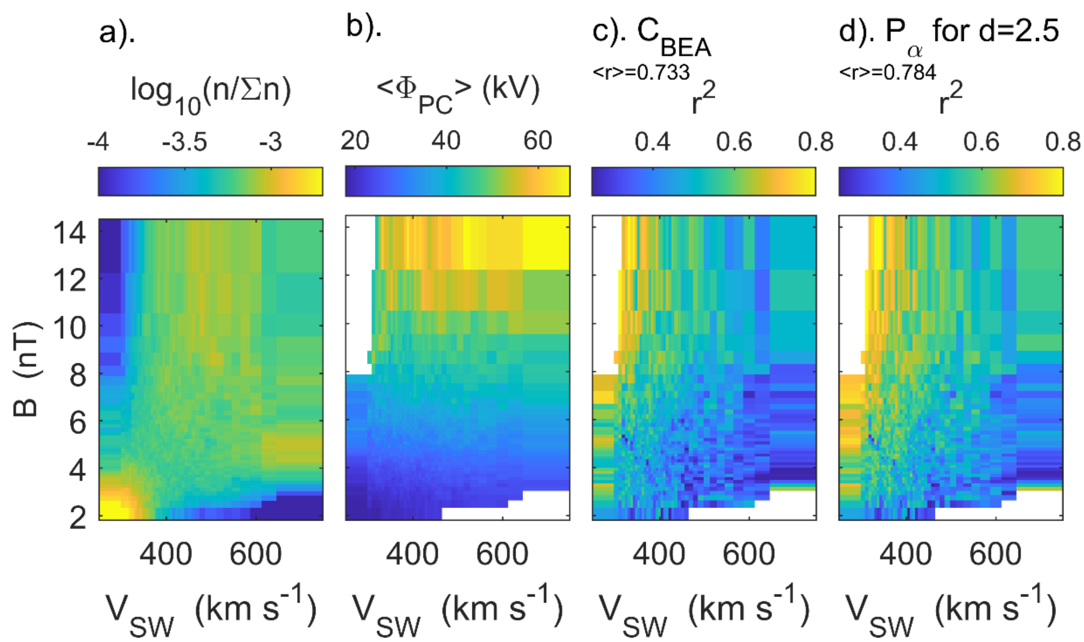


Figure 12. Analysis of the fraction of the variance r^2 of the transpolar voltage Φ_{PC} explained by coupling functions (c) C_{BEA} and (d) P_α for d = 2.5 , where r is the correlation coefficient. Data are divided into 40 inter-quantile ranges of both the solar wind speed V_{sw} and IMF B. The peak of the lag correlograms r is found and r^2 plotted as a function of V_{sw} (along the x-axis) and B (along the y axis). Part (a) gives the fraction of samples in each bin (on a logarithmic scale); (b) gives the mean transpolar voltage in each bin.

Funding

The work presented in this paper was supported by two UKRI grants to the University of Reading: the UK Science and Technology Facilities Council (STFC) consolidated grant number ST/M000885/1 and the UK Natural Environment Research Council (NERC) Directed Highlight Topic “Space Weather Impact on Ground-based Systems (SWIGS)", Grant number NE/P016928/1/

Data Availability

The datasets used in this study are publicly available. The Omni interplanetary are available from NASA’s Space Physics Data Facility **http://omniweb.gsfc.nasa.gov/ow.html** the Themis-B data from NASA’s Coordinated Data Analysis Web (CDAWeb) **https://cdaweb.gsfc.nasa.gov/index.html/**; satellite locations from NASAs satellite Situation Center **https://sscweb.gsfc.nasa.gov**; the *AL* index data from World Data Center for Geomagnetism, Kyoto **http://wdc.kugi.kyoto-u.ac.jp/aeasy/index.html** and the *am* index form the International Service of Geomagnetic Indices (ISGI) **http://isgi.unistra.fr/data_download.php;** the *SML* and *SMR* geomagnetic indices from the SuperMAG project at The Johns Hopkins University: **https://supermag.jhuapl.edu/**. The SuperDARN radar data and associated processing software is available from Virginia Polytechnic Institute and State University **http://vt.superdarn.org/tiki-index.php?page=Data+Access** or from PI groups participating in the project.