# *Does adding solar wind Poynting flux improve the optimum solar wind - magnetosphere coupling function?*

Article

Accepted Version

Lockwood, M. ORCID: https://orcid.org/0000-0002-7397-2172 (2019) Does adding solar wind Poynting flux improve the optimum solar wind - magnetosphere coupling function? Journal of Geophysical Research: Space Physics, 124 (7). pp. 5498-5515. ISSN 2169-9402 doi: https://doi.org/10.1029/2019JA026639 Available at https://centaur.reading.ac.uk/84889/



To link to this article DOI: http://dx.doi.org/10.1029/2019JA026639

Publisher: American Geophysical Union

*submitted to J. Geophys. Res. 19th February 2019*
*Accepted version (with authorized late added sentence) 2 July 2019*

# Does adding solar wind Poynting flux improve the optimum solar wind - magnetosphere coupling function?

Mike Lockwood* (1)

*(1) Department of Meteorology, University of Reading UK*

**Abstract**. We study the contribution of the solar wind Poynting flux $\vec{S}_{sw}$, to the total power input into the magnetosphere. The dominant power delivered by the solar wind is the kinetic energy flux of the particles which is larger than $S_{sw}$ by a factor of order $M_A^2$, where $M_A$ is the Alfvén Mach number The currents $\vec{J}$ flowing in the bow shock and magnetosheath and the electric field $\vec{E}$ of the solar wind give regions where $\vec{J}.\vec{E} < 0$, which are sources of Poynting flux, generated from the kinetic energy flux. For southward IMF, $\vec{E}$ is duskward and the currents in the high-latitude tail magnetopause are also sources of Poynting flux. We show transfer of kinetic energy into the magnetosphere is less efficient than direct entry of $\vec{S}_{sw}$ by a factor $M_A$. Because $M_A$ is typically of order 10, this means that although the power density in the solar wind due to SSW is typically only 1%, it is responsible for of order 10% of the energy input to the magnetosphere. To investigate the effect of this, we add a term to the solar wind-magnetosphere energy coupling function that allows for $\vec{S}_{sw}$ which increases the correlation with the geomagnetic am index for 1995-2017 (inclusive) from 0.908 to 0.924 for 1-day averages and from 0.978 to 0.979 for annual means. The increase for means on daily or smaller timescales is a small improvement but is significant (at over the 3-σ level), whereas the improvement for annual or Carrington-rotation means is not significant.

## 1. Introduction

The basic physics of energy flow into the magnetosphere was elegantly summarized using Poynting's theorem by *Cowley* (1991). He considered only the steady-state case when the near-Earth Interplanetary Magnetic Field (IMF) points southward. This was generalized to cover the substorm phases and northward IMF conditions (which are all inherently non-steady state cases) by *Lockwood* (2004). Global magnetohydrodynamic (MHD) numerical simulations give a unique way of studying the details of this energy flow into the magnetosphere, the storage and deposition in the magnetosphere, and its return from the magnetosphere to the interplanetary medium (*Palmroth et al*., 2003; *Ebihara et al*. (2019). These studies confirm the fundamental physical expectation that the energy density in the solar wind is dominated by the kinetic energy flux of the bulk flow of the particles, it being roughly 2 orders of magnitude larger than the magnitude of the Poynting flux vector $|\vec{S}_{SW}| = S_{SW}$, (and larger than both the magnetic energy flux and the thermal energy flux by about the same factor). The simulations also confirm the expectation that between about 2% and 7% of the solar wind kinetic energy that is incident on the effective cross-sectional area that the magnetosphere presents to the solar wind enters the magnetosphere (*Koskinen and Tanskanen*, 2002).

An interesting point that emerges from the simulations by *Ebihara et al*. (2019) is that, although the fraction of total energy flux in the solar wind that is in the form of Poynting flux is very small, the fraction of power that is delivered to the magnetosphere that originates from that solar wind Poynting flux may not be as small because of the relative inefficiency with which kinetic energy of the solar wind is converted into Poynting flux by currents flowing in the bow shock, magnetosheath and magnetopause (*Cowley*, 1991).

*Vasyluinas et al.* (1982) used dimensional analysis of power input into the magnetosphere to derive a coupling function between the solar wind and the magnetosphere. This coupling function has just one free fit parameter (the coupling exponent $\alpha$) which means it is much less prone to errors associated with statistical "overfitting". *Lockwood et al.* (2019a) have shown that data gaps in the interplanetary data series cause considerable noise in solar wind – magnetosphere correlation studies and this noise can cause major overfitting problems because fits with too many free parameters are fitting the noise and therefore are not robust when considering general datasets. In short, overfitting is damaging predictive power. The theory by *Vasyluinas et al.* (1982) introduces the one free fit parameter, $\alpha$, through a dimensionless term $M_A^{-2\alpha}$ which means that it appears in the exponents for terms in the resulting expression for magnetospheric power input in solar wind velocity ($V_{sw}$), mean ion mass ($m_{sw}$), number density ($N_{sw}$), and the IMF field strength ($B$). The formulation by *Vasyluinas et al.* (1982) was shown to be optimum over a very large range of timescales by *Finch and Lockwood* (2004), although for long averaging timescales (approaching 1 year) the simpler form $V_{sw}^2 B$ (with no IMF orientation factor) performs equally well. This formulation has been used a great many times in diverse areas, for example: to investigate which coupling function best predicts geomagnetic storms (*Gonzalez et al*, 1989); to compute long term change in open solar flux (*Lockwood et al.,* 1999); to understand transpolar voltage saturation (*Siscoe et al*., 2002) and many others. Recently, the realisation of the problems associated with overfitting has made the *Vasyluinas et al.* formulation, with its single free fit parameter, important and it has been used to study remote sensing of the geoeffectiveness of CMEs (*Owens et al.,* 2018) and to reconstruct the numbers of storms and substorms back to (and including) the Maunder minimum (*Lockwood et al*., 2017, 2018a).

The *Vasyluinas et al.* (1982) theory, and so the applications that employed it, is based on the approximation that all energy fluxes in the solar wind can be neglected except the kinetic energy flux. This has been very successful: for example, by careful analysis to minimise the effect of data gaps, *Lockwood et al*. (2019a) have shown that the correlation with geomagnetic activity is 0.990 ± 0.007 for annual means, 0.897 ± 0.004 for daily means, 0.79 ± 0.03, for 3-hourly means, and 0.7046 ± 0.0004 for one-minute means (the uncertainties being at the 2σ level and the lower correlation for 1-minute data is largely due to the variability around the average of the substorm growth-phase lag). Given that the study of *Ebihara et al*. (2019) shows that the fraction of the solar wind power entering the magnetosphere due to $S_{sw}$ is larger than the fraction of the energy flux that it carries in the solar wind, it is interesting to see if solar wind coupling functions based on energy input into the magnetosphere can be improved by adding a term to allow for solar wind Poynting flux, $S_{sw}$. This is investigated.

## 2. Total Power into the magnetosphere

Poynting flux in a plasma (for which $\mu_r$ = 1) is given by

$$\vec{S} = \vec{E} \times \vec{H} = \frac{\vec{E}\times\vec{B}}{\mu_o} \quad . \qquad (1)$$

Assuming that the solar wind flows radially to Earth (i.e. the solar wind velocity $\vec{V}_{sw}$ is in the $-X$ direction of the Geocentric Solar Magnetospheric (GSM) reference frame) and using ideal Magnetohydronamics (MHD) so that $\vec{E}_{sw} = -\vec{V}_{sw} \times \vec{B}$, where $\vec{B}$ is the interplanetary magnetic field, IMF), the Earthward-directed Poynting flux is

$$S_{sw} = \frac{V_{sw}B_{\perp}^2}{\mu_o} = \frac{V_{sw}B^2}{\mu_o} cos^2\,(\varphi) \qquad (2)$$

where $B_\perp$ is the component of $\vec{B}$ that is transverse to the flow and so $\varphi$ is the angle between $\vec{V}_{sw}$ and $\vec{B}$. Hence the ratio of the kinetic energy flux of particles to the Poynting flux in the solar wind is:

$$\frac{F_{KE}}{S_{sw}} = \frac{\frac{1}{2} m_{sw} N_{sw}\ V_{sw}^3}{\frac{1}{\mu_o} V_{sw} B^2 cos^2(\varphi)} = \frac{M_A^2}{2cos^2(\varphi)} \qquad (3)$$

where $m_{sw}$, $N_{sw}$ and $M_A$ are the mean ion mass, number density and Alfvén Mach number of the solar wind, respectively. Because $M_A$ is typically 10 and $2cos^2(\varphi) \approx 1$, this means that the Poynting flux in the solar wind is of order 1% of the kinetic energy flux.

Poynting's theorem for a plasma (in which there is no displacement current and no permanent magnetism) is

$$-\int_{A_1} \vec{S}.\overrightarrow{dA} = dW_B/dt + \int_{\tau_1} \vec{E}.\vec{J}\, d\tau \qquad (4)$$

where: $W_B$ is the energy density stored in the magnetic field, $W_B = B^2/(2\mu_o)$; $A_1$ is a surface area surrounding a unit volume $\tau_1$; $\vec{E}$, $\vec{J}$, and $\vec{S}$ are the electric field, current density and Poynting flux vectors, $\overrightarrow{da}$ is an element of the surface area $A_1$, $d\tau$ is an element of volume and $\tau_1$ is the volume inside the surface area $A_1$. This means that the negative of the divergence of the Poynting flux equals the sum of the rates at which energy is given to the magnetic field and to the particles (it can be shown that the last term in (4) is the sum of ohmic heating and the work done by/against the so-called $\vec{J} \times \vec{B}$ force). Hence this is a statement of conservation of energy. Regions with current that is aligned with the electric field, and/or with an increasing the magnetic field, are sinks of Poynting flux and conversely regions with a current anti-parallel to the electric field and/or a falling magnetic field, are sources of Poynting flux. In the steady-state discussed by *Cowley* (1991), $dW_B/dt = 0$ and so sinks of Poynting flux are regions where particles are accelerated or heated ($\vec{E}.\vec{J} > 0$) and sources of Poynting flux are where particles are slowed or cooled ($\vec{E}.\vec{J} < 0$).

### *2.1 Poynting Flux for Southward IMF*

Figure 1a illustrates the steady-state case for when the IMF is pointing southward in the GSM frame (in the $-Z_{GSM}$ direction) so the motional electric field of the solar wind in the Earth's frame, $E_{sw}$, points from dawn to dusk in the $+Y_{GSM}$ direction (after *Cowley*, 1991). In this steady-state case, the magnetic field is constant everywhere and so, by Faraday's law, the electric field is curl-free which means it is the same at all points in the noon-midnight plane shown in Figure 1a.

Before entering the magnetosphere, the kinetic energy flux is converted into Poynting flux by the currents $\vec{J}$ that flow in the bow shock, magnetosheath and tail magnetopause in regions where $\vec{J}.\vec{E}_{sw} < 0$, as shown in in Figure 1a. This is added to the pre-existing solar wind Poynting flux. Because $\vec{S}$ is perpendicular to the magnetic field, the draped IMF in the magnetosheath deflects Poynting flux towards the magnetosphere. This occurs irrespective of the IMF orientation. The major extraction of energy by the magnetosphere from the magnetosheath takes place during intervals of southward IMF along the north and south flanks of the tail lobes where the magnetopause currents are from dusk to dawn and so anti-parallel to the electric field (i.e., $\vec{J}.\vec{E} < 0$). Note that these magnetopause currents are orthogonal to the electric field at low magnetospheric latitudes (close to the equatorial plane) and so $\vec{J}.\vec{E} = 0$ and no energy is extracted there.

Figures 1b and 1c show how this is modified by non-steady conditions during the substorm cycle. (after *Lockwood*, 2004). During the growth phase, energy is stored in the increasing field in the tail lobes which are therefore sinks of Poynting flux; in the expansion phase this stored energy is released and deposited in the plasma sheet, ring current and ionosphere, whilst some is propagated down the far tail and returned to the interplanetary medium. Note in these cases, the changing magnetic field in the tail lobes means that the electric field is not curl-free and this induction effect decouples the electric field at high latitudes in the tail lobe from that in the cross-tail current sheet: in the growth/expansion phase the electric field at the cross-tail current sheet is lower/higher than that at the high-latitude magnetopause, respectively, and hence the sink of Poynting flux in the cross-tail current sheet is smaller/greater than the source at the high-latitude magnetopause for growth/expansion phases as tail lobe energy is stored/released.

### *2.2 Poynting Flux for Northward IMF*

Figure 1d shows the situation for persistent northward IMF ($\vec{B}$ points in the $+Z_{GSM}$ direction). This reversal in the field direction compared to figure 1a, 1b and 1c also reverses the electric field in the solar wind and the direction of the currents that flow in the bow shock and magnetosheath. Hence these regions remain sources of Poynting flux. The draping of the interplanetary magnetic field again deflects Poynting flux towards the magnetosphere. The major difference in Figure 1d compared to the southward IMF cases is that the Chapman-Ferraro currents do not reverse in direction when the IMF points northward because the dominant cause of the magnetic shear across the magnetopause is the difference in the magnitudes of the terrestrial and magnetosheath fields. Hence during northward IMF these magnetopause currents may weaken but do not reverse in direction. Another key point is that the geomagnetic tail never disappears because the timescales for that to happen are very much longer than any intervals between the periods of southward IMF that generate open lobe magnetic flux. This means that the dawnward electric field in the solar wind and magnetosheath during northward IMF makes the north and south flanks of the tail sinks of Poynting flux in this case ($\vec{J}.\vec{E} > 0$). The magnetic shear persists across the cross-tail current sheet in which reconnection will still occur, albeit at a reduced rate. This curl in the electric field shows the tail lobe field is reducing, the stored energy being released to both the magnetopause and cross-tail current sheet sinks of Poynting flux. Hence this northward IMF case is an inherently non-steady-state situation and the curl in the electric field associated with the decay of the field in the lobe allows $\vec{E}$ to be in opposite directions in the centre and flanks of the tail. This would persist for as long as the tail does not disappear, and as the timescale for its loss is much, much greater than the duration of any interval of northward IMF, this never occurs. In Figure 1d, the dayside magnetopause currents are drawn as weakened because field has built up in the magnetosheath plasma depletion layer, reducing the magnetic shear across the nose of the magnetosphere. The dayside acts as a sink of Poynting flux partly because the magnetic flux that is lost from the tail accumulates on the dayside, pushing the magnetopause at the nose of the magnetosphere in the $+X_{GSM}$ direction (i.e. the stand-off distance increases). However, the key point in Figure 1d is that energy is no longer extracted at the high-latitude flanks of the long geomagnetic tail. As the effect of the IMF orientation influences all Poynting flux, whether it was generated from the solar wind kinetic energy flux by the bow shock/magnetosheath or was present in the incident solar wind, we should expect the extraction of energy from these two sources by the magnetosphere to share the same IMF orientation dependence.

### *2.3 Adding Solar Wind Poynting Flux*

The efficiency of the transfer of the solar wind kinetic energy flux into the magnetosphere is accounted for in the dimensional analysis of *Vasyluinas et al.* (1982) using the dimensionless transfer function (the fraction of incident flux extracted by the magnetosphere)

$$t_{KE} = M_A^{-2\alpha} \, sin^4(\theta/2) \quad (5)$$

where $\theta$ is the IMF "clock angle" in the GSM frame and $\alpha$ is called the "coupling exponent" and is determined empirically. The clock angle $\theta$ is defined as $\arctan(B_{YM}/B_{ZM})$, where $B_{YM}$ and $B_{ZM}$ are the *Y* and *Z* components of the IMF in the GSM frame of reference. From careful analysis that minimises the effect of data gaps in the interplanetary data, *Lockwood et al.* (2019a) derived a value for $\alpha$ of 0.44 for all averaging timescales between 1 minute and 1 year for the mid-latitude *am* geomagnetic index. The term $sin^4(\theta/2)$ allows for the effect of IMF orientation on the transfer of energy into the magnetosphere. The supporting information for the paper by *Lockwood et al.* (2019b) confirms that this is the optimum IMF orientation factor to use when considering the extraction of the kinetic energy flux, $F_{KE}$, by the magnetosphere. The use of the $sin^4(\theta/2)$ factor to quantify the effect of IMF orientation is optimum because it is a function that is not discontinuous in gradient and which allows for a low rate of reconnection taking place in the dayside magnetopause (thereby generating some open magnetospheric flux) when the IMF points northward (i.e., when the clock angle $\theta$ factor is less than 90 degrees): this has been demonstrated to be the case in a number of studies, most conclusively by observations of $O^+$ ions of ionospheric origin flowing out through the dayside magnetopause (*Chandler et al.,* 1999).

The solar wind Poynting flux needs no conversion at the bow shock/magnetosheath equivalent to that needed for kinetic energy flux but, as discussed in the last section, its ability to enter the magnetosphere should also depend on the same IMF orientation factor. Hence the $M_A^{-2\alpha}$ term is not required whereas an IMF orientation factor is required. From the above arguments we expect the $sin^4(\theta/2)$ term to apply to $S_{sw}$ as it did to the transfer of kinetic energy flux. Later in this paper we will confirm that this is true to a good approximation and hence the transfer function for the Poynting flux can be written as

$$t_S = f_s \, sin^4(\theta/2) \, , \quad (6)$$

where $f_s$ is the fraction of the total solar wind Poynting flux power (incident on the same cross-sectional area as used to compute the total kinetic energy flux) that enters the magnetosphere. This area is $c\pi L_o^2$, where $L_o$ is the stand-off distance of the nose of the magnetosphere and $c$ is an approximately constant area factor that allows for the shape of the dayside magnetosphere. Note that $f_s$ can be greater than unity because the relevant cross-sectional area of the bow shock may be greater than the area $c\pi L_o^2$ relevant to kinetic energy capture by the magnetosphere. It is very difficult to estimate $f_s$ directly without using a global MHD model of the magnetosphere: as mentioned above, the degree to which Poynting flux is deflected towards the magnetosphere depends on the degree of draping of the IMF over the nose of the magnetosphere in the magnetosheath. Furthermore, the $X_{GSE}$ at which the Poynting flux arrives in the tail matters because if it arrives beyond the relevant tail reconnection point (i.e. the one that is closing open flux) it will almost certainly be returned to the interplanetary medium and not contribute to near-Earth phenomena such as the substorm current wedge, and hence to geomagnetic indices such as *ap* and *am* (*Lockwood*, 2013). However, these geometric considerations apply equally, and in the same way, to Poynting flux generated by the bow shock currents and magnetosheath from the kinetic energy density flux of the solar wind. Later in this paper we empirically find optimum values for $f_s$ that range from 0.74 for hourly data to 0.3 for annual means (although the results for timescales great than a day are found not to be statistically significant). From equations (3), (4) and (6), we can compute the ratio of total power inputs into the magnetosphere

$$\psi = \frac{(c\pi L_o^2)\, S_{SW}\, t_S}{(c\pi L_o^2)\, F_{KE}\, t_{KE}} = \frac{2 f_s cos^2(\varphi)}{M_A^{(2-2\alpha)}} \quad (7)$$

using $2 f_s cos^2(\varphi) = 1$ and $\alpha = 0.44$ gives a value of this ratio $M_A^{1.12} \sim M_A$. Hence a typical $M_A$ value of 10 means that the total power entering the magnetosphere due to solar Poynting flux may be of order a tenth of that due to the kinetic power, even though the ratio of the flux densities in the solar wind is of order 1/100.

The formulation of *Vasyliunas et al.* (1982) computes the area ($\pi L_o^2$) by assuming the dayside magnetopause is hemispheric in shape and in equilibrium so area $L_0$ is the equilibrium stand-off distance of the dayside magnetopause. *Lockwood et al.* (2019a) generalised this by using a constant multiplicative area factor $c$ so the area presented to the solar wind is ($c\pi L_o^2$). (This does not add to the number of free fit parameters because it is eventually cancelled by normalising the power input to its overall average value). Assuming the dayside magnetopause is in equilibrium, pressure balance at the dayside magnetopause gives and expression for $L_o$. The power input in to the magnetosphere due to kinetic energy density of the solar wind can then be estimated (see *Vasyluinas et al.,* 1982; *Lockwood et al.* (2019a; 2019b):

$$P_\alpha = (c\pi L_o^2)\, F_{KE}\, t_{KE}$$

$$= \left(\pi c k_2 k_1^2 M_E^{2/3} \mu_0^{-1/3}\right) m_{sw}^{(2/3-\alpha)} N_{sw}^{(2/3-\alpha)} V_{sw}^{(7/3-2\alpha)} B^{2\alpha}\, sin^4(\theta/2) \qquad (8)$$

where $k_1$ and $k_2$ are constants and $M_E$ is the magnetic moment of the Earth which can be computed for a given time using the IGRF-15 Model (*Thébault et al*., 2015). Because the variation of $M_E$ with time is small and approximately linear we can treat the term in brackets as a constant that we can later cancel out by normalising $P_\alpha$ to its average value over the whole period $P_o$ to give $P_\alpha/P_o$.

From (6) we can add the power input due to solar wind Poynting flux to get the combined power input

$$P_{\alpha 1} = P_\alpha \{1 + \psi\} = P_\alpha \left\{1 + 2 f_s cos^2(\varphi)\, M_A^{(2\alpha-2)}\right\} \qquad (9)$$

Equations (7) and (8) can be used to compute the normalised total power into the magnetosphere, from both solar wind kinetic energy and solar wind Poynting flux, $(P_{\alpha 1}/P_{o1})$ from measured solar wind parameters, with two free parameters that need to be derived empirically, $\alpha$ and $f_s$.

Later in this paper we test the $sin^4(\theta/2)$ IMF orientation factor used by employing a factor $G$ defined by:

$$G = \left(\pi c k_2 k_1^2 M_E^{2/3} \mu_0^{-1/3}\right) m_{sw}^{(2/3-\alpha)} N_{sw}^{(2/3-\alpha)} V_{sw}^{(7/3-2\alpha)} B^{2\alpha} \left\{1 + 2 f_s cos^2(\varphi)\, M_A^{(2\alpha-2)}\right\}. \quad (10)$$

To remove the constants and produce a simpler coupling function, we normalise $G$ to its overall mean $G_o$.

As pointed out by *Vasyluinas et al.* (1982), if a plot of the observed ratio $\{am/(G/G_o)\}$ against the observed $sin^4(\theta/2)$ and get a proportional variation with slope $s$, we can write

$$am = s\,(G/G_o)\, sin^4(\theta/2) = (s/G_o)\, P_{\alpha 1} = s_1\, P_{\alpha 1} = s_2\, (P_{\alpha 1}/P_{o1}) \qquad (11)$$

where $$s_2 = s\,(P_o/G_o) \qquad (12)$$

Hence if we find proportionality $s$ is constant and, because $P_o$ and $G_o$ are both constants, $s_2$ is also a constant and hence $(P_{\alpha 1}/P_{o1})$ is a proportional predictor of $am$.

## 3. Observations

The study presented here is based on 12,097,440 1-minute samples of interplanetary parameters observed in the interval 1995-2017 (inclusive), downloaded from the Omni database compiled and maintained by the Space Physics Data Facility at NASA's Goddard Space Flight Center. If any of the parameters used in compiling the power input estimates are missing, this generates a data gap and, as a result, the 1-minute dataset contains 9,530,831 valid samples of both the old and new power input estimates $P_\alpha$ and $P_{\alpha 1}$, an availability of 78.78%. These were then used to generate hourly means. We require the number of minute samples within the hour, $N$, to exceed a limit $N_{lim}$which depends on the parameter in question and the accuracy required. *Lockwood et al.* (2019a) carried out a Monte-Carlo study on data from 22-years in which they introduced data gaps at random into data for hours when all 60 1-minute samples were available, and so derived the $N_{lim}$ values needed to give uncertainties of 2% and 5%. We here adopt the 5% criteria. The autocorrelation time of the solar wind speed, $V_{sw}$, is sufficiently large that a single sample in the hour meets the 5% requirement and so $N_{lim}$ = 1. On the other hand, the IMF orientation clock angle, $\theta$, has a very short autocorrelation time that gives $N_{lim}$= 56. Applying the $N \geq N_{lim}$ criteria (to give uncertainties in hourly means of all parameters in $P_\alpha$ and $P_{\alpha 1}$, that are below 5%), gave 161,627 hourly mean estimates of $P_\alpha$ and $P_{\alpha 1}$, out of a possible 201,624, an availability of 80.16%. (Because the criteria have the same effect on $P_\alpha$ and $P_{\alpha 1}$ this number applies to the power input estimates with and without the solar wind Poynting flux. We then made 3-hourly means only if all three of the hourly means are available. The interval covers 8401 days. We use a mean value for a day if just one 3-hourly mean is available from the day, but employ the piecewise removal of all the geomagnetic am index data for times when the corresponding $P_\alpha$ value is missing (*Finch and Lockwood*, 2007), this gives 8401 daily samples, and availability of 99.69%. The available hourly means were also averaged to give 23 annual means and 309 averages over Carrington synodic solar rotation periods of 27.26 days (654 hours).

The *am* geomagnetic index data are generated and made available by The International Service of Geomagnetic Indices (ISGI), France and collaborating institutes. The stations used to compile the *am* index (*Mayaud*, 1980) are situated at sub-auroral latitudes close to corrected geomagnetic latitude $\Lambda_{CG} = 50°$. There are 15 stations in current use in the northern hemisphere and 10 in the southern. They are grouped into longitude sectors, with 5 such groups in the Northern hemisphere, and 4 in the Southern. The *K* indices for stations in a longitude sector are averaged together and the result is converted into a sector $a_K$ value using the standard K2aK scale. Weighted averages of these sector $a_K$ values are then generated in each hemisphere giving *an* and *as*, the weighting factors accounting for the differences in the longitude extents of the sectors. The index *am* is equal to (*an*+*as*)/2. Note that we here employ all available *am* data up to the end of 2017 and that after the end of 2014 these data are classed as "provisional" which means they have passed initial quality checks and can be used, but not yet been through the final review that defines them as "definitive". *Lockwood et al.* (2018b) have developed a model of the sensitivity of a geomagnetic station to solar wind forcing that can be used to derived the time-of-day/time-if-year response of any geomagnetic index provided that, like *am*, is compiled using an analytic algorithm. This was used by *Lockwood et al.* (2019d) to show that the *am* index response is exceptionally uniform, the standard deviation being just 0.65% of the mean value. Analysis has demonstrated how mid-latitude "range" indices such as *am* respond primarily to the substorm current wedge of substorm expansion phases (see *Finch et al.,* 2008; *Lockwood*, 2013; and the Supporting Information file associated with *Lockwood et al.,* 2019a).

Figure 2 summaries the datasets used. The left hand plots show daily means (averaging timescale $\tau$ = 1 day) and the right hand plots show annual means ($\tau$ = 1 year) for 1995-2017 (inclusive). From top to bottom: (a and b) the *am* index; (c and d) the normalised power input to the magnetosphere ($P_\alpha/P_o$ , where $P_o$ is the mean of $P_\alpha$ for the whole interval); (e and f) the additional Poynting flux term for unit $f_S$, $(\psi/f_S)$ (see equation 9 of text); (g and h) the solar wind Alfvén Mach number, $M_A$ ; and (i and j) the cosine of the angle $\varphi$ between the IMF vector $\vec{B}$ and the solar wind velocity vector $\vec{V}_{sw}$ . It can be seen that $(\psi/f_S)$ is highly anti-correlated with $M_A$.

## 4. Correlation Analysis

### *4.1 Daily and Annual Averaging Timescales*

Figure 3 shows correlations between the *am* index and the power input to the magnetosphere estimate, $P_{\alpha 1}$ which allows for the solar wind Poynting flux (using equation 9). The correlation coefficients are colour-contoured as function of the coupling exponent $\alpha$ (vertical axis) and the transfer fraction for solar wind Poynting flux, $f_S$ (horizontal axis). Note that the left hand edge of these plots is for $f_S = 0$, when $P_{\alpha 1}$ reduces to $P_\alpha$, the power input from solar wind kinetic energy flux alone. The plot on the left is for daily data (averaging timescale, $\tau$ = 1 day) and the right hand plot is for annual data ($\tau$ = 1 year). Both plots show that the $\alpha$ giving peak correlation falls as $f_S$ is increased and that for $f_S > 1$ the peak correlation also falls. The black dot in (a) marks the peak correlation for $\tau$ = 1 day which is at $\alpha$ = 0.36 and $f_S$ = 0.68 for which the correlation is $r$ = 0.924. The cyan diamond is for $f_S$ = 0 and is at $\alpha$ = 0.42 for which the correlation is slightly lower, being $r$ = 0.908. The mauve dot in (b) marks the peak correlation for $\tau$ = 1 year which is at $\alpha$ = 0.32 and $f_S$ = 0.30 for which the correlation is $r$ = 0.980. The green diamond is for $f_S$ = 0 and is at $\alpha$ = 0.34 for which the correlation is very close to the peak value, being $r$ = 0.979.

The question arises if these small increases in correlation, achieved by allowing for the solar wind Poynting flux, are statistically significant. This question is addressed by Figure 4. We here look at the significance of the difference between two correlations by computing the *p*-value of the null hypothesis that they are the same. We use the Meng-Z test for the difference between the correlations between A and B and between A and C which allows for the inter-correlation of B and C (*Meng et al.*, 1992). We test against the AR1 red-noise model by using the effective number of independent data pairs, $N_{eff}$ given by:

$$N_{eff} = N \frac{(1 - a_1)}{(1 + a_1)} \qquad (13)$$

where $N$ is the actual number of data pairs and $a_1$ is the autocorrelation of A at lag 1 (*Wilks*, 1995) .

The top panels of Figure 4 are correlograms showing linear correlation coefficient $r$ as a function of $\alpha$ for the (fixed) best-fit value of $f_S$ - hence they are vertical slices through Figure 3. As in Figure 3, the plot on the left is for $\tau$ = 1 day and the right hand plot is for $\tau$ = 1 year. The line colours are coded in the same way as the points in Figure 3, so the black and mauve lines are for the optimum $f_S$ whereas the cyan and green lines are for $f_S$ = 0. In each case, the $\alpha$ giving peak correlation is marked with a vertical dashed line and the symbols used in Figure 3. The second row of panels uses the same colour scheme to show the $p$-values for the null hypothesis that the correlation is not significantly higher if the solar wind Poynting flux is added (using the optimum combination of $\alpha$ and $f_S$), as computed using the Meng-Z test. Hence at the peak $r$, the black curve in the middle-left panel gives $p$ = 1 but the $p$-value falls either side of this as $r$ falls below its peak value. The horizontal dashed lines mark the level at which this drop in $r$ becomes significant at the $1\sigma$ and $2\sigma$ levels. The cyan line gives the corresponding $p$-value for the difference between $r$ (at a given $\alpha$ and for $f_S$ = 0) and the peak $r$ for the optimum $f_S$ ( = 0.68) and $\alpha$ (= 0.36). These $p$-values are very small and so they have been plotted times 100 in Figure 4c: there is a small peak at the peak $r$ for $f_S$ = 0 at $\alpha$ = 0.42, but even at that peak, the $r$ is significantly lower than for $f_S$ = 0.68 at greater than the $3\sigma$ level. Hence adding the solar wind Poynting flux has significantly improved the correlation for $\tau$ = 1 day, even though the increase in $r$ is actually rather small. This is stressed by the similarity of the scatter plots shown in the bottom left panel for the $\alpha$ of peak correlation for $f_S$ = 0.68 (black points) and for $f_S$ = 0 (cyan points).

The right hand panels are the corresponding plots for $\tau$ = 1 year. In this case the difference for $f_S$ = 0 is not significant at even the $1\sigma$ level at almost all $\alpha$ (green line in the middle-right panel). Hence although significant improvement can be made for 1-day data by adding the solar wind Poynting flux, this is not true for annual data.

### *4.2 Three-Hourly and Hourly Averaging Timescales*

The power input to the magnetosphere has a characteristic variation with both time-of-year *F* and UT because of the Russell-McPherron effect of the angle of rotation between the Geocentric Solar Ecliptic (GSE) and GSM reference frames (*Russell and McPherron*, 1973). In addition, there are dipole tilt effects on the geomagnetic response which means that *am* shows a marked equinoctial variation (*Cliver et al*., 2000). (See *Lockwood et al.* (2016) for a discussion of the proposed mechanisms). Hence there are strong UT dependencies in the solar wind-magnetosphere-ionosphere system that make it desirable to repeat the study for sub-daily averaging timescales. The problem is that there is a variable lag in the *am* response that becomes an increasingly significant factor as the averaging timescale is reduced. For the 3-hourly resolution of the *am* data, neglecting a response lag of 1 hour would mean that a third of the data in the interplanetary average (33%) was not data relevant to the corresponding three-hourly *am* data. By way of comparison, neglecting the same lag for daily averages would mean that (1/24) of the data were not relevant (4.1%), and this falls to 0.15% for averages over Carrington Rotation (CR) intervals and $(4\times10^{-4})$% for annual means. There are high time geomagnetic resolution indices, such as the auroral electrojet indices (*AE*, *AU* and *AL*) and their SuperMAG equivalents (*SME*, *SMU* and *SML*) which have 1-minute resolution that could be used in this context, but these are all generated from northern hemisphere stations only which means that they have a uneven response and give a spurious annual modulation which *am*, on the other hand, reduces to very low levels by optimum choice of stations and the use of longitude sector weightings (*Lockwood et al.,* 2019d).

Rather than move to a different geomagnetic index in order to study sub-daily timescales, we here adopt a different approach that enables us to continue to use the *am* index. The Omni interplanetary dataset has been lagged from the time and place of observation to the nose of Earth's bow shock (*Case and Wild*, 2012). The *am* index responds after a lag *dt* that is the sum of the propagation delay from the nose of the bow shock to the dayside magnetopause, the duration of the substorm growth phase (when energy accumulates in the near-Earth tail lobe magnetic field), the propagation time from current disruption in the near-Earth cross tail current sheet to the nightside auroral oval, and the time between substorm onset and the peak response of *am*.

To derive the optimum lag *dt* we use the *am* data, assigned to the times $t_{am}$ of the mid-points of the 3-hour intervals ($\tau$ = 3 hours) over which each *am* estimate is made (i.e., 1.5 UT for the 0-3 UT interval, 4.5 UT for 3-6 hrs, up to 22.5 UT for 21-24 hrs UT). Three-hourly means of one-minute interplanetary data were generated over intervals between ($t_{am} - \tau/2 - dt$) and ($t_{am} + \tau/2 - dt$) for a given lag *dt*. To construct the 3-hourly means of interplanetary data we make 3 one-hour means and use the criteria derived by *Lockwood et al*, (2019a) of the numbers of samples in the hour that are required to make the uncertainty in the hourly mean of that parameter 5% or less. These three one-hour means are then averaged to give the 3-hourly mean and valid data points require that all three one-hour means are available. The correlations between $P_{\alpha}(t_{am} - dt)$ and $am(t_{am})$ for each *F*-UT bin were then evaluated for lags *dt* which was varied between zero and 300 min (5 hours). For each *F*-UT bin, the lag giving peak correlation was determined and the distribution of *dt* for the 160 bins is shown in Figure 5a. The peak correlations were always above 0.65 and most correlations near the mode *dt* value exceeded 0.95.

To look for an effect of activity level on the lag *dt*, the *am* data were then further sub-divided into quantile ranges. We use the notation that q(n) is the (100×n)% quantile of the distribution of *am* values. Five of the quantile ranges used each contained 20% of the *am* data being: $0 \le am \le q(0.2)$; $q(0.2) < am \le q(0.4)$; $q(0.4) < am \le q(0.6)$; $q(0.6) < am \le q(0.8)$; and $q(0.8) < am \le q(1)$. In addition, we looked at three quantile ranges to study the most active periods covering the top 10% *am* values, $q(0.9) < am \le q(1)$; the top 5%, $q(0.95) < am \le q(1)$; and the top 1%, $q(0.99) < am \le q(1)$. The results are shown in Figure 5b. The peak of the distribution is near $dt = 60$ min for $0 \le am \le q(0.2)$; and rises with *am* to near 70 min for $q(0.2) < am \le q(0.4)$; $q(0.4) < am \le q(0.6)$; and $q(0.6) < am \le q(0.8)$. However, for the largest *am* values the peak *dt* falls again and is in the range 40-50 min.  In all cases there are only a few examples that give $dt < 0$: these are almost non-existent for low *am* levels and occur mostly for high activity levels when persistent solar wind forcing is likely to generate them from chance occurrences in the data series.  A number *F*-UT combinations and *am* activity levels do give *dt* smaller than about 10 minutes which could indicate a directly-driven response in *am*.  For low *am* levels these low-*dt* cases are extremely rare but for largest 1% of *am* values they are almost as common as the occurrences of the peak lag suggesting prior energy input resulting in a very large total energy stored in the near-Earth tail lobes (as for the $dt < 0$ cases).  There are also some cases of very long lags for the top 1% of am data, suggesting that following very large energy input to the magnetosphere a series of substorms are required to return the system to lower stored energy levels.  We here use $dt = 60$ min. as an overall average response lag but note that there is considerable variability in *dt* about this value which will lower correlations and increase noise in statistical studies.

We also make a study on hourly timescales ($\tau = 1$ hour). To do this we take means of the interplanetary data over one-hour intervals centered on times $t_{SW}$ and compare these to *am* values that are interpolated from the observed 3-hourly *am* data to the times ($t_{SW}+dt$), where *dt* is the derived optimum lag of 60 min., the interpolations being carried out using the Piecewise Cubic Hermite Interpolating Polynomial (PCHIP) procedure. The reason these hourly data are instructive is because we find *dt* is of order this timescale.  The *am* index responds primarily to the substorm current wedge during substorm expansion phases, which means that fluctuations in the power input to the magnetosphere on timescales below *dt* are of lesser importance because they are averaged out as energy is accumulated in the geomagnetic tail lobe field during the substorm growth phases, which is the largest contribution to the overall response lag *dt*. Hence by extending the analysis down to $\tau = 1$ hour we are covering the full range of timescales over which  fluctuations in power input to the magntosphere are not averaged out by the accumulation of magnetic energy in the tail lobes.

The left hand panels of Figure 6 are equivalent to Figure 4 for the 3-hourly averaged data obtained for this optimum *dt* and show the behaviour is quite similar to that for the daily data. The correlation is lower (as we would expect because of the variability in actual *dt* values) and the difference between with and without the solar wind Poynting flux is again very small. However the difference in the optimum α for these two cases is larger than for daily data. For hourly data (not shown) the results are very similar, but the scatter is greater and the correlation coefficients lower.

### *4.3 Dependence on Averaging Timescales*

To fill the large gap in timescales between 1 day and 1 year, the right-hand panels in Figure 6 repeat the same analysis for averages over Carrington synodic solar rotation periods of 27.26 days. Table 1 summarises the results for the 4 averaging timescales analysed in Figures 4 and 6 plus the one-hour data obtained by applying interpolation to the *am* observations. The optimum α required is almost constant if Poynting flux is included but increases with decreasing timescales if Poynting flux is omitted. The correlations are always very slightly higher if Poynting flux is included. The *p*-value that the correlations are not significantly different is high for annual and Carrington means, but low for daily, 3-houry and hourly averages.

## 5. The IMF orientation factor

It was argued in Section 2 that the Poynting flux entry into the magnetosphere has a transfer function that depends on IMF orientation with the same $sin^4(\theta/2)$ dependence for both energy derived from the kinetic energy flux of solar wind particles and from the Poynting flux in the solar wind. Where here test this using the method recommended by *Vasyluinas et al.* (1982), namely plotting the observed ratio $\{am/(G/G_o)\}$ against the observed values of $F_\theta = sin^4(\theta/2)$, where $G$ is given by equation (10) and is the total proposed coupling function, (allowing for both the kinetic energy flux and Poynting flux in the solar wind) without the IMF orientation factor, $F_\theta$. $G_o$ is the overall mean of $G$. The results are shown in Figure 7 for τ = 1 day and Figure 8 for τ = 3 fours . The study cannot be performed for τ = 27.26 days nor τ = 1 year because the central limit theorem means that the width of the distribution of $F_\theta$ reduces with increased τ: Figure 2 of *Lockwood et al.* (2017) shows that at τ > 1 day a full range of $F_\theta$ values is not present (and even at τ = 1 day samples of $F_\theta = 0$ and, in particular, of $F_\theta = 1$ are rare). The narrowing of the distribution is such that $F_\theta$ can even be considered to be have a single, constant, value to within an accuracy of ±4.9% (at the 1-σ level) for τ = 1 year and to within ±10.3% for τ = 27 days. Note that the equivalent plots to Figure 8 for magnetospheric power input, without the solar wind Poynting flux term, are shown in the supporting information file attached to the paper by *Lockwood et al.* [2019b]. The daily data (grey dots) in figure 7 and show considerable scatter but a linear trend, emphasised by the black dots that are means in non-overlapping bins of $F_\theta = sin^4(\theta/2)$ that are 0.05 wide (the error bars are plus and minus one standard deviation). A discussed in section (2) this linearity shows that $sin^4(\theta/2)$ remains an appropriate form for the IMF orientation term $F_\theta$ after we have allowed for solar wind Poynting flux. Figure 8 repeats this study for the 3-hourly data. It can be seen that the scatter in the individual data points is greater and the binned means do not agree quite as well with the linear fits as they do for the τ = 1 day case. (The root mean square deviation of the binned means for τ = 3 hours is $\Delta_{rms} = 0.071$ whereas for τ = 1 day $\Delta_{rms} = 0.061$). Figure 9 shows how the $sin^4(\theta/2)$ factor performs relative to other proposed alternatives. The observed ratio $\{am/(G/G_o)\}$ for τ = 1 day in the 24 bins of IMF clock angle $\theta$ that are 7.5° wide are scaled to be between 0 and 1 and then plotted as a function of $\theta$. As discussed by *Lockwood et al.* (2019b), the exponent $n$ in $sin^n(\theta/2)$ has been proposed to be between 2 and 6. In Figure 9 we also test the frequently used function $U(\theta)\cos(\theta)$ where $U(\theta) = 1$ for southward IMF ($\theta$ > 90°) and $U(\theta) = 0$ for northward IMF ($\theta \leq 90°$). It can be seen that $n = 4$ gives the best fit, although the agreement is not quite as good as for the case when solar wind Poynting flux is omitted (see Figure S-12 of the Supporting Information file associated with *Lockwood et al.,* 2019b). A slightly better fit would be obtained using $n$ of 3.8 but this is not statistically significant. The implications of the lack of a major effect of averaging timescale on the analysis of the optimum IMF orientation factor is discussed in the next section.

## 6. Discussion and Conclusions

We have used the near-continuous interplanetary data available for 1995-2017 (inclusive) to investigate if allowing for the solar wind Poynting flux significantly increases the correlation with geomagnetic activity. The use of near continuous solar wind data is very important because *Lockwood et al.* (2019a) have shown that the frequently-made assumption that the effect of data gaps just "averages out" is invalid and causes serious error. In effect, the presence of data gaps introduces noise into correlation studies which gives "overfitting", where the derived coupling function is fitted to the noise in the data and the overfitted refinements to the coupling function have no predictive power as they are not robust. This is important because adding the solar wind Poynting flux to the kinetic energy input adds an additional unknown parameter that has to be fitted empirically and it is the use of too many free fit variables that makes overfitting a particular problem.

It has been shown that a very slight improvement to the energy input to the magnetosphere coupling function can be made by allowing for the Poynting flux that is in the solar wind and adding it to the dominant solar wind energy flux (namely the kinetic energy density flux of the particles). The improvement is small (raising the correlation with the *am* index from 0.908 to 0.924 for daily means and from 0.979 to 0.980 for annual means) and statistically significant for data averaged on daily timescales or less but not significant for annual means or means taken over Carrington Rotation intervals.

We derive empirically a value for the coupling fraction for solar wind Poynting flux, $f_S$, of 0.3 for annual means and so from the $(\psi/f_S)$ variation for annual means shown in the right column of Figure 2, we estimate that $\psi$ varies between 0.016 at sunspot minimum and 0.035 at sunspot maximum, i.e. the contribution of solar wind Poynting flux to total power input varies over the range 1.6-3.5% on these annual timescales. For daily values, the best estimate of $f_S$ is 0.68 which gives a larger range for the contribution of solar wind Poynting flux to total power input of 0-47%. The distribution of daily $\psi$ varies is lognormal in form with a mode value of 0.060 and 10 and 90 percentile values of 0.030 and 0.119. Hence the most common percentage of total power into the magnetosphere that arises from solar wind Poynting flux on daily averaging timescales is 6% and of all daily values, 80% lie in the range 3-12%.

Note that the frequently-used epsilon coupling function, $\varepsilon$, is based on the incorrect assumption that the relevant energy flux in the solar wind is the Poynting flux and, although this can be made consistent with the energy coupling function based on the dominant solar wind kinetic energy flux, $P_\alpha$ , using an extreme value of the coupling exponent $\alpha$, this does not give as good agreement with geomagnetic indices as does the optimum value of $\alpha$. This is why $\varepsilon$ performs considerably less well than $P_\alpha$ (and hence $P_{\alpha1}$) on all averaging timescale (see *Finch and Lockwood*, 2007).

Lastly, we noted in the last section that the optimum form of the IMF orientation factor $F_\theta$ is, as expected, the same for the Poynting flux that is generated at the bow shock (or in the magnetosheath) from the solar wind kinetic energy density and for the Poynting flux which was present in the solar wind prior to it hitting the bow shock. In addition, the optimum form of the IMF orientation factor at $F_\theta$ was found not change with timescale, although the noise in the analysis is greater at low τ (largely caused by the increased importance of the variable response lag of $am$) and at large τ the distribution of $F_\theta$ narrows to an almost constant value, as shown in Figure 2 of *Lockwood et al*., 2017). At high time resolution (1 minute) the distribution of the optmum $F_\theta$ has an unexpected form with a great many samples in a narrow spike at $F_\theta$ = 0 (see explanation in Figure 9 of *Lockwood et al.,* 2019b). Figure 8 of *Lockwood et al.* (2019b) and Figure 4 of *Lockwood et al.* (2019c) show that it is the variability and distribution of $F_\theta$ which sets the distribution of power input to the magnetosphere at high resolutions (1-minute) and that averaging causes these distributions to evolve towards a log-normal form at τ = 1 day which matches closely that in the *am* and *ap* geomagnetic indices. For timescales τ up to the response lag *dt* ~ 60 min., the geomagnetic response closely follows the average of the IMF orientation factor because during substorm growth phases the effects the storage of energy integrate, and hence average out, the effects of the rapid fluctuations in power input to the magnetosphere. Hence it is significant that $F_\theta = sin^4(\theta/2)$ factor works well right down to τ of 1 hour. At τ above 1 day, the $F_\theta$ factor becomes increasing less important as it tends towards its quasi-constant value at τ = 1 year.

**Acknowledgements**. The author is grateful to the staff of Space Physics Data Facility, NASA/Goddard Space Flight Center, who prepared and made available the OMNI2 dataset used. The data were downloaded from http://omniweb.gsfc.nasa.gov/ow.html. They are also grateful to the staff of 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. The work presented in this paper is supported by STFC consolidated grant number ST/M000885/1, and by the SWIGS NERC Directed Highlight Topic Grant number NE/P016928/1/.

## References

Case, N. A., and J. A. Wild (2012) A statistical comparison of solar wind propagation delays derived from multispacecraft techniques, *J. Geophys. Res*., **117**, A02101, doi:10.1029/2011JA016946.

Chandler, M.O., S.A. Fuselier, M. Lockwood and T.E. Moore (1999) Evidence of component magnetic merging equatorward of the cusp, *J. Geophys. Res*., **104** , 22623-22648, doi: 10.1029/1999JA900175

Cowley, S. W. H. (1991) Acceleration and heating of space plasmas - Basic concepts, *Annales Geophysicae*, 9, 176-187

Cliver, E. W., Kamide, Y., and Ling, A. G. (2000) Mountains versus valleys: Semiannual variation of geomagnetic activity, *J. Geophys. Res.*, **105**(A2), 2413– 2424, doi:10.1029/1999JA900439.

Ebihara, Y., Tanaka, T., & Kamiyoshikawa, N. (2019). New diagnosis for energy flow from solar wind to ionosphere during substorm: Global MHD simulation. Jour*nal of Geophysical Research: Space Physics*, **124**, 360–378. doi: 10.1029/2018JA026177

Finch, I. D., & Lockwood, M. (2007). Solar wind-magnetosphere coupling functions on timescales of 1 day to 1 year, *Annales Geophysicae*, **25** (2), 495–506. doi: 10.5194/angeo-25-495-2007

Finch, I.D., M. Lockwood, A. P. Rouillard (2008) The effects of solar wind magnetosphere coupling recorded at different geomagnetic latitudes: separation of directly-driven and storage/release systems, *Geophys. Res. Lett*., **35**, L21105, doi:10.1029/2008GL035399

Gonzalez, W. D., B.T. Tsurutani, A.L.C. Gonzalez, E.J. Smith, F. Tang, and S.-I. Akasofu (1989) Solar wind-magnetosphere coupling during intense magnetic storms (1978-1979), J*ournal of Geophysical Research*, **94** (A7), 8835-8851. doi: 10.1029/JA094iA07p08835

Koskinen, H. E. J., and E. Tanskanen (2002) Magnetospheric energy budget and the epsilon parameter, *J. Geophys. Res.*, **107** (A11), 1415. doi:10.1029/2002JA009283, 2002

Lockwood, M. (2004) Solar Outputs, their variations and their effects of Earth in *The Sun, Solar Analogs and the Climate*, Proc. Saas-Fee Advanced Course, 34 by J.D. Haigh, M. Lockwood and M.S. Giampapa, eds. I. Rüedi, M. Güdel, and W. Schmutz, pp107-304, Springer, ISBN: 3-540-23856-5, 2004

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

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

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

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

Lockwood, M.J. Owens, L.A. Barnard, C.J. Scott, C.E. Watt and S. Bentley (2018a) Space Climate and Space Weather over the past 400 years: 2. Proxy indicators of geomagnetic storm and substorm occurrence, *J. Space Weather Space Clim*., **8**, A12, doi: 10.1051/swsc/2017048

Lockwood, M., I.D. Finch, A. Chambodut, L.A. Barnard, M.J. Owens, and E. Clarke (2018b) A homogeneous aa index: 2. hemispheric asymmetries and the equinoctial variation*, J. Space Weather Space Clim.,* **8**, A58, doi: 10.1051/swsc/2018044

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**. doi: 10.1029/2018SW001856

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

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

Meng, X.-I., Rosenthal, R., & Rubin, D. B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin,* **111** (1), 172–175. doi: 10.1037//0033-2909.111.1.172

Owens, M.J., M. Lockwood and L.A. Barnard (2018) Ion charge states and potential geoeffectiveness: The role of coronal spectroscopy for space-weather forecasting, *Space Weather*, **16**, doi: 10.1029/2018SW001855, (2018)

Palmroth, M., Pulkkinen, T. I., Janhunen, P., & Wu, C. C. (2003) Stormtime energy transfer in global MHD simulation. *Journal of Geophysical Research: Space Physics*, **108**(A1), doi.org/10.1029/2002JA009446

Russell, C. T., and McPherron, R. L. (1973) Semiannual variation of geomagnetic activity, *J. Geophys. Res*., **78**(1), 92– 108, doi:10.1029/JA078i001p00092.

Siscoe, G. L., G. M. Erickson, B. U. Ö. Sonnerup, N. C. Maynard, J. A. Schoendorf, K. D. Siebert, D. R. Weimer, W. W. White, and G. R. Wilson (2002) Hill model of transpolar potential saturation: Comparisons with MHD simulations, J. Geophys. Res., 107(A6), doi:10.1029/2001JA000109.

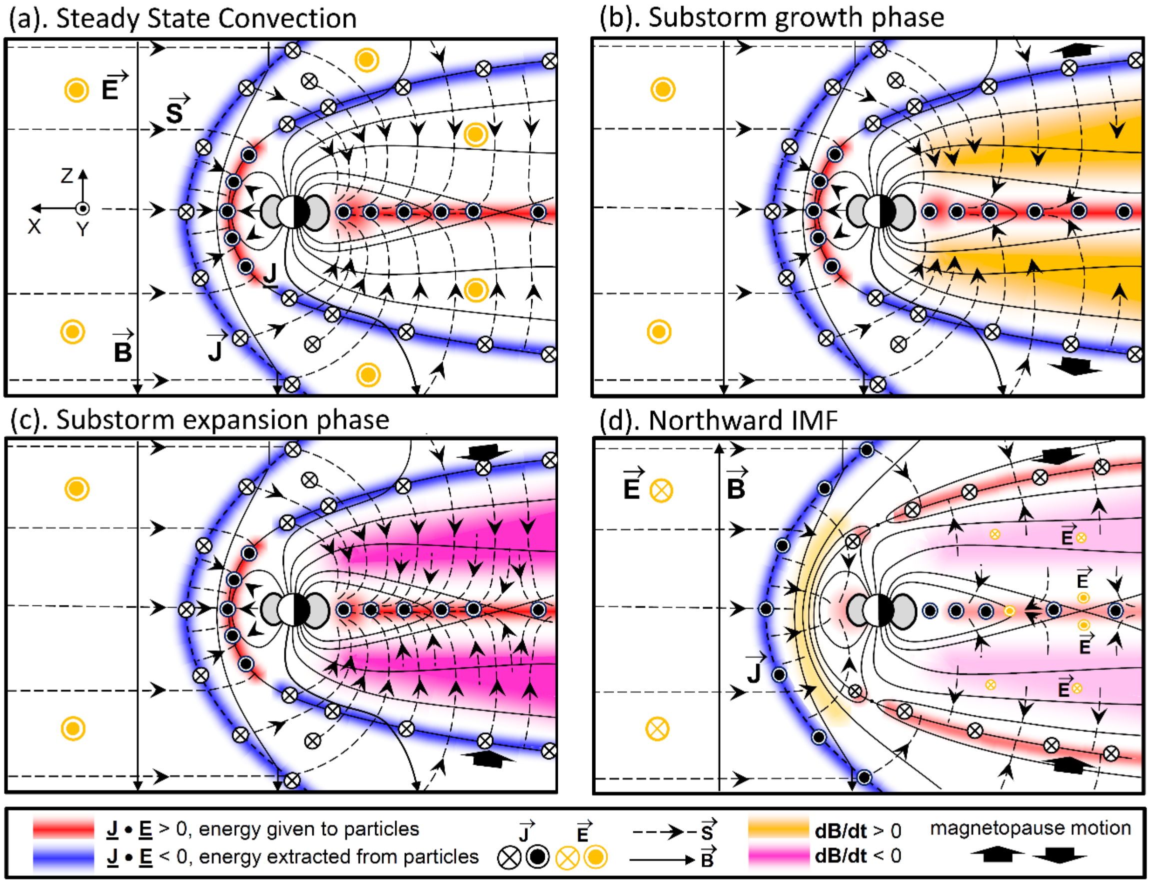
Thébault E, C.C. Finlay, C.D. Beggan, P. Alken, J. Aubert et al. (2015) International Geomagnetic Reference Field: the 12th generation, *Earth Planet Space* **67**, 79. doi: 10.1186/s40623-015-0228-9.

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

Wilks, D.S. (1995) Statistical methods in the atmospheric sciences, International Geophysical Series, Volume 59, ISBN-10: 0127519653. ISBN-13: 978-0127519654, Academic Press, London.

**Table 1.** Summary of correlations between estimated power input to the magnetosphere and the *am* geomagnetic index for 1995-2017 (inclusive) and averaging timescales, $\tau$, of 1 year, 1 (synodic) Carrington rotation (CR = 27.28 days), 1 day, 3 hours and 1 hour. (For $\tau = 1$ hour, 3-hourly am index data are interpolated using PCHIP interpolation and allowing for the derived optimum 60-minute response lag of *am* with respect to the solar wind data). For each $\tau$, analysis is done with and without including solar wind Poynting flux. In each case, the number of samples *N*, the best fit factor $f_S$ (by definition $f_S$ is zero if *S* is not included), the best-fit coupling exponent $\alpha$, and the peak correlation, $r_p$, are given. The last column gives the *p*-value that the difference in the peak correlation without *S* is the same as that with *S*. The number of the figure showing the data and the line colour used in that figure are given for each case.

| $\tau$ | with *S*? | figure # | line colour | *N* | $f_S$ | $\alpha$ | peak *r* | *p*-value of difference in peak *r* |
|---|---|---|---|---|---|---|---|---|
| 1 year | N | 4 | green | 23 | 0 | 0.34 | 0.978 | 0.65 |
| | Y | | mauve | | 0.30 | 0.32 | 0.979 | |
| 1 CR | N | 6 | green | 309 | 0 | 0.37 | 0.9315 | 0.98 |
| | Y | | mauve | | 0.65 | 0.35 | 0.9316 | |
| 1 day | N | 4 | cyan | 8401 | 0 | 0.42 | 0.908 | 0.001 |
| | Y | | black | | 0.68 | 0.36 | 0.924 | |
| 3 hours | N | 6 | cyan | 67,208 | 0 | 0.43 | 0.842 | 0.03 |
| | Y | | black | | 0.69 | 0.39 | 0.843 | |
| 1 hour | N | - | - | 161,627 | 0 | 0.49 | 0.7866 | 0.02 |
| | Y | | - | | 0.74 | 0.38 | 0.7886 | |




673 **Figure 1**. Schematic noon-midnight cross-sections of the magnetosphere for (a) steady-state
674 magnetospheric convection during southward IMF; (b) a substorm growth phase; (c) a
675 substorm expansion phase and (d) persistent northward IMF. Part (a) is after *Cowley* (1991),
676 parts (b)-(d) after *Lockwood* (2004). See text and key at the base of the figure for explanation
677 of coloured regions and symbols.

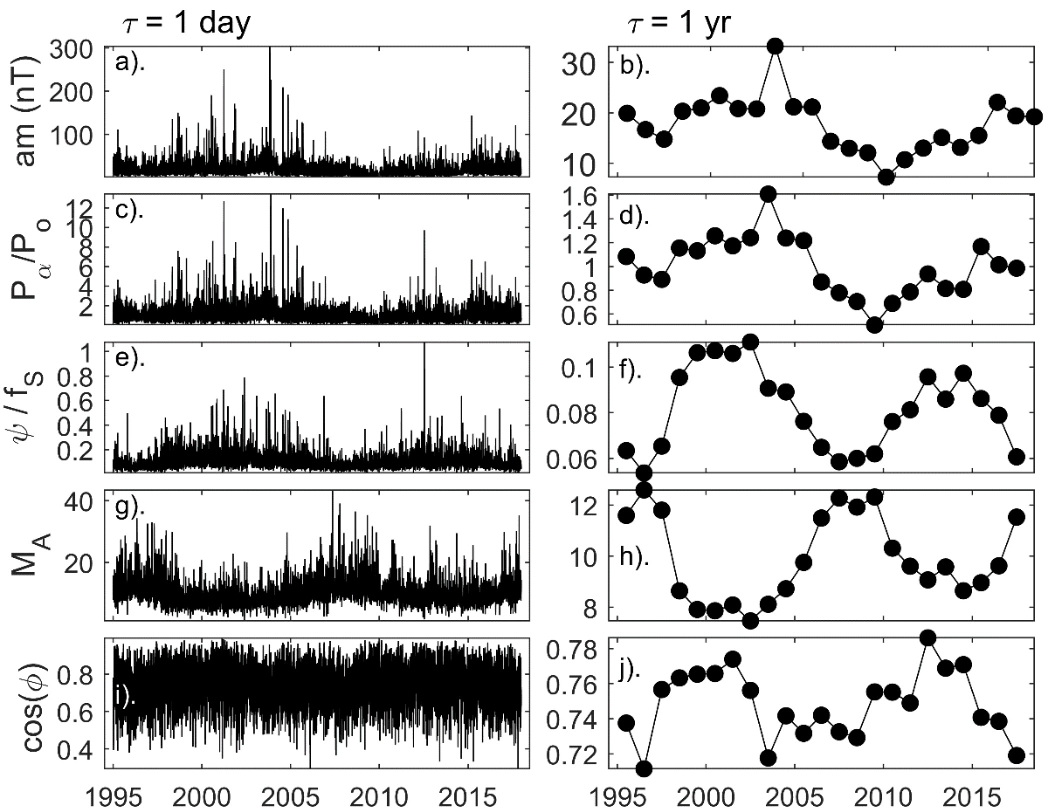



**Figure 2.** Daily (left) and annual (right) means of time-series data used in this paper. From top to bottom: (a and b) the *am* geomagnetic index; (c and d) the power input to the magnetosphere based on solar wind kinetic energy flux normalised to its overall mean, $P_\alpha/P_o$; (e and f) a factor relating to solar wind Poynting flux entering the magnetosphere $\psi/f_S$ (see text for details and equation (6) for definition); (g and h) the solar wind Alfvén Mach number $M_A$; and (i and j) $\cos(\varphi) = (B_z^2 + B_y^2)^{1/2}/B$ where $B_z$ and $B_z$ are two orthogonal components of the IMF $B$ that are perpendicular to the solar wind flow direction.

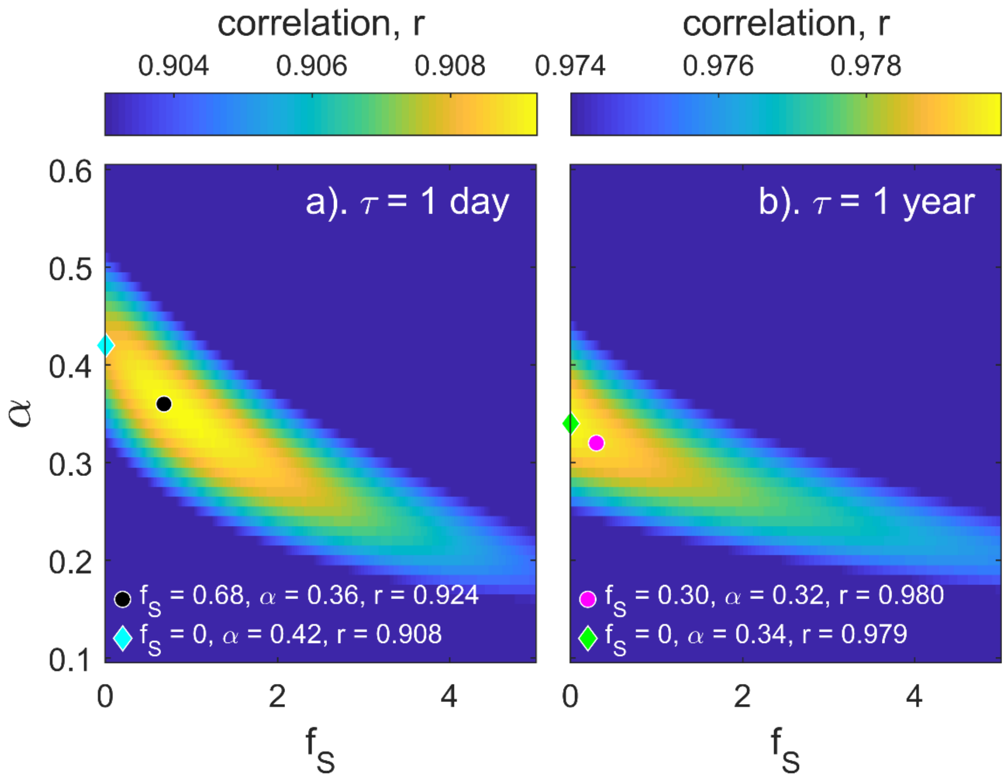



**Figure 3.** Linear correlation coefficients $r$ between the *am* index and $P_{\alpha 1}$ (computed using equations 7 and 8), colour contoured as a function of the fit parameters $\alpha$ (along the vertical axis) and $f_S$ (along the horizontal axis), for averaging timescale $\tau$ of (a) 1 day and (b) 1 year. The black dot marks the peak correlation in (a) and the cyan diamond the peak for $f_S = 0$, (for which $P_{\alpha 1} = P_{\alpha}$ as the solar wind Poynting flux is not included). The corresponding points for annual mean data are shown by a cyan dot and a green diamond in (b). The values of $\alpha$ and $f_S$ at these points and the resulting correlation $r$ are given at the bottom of each panel.

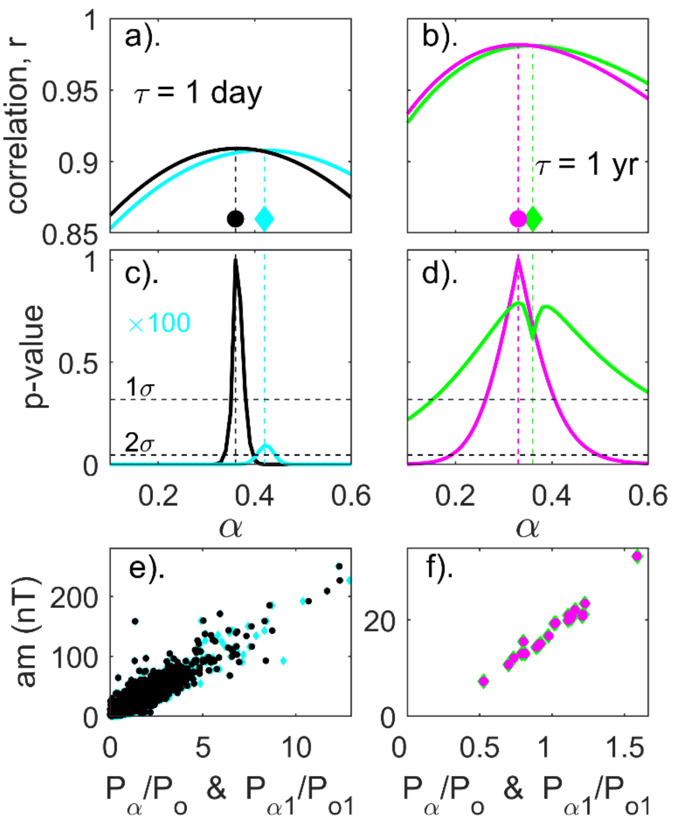



**Figure 4.** Comparison of the $r - \alpha$ correlograms for the optimum $f_S$ and for $f_S = 0$ . The left hand panels are for 1-day means ($\tau$ = 1 day) and the right hand panels are for annual means ($\tau$ = 1 yr). Lines are coloured using the same colour scheme as the points in Figure 3. The top panels show the linear correlation coefficient $r$ as a function of the coupling exponent $\alpha$: the black line for $\tau$ = 1 day is for the optimum $f_S$ of 0.68, the cyan line is for $f_S$ = 0; the mauve line for $\tau$ = 1 yr is for the optimum $f_S$ of 0.30 and the green line is for $f_S$ = 0. The vertical dashed lines mark the peak correlations. The middle panels show the corresponding $p$-values for the null hypothesis that the correlation is as good as for the peak value at the optimum $f_S$: the $1\sigma$ and the $2\sigma$ levels are shown by horizontal dashed black lines. Note that the $p$ value for $\tau$ = 1 day and $f_S$ = 0 (the cyan line) has been multiplied by 100 and that even at the optimum $\alpha$, the correlation is significantly lower than the peak for $f_S$ = 0.68 at more than the $3\sigma$ level. On the other hand, for $\tau$ = 1 yr and $f_S$ = 0 (the green line) the correlation is not significantly lower at even the $1\sigma$ level at almost all values of $\alpha$. The bottom panels shows the scatter plots for the optimum $\alpha$ values for the optimum and zero $f_S$ cases, using the same colour scheme.

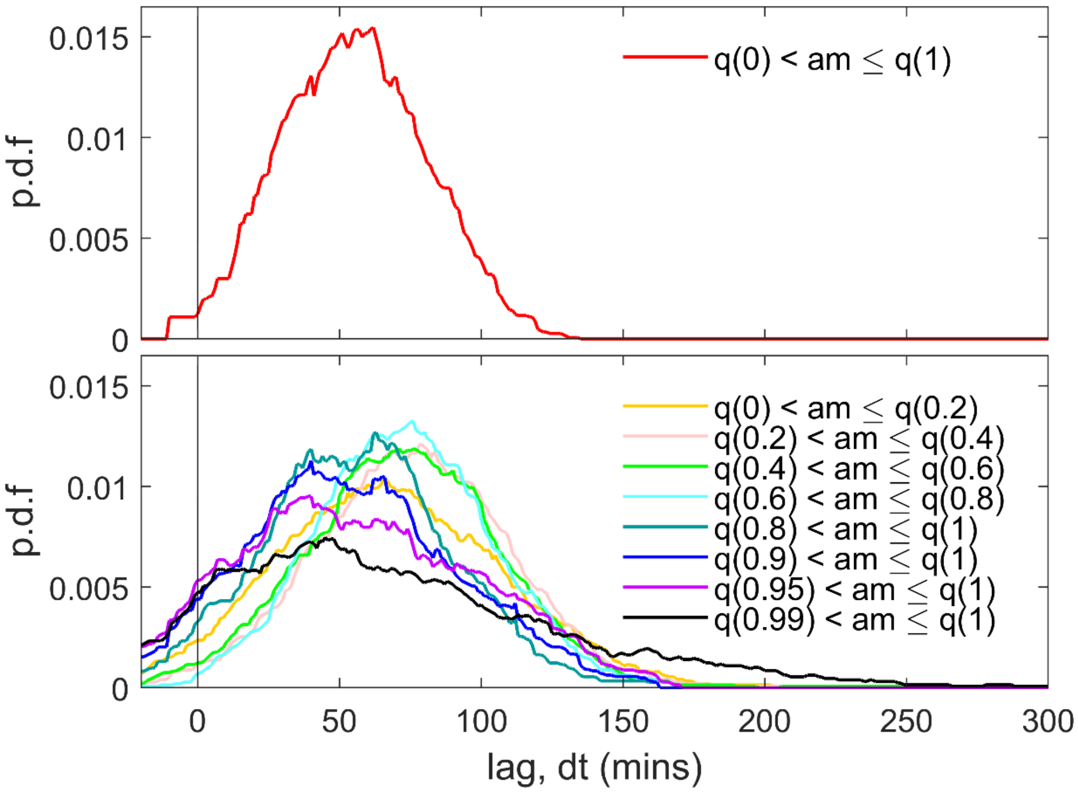



**Figure 5.** Analysis of response lags for the *am* index to power input to the magnetosphere for 1995-2017, inclusive. (Top). The 3-hourly *am* data have been sorted into 20 time-of year bins ($F$) and 8 UT bins. One-minute $P_\alpha$ data have then been averaged into 3-hourly intervals, centred on the mid-points of the *am* data intervals, minus a response lag $dt$ that was varied between zero and 5 hours in steps of one minute. At each $dt$ the correlation between $P_\alpha(t-dt)$ and $am(t)$ was evaluated and the red line shows the distribution of the $dt$ giving peak correlation for the 160 $F$-UT bins. (Bottom) The same analysis but the *am* data have been further subdivided into 8 quantile ranges: the first 5 of these quantile ranges each contain 20% of the data, for example the first is for *am* between zero and its 0.2 quantile, $q(0.2)$. The last three study the optimum lags for the largest *am* values, studying the top 10%, 5% and 1% of all *am* values.

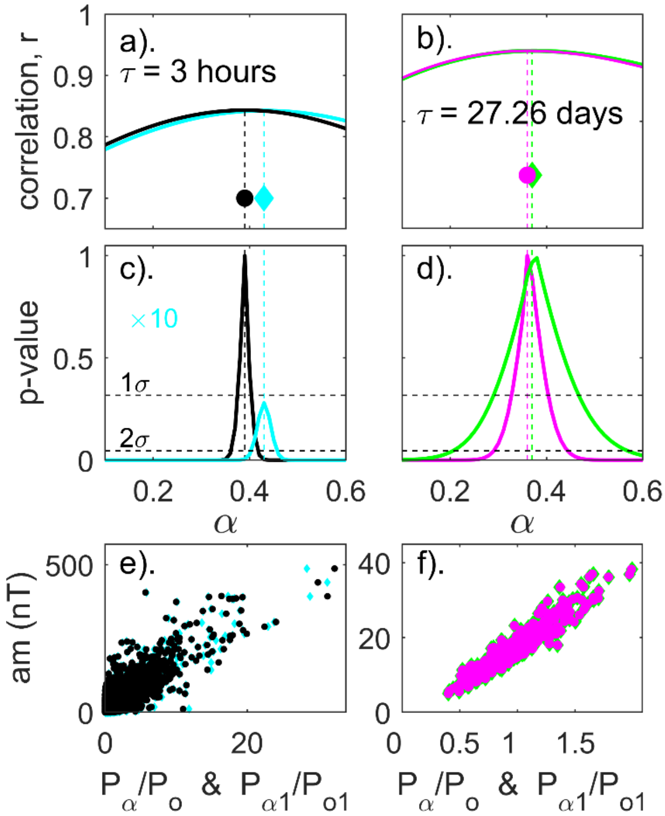



**Figure 6**. The same as Figure 4 but for (left hand panels) 3-hourly data and (right hand panels) averages over Carrington rotation intervals (27.26 days). For the 3-hourly timescale, the solar wind data are averaged over intervals a response delay of $dt$ = 60 min before the three hours over which the *am* values are evaluated (see Figure 5).

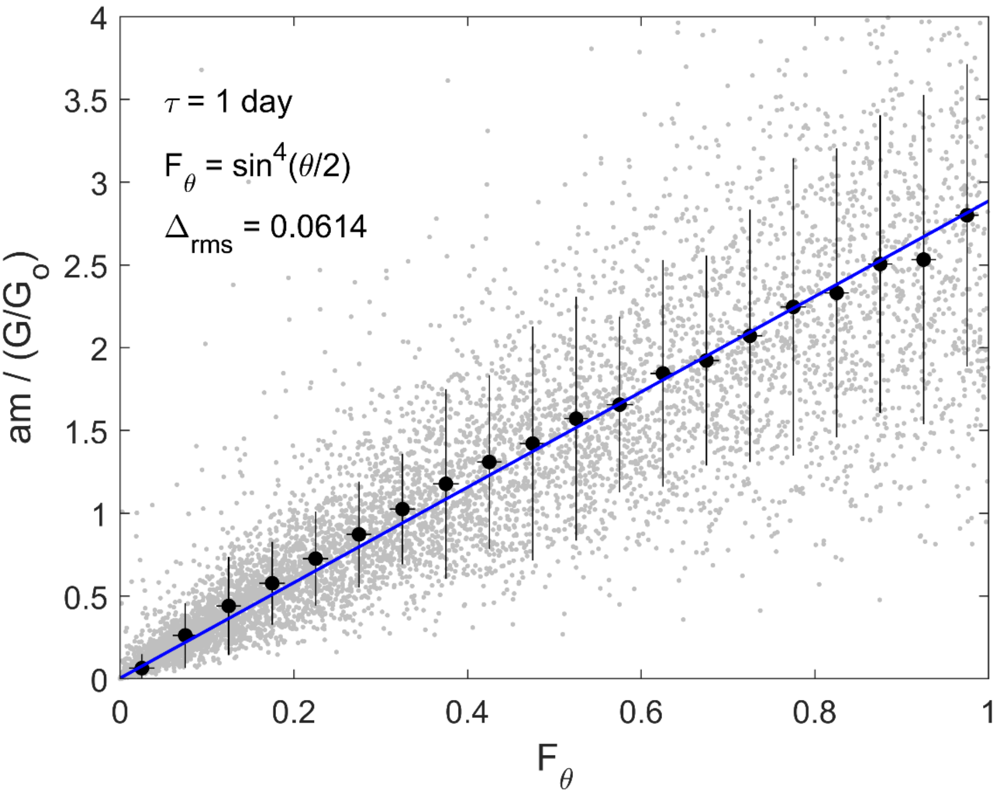



**Figure 7.** Test of the IMF orientation factor in the transfer function $F_\theta = sin^4(\theta/2)$ , where $\theta$ is the IMF clock angle, $\theta = \arctan(B_{YM}/B_{ZM})$, where $B_{YM}$ and $B_{ZM}$ are the Y and Z components of the IMF in the GSM frame of reference. These data are for averaging timescale τ = 1 day. As shown by equations (11) and (12) of the text, the proportionality of the two confirms $sin^4(\theta/2)$ is a good IMF orientation factor in the transfer function employed in $P_{\alpha 1}$. The grey dots are for daily means and the back dots are means for non-overlapping bins of $F_\theta$ of width 0.05 and both horizontal and vertical error bars are plus an minus one standard deviation. The blue line is the best-fit linear regression to the binned data.

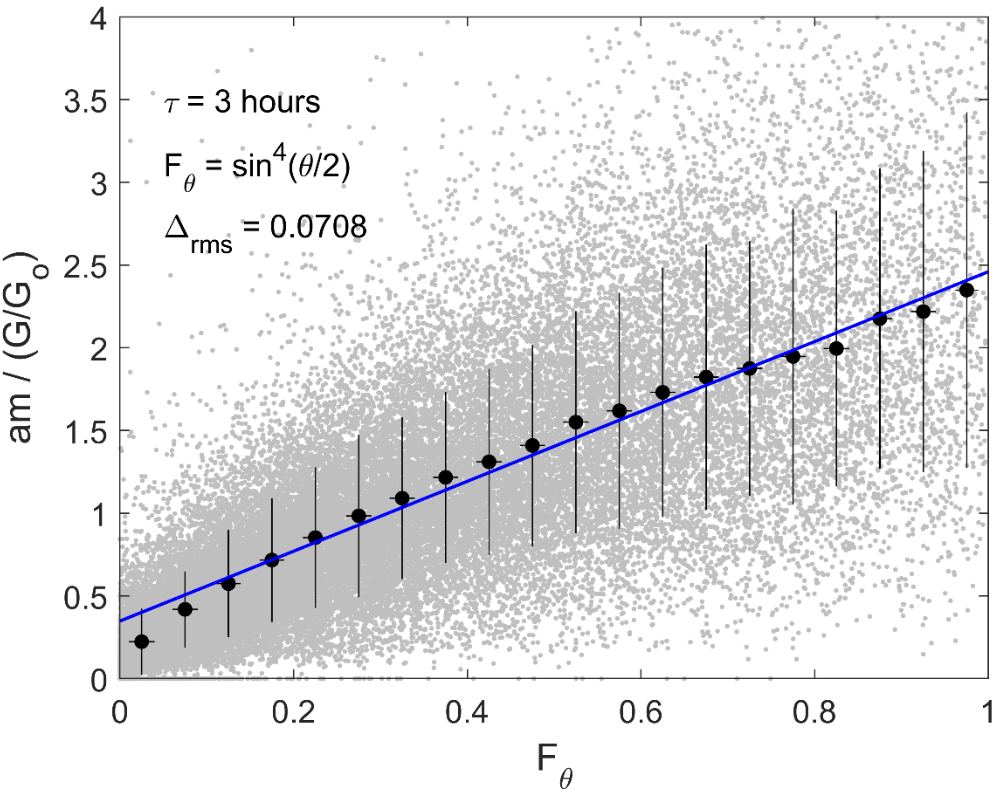



**Figure 8.** The same as Figure 7 for an averaging timescale τ = 3 hours and allowing for a 60-minute response lag of *am* to the solar wind power input data.

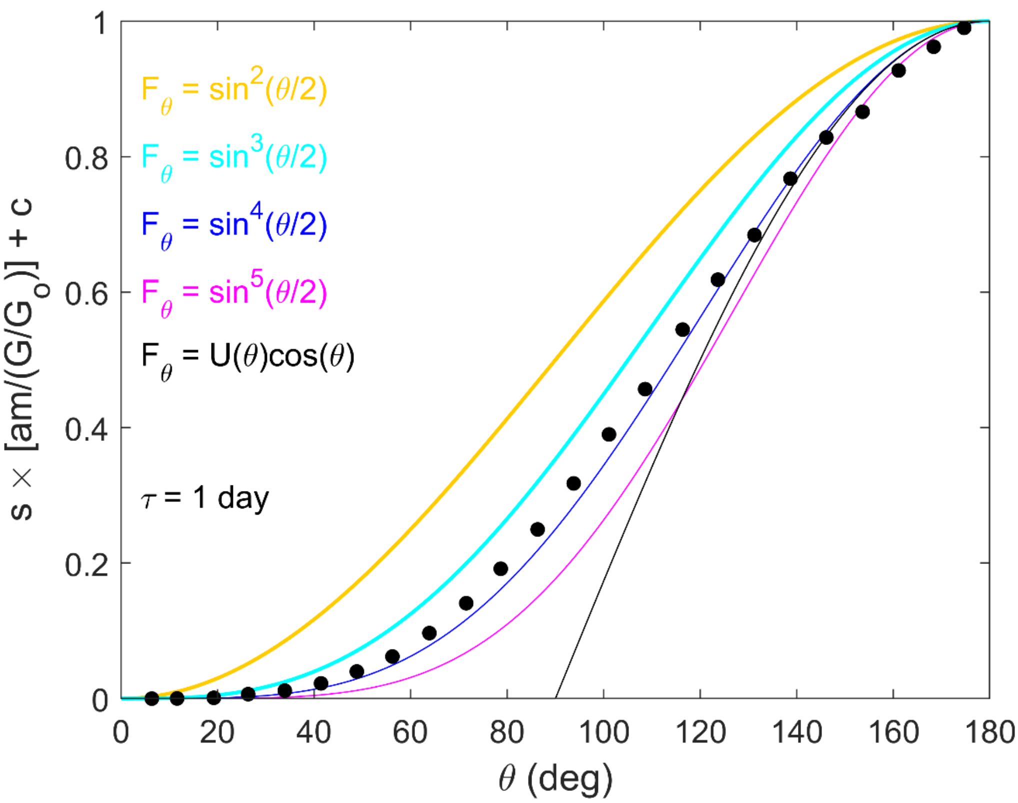



**Figure 9.** Comparison of the observed IMF orientation transfer factor with various forms suggested and/or used in the literature. The black dots are the means of the observed $\theta$ and the scaled ratio $\{am/(G/G_o)\}$ for τ = 1 day and bins of $\theta$ that are 7.5° wide. The coloured lines show 5 proposed analytic forms: (orange) $sin^2(\theta/2)$; (cyan) $sin^3(\theta/2)$; (blue) $sin^4(\theta/2)$; (mauve) $sin^5(\theta/2)$ and $U(\theta)\cos(\theta)$ where $U(\theta) = 0$ for $\theta \leq 90°$ and $U(\theta) = -1$ for $\theta > 90°$.

## Plain English Summary

Space weather is caused by energy extracted from the solar wind by the magnetosphere, the volume of space surrounding the Earth that is dominated by Earth's magnetic field. That energy arrives in two main forms in the solar wind: the kinetic energy of the particle flow and an electromagnetic energy flux. The most successful predictors of space disturbances have considered only the kinetic energy flux. The paper shows typically 10% of the power input comes from the electromagnetic energy flux in the solar wind and allowing for this can make a small, but significant, improvement to our ability to predict terrestrial space weather disturbances.