Solar wind-magnetosphere coupling functions on timescales of 1 day to 1 year

I. Finch, M. Lockwood

Ann. Geophys., 25, 495–506, 2007
www.ann-geophys.net/25/495/2007/

Annales Geophysicae

Solar wind-magnetosphere coupling functions on timescales of 1 day to 1 year

I. Finch[1,*] **and M. Lockwood**[1,*]

[1]Space Science and Technology Department, Rutherford Appleton Laboratory, Oxfordshire, UK
[*]also at: School of Physics and Astronomy, University of Southampton, Southampton, UK

Received: 23 June 2006 – Revised: 22 January 2007 – Accepted: 25 January 2007 – Published: 8 March 2007

Abstract. There are no direct observational methods for determining the total rate at which energy is extracted from the solar wind by the magnetosphere. In the absence of such a direct measurement, alternative means of estimating the energy available to drive the magnetospheric system have been developed using different ionospheric and magnetospheric indices as proxies for energy consumption and dissipation and thus the input. The so-called coupling functions are constructed from the parameters of the interplanetary medium, as either theoretical or empirical estimates of energy transfer, and the effectiveness of these coupling functions has been evaluated in terms of their correlation with the chosen index. A number of coupling functions have been studied in the past with various criteria governing event selection and timescale. The present paper contains an exhaustive survey of the correlation between geomagnetic activity and the near-Earth solar wind and two of the planetary indices at a wide variety of timescales. Various combinations of interplanetary parameters are evaluated with careful allowance for the effects of data gaps in the interplanetary data. We show that the theoretical coupling, P_α, function first proposed by Vasyliunas et al. is superior at all timescales from 1-day to 1-year.

Keywords. Geomagnetism and paleomagnetism (Time variations, diurnal to secular) – Interplanetary physics (Interplanetary magnetic fields; Solar wind plasma) – Magnetospheric physics (Solar wind-magnetosphere interactions)

1 Introduction

Studies of correlations between individual parameters (and combinations of parameters) of the solar wind and geophysical indices have been conducted since in-situ measurements of the solar wind first became available. An early study was conducted by Snyder et al. (1963) who found a correlation between solar wind velocity and the K_p index at daily timescales. Further correlation studies have been conducted at a number of timescales, from minutes (Meng et al., 1973; Burton et al., 1975; Baker et al., 1981), to years (Crooker et al., 1977; Stamper et al., 1999). Studies have used a variety of geomagnetic indices, the previously mentioned studies using *AE*, D_{ST}, *AE*, *Ap* and *aa*, respectively.

A review of the coupling functions that have been previously investigated has been given by Baker (1986), while a more detailed analysis of the relationship between many of these functions was presented by Gonzalez (1990). We here use approximately the same set as that selected by Stamper et al. (1999) (but have added two additional coupling functions, $|\boldsymbol{B}|$ and $v_{sw}^2\,|\boldsymbol{B}|$, where $|\boldsymbol{B}|$ is the IMF magnitude and v_{sw} is the solar wind speed). The origin and physical meaning, if any, of these coupling functions is examined in more detail in Sect. 3 of this paper. In Table 1 we compare our correlation results at the averaging timescale of one year, as used by Stamper et al., and find that our results are in line with theirs. We here extend the work of Stamper et al. by systematically studying the dependence on timescale.

We expect a study at a particular timescale to be most sensitive to mechanisms and events with appropriate characteristic timescales, e.g. a study with weekly resolution would be sensitive to recurrent storms due to solar rotation but will not detect features due to minute-level turbulence. We are not aware of any other studies which have been made over such a wide range of timescales or that have looked in detail at coupling function correlations at timescales between 1-day and 1-year.

Baker (1986) discusses the types of phenomena that are revealed by correlation studies at a given timescale. According to his survey the highest temporal resolution considered here (1 day) will give access to storm timescales, and although this temporal resolution can reveal gross coupling relationships it is insufficient to study directly the physical

Correspondence to: I. Finch
(i.d.finch@rl.ac.uk)

Published by Copernicus GmbH on behalf of the European Geosciences Union.

Table 1. Correlations of Annual Means of Interplanetary Coupling Functions with Geophysical Indices *aa* and *am*. The interplanetary parameters are: B_S, the southward IMF component (in the GSM frame); $|\boldsymbol{B}|$, the magnitude of the IMF; v_{sw}, the solar wind velocity; P_{sw}, the solar wind dynamic pressure $=mN_{sw}v_{sw}^2$, where m is the mean ion mass and N_{sw} is the solar wind plasma density; ε, Akasofu's epsilon parameter ($\propto v_{sw}|\boldsymbol{B}|^2 \sin^4(\theta/2)$), where θ is the IMF clock angle in the GSM frame; and P_α, Vasyliunas' parameter described in Sect. 3.

Interplanetary coupling function	Correlation Coefficient, r (*aa*) Stamper et al. (1999)	Correlation coefficient, r (*aa*)	Correlation coefficient, r (*am*)	r am_w and OMNI 2
$<B_S>$	0.519	0.43	0.488	0.392
$<\|\boldsymbol{B}\|>$		0.622	0.675	0.613
$<v_{sw}^2>$	0.663	0.745	0.697	0.756
$<P_{sw}>=<N_{sw}v_{sw}^2>$	0.743	0.521	0.533	0.532
$<v_{sw}^2 B_s>$	0.869	0.856	0.879	0.83
$<\varepsilon>=<v_{sw}B^2\sin^4(\theta/2)>$	0.781	0.742	0.786	0.738
$<P_m>$	0.938	0.942	0.954	0.958

mechanisms producing that coupling. We note that some important timescales, such as ring current growth and decay and radiation belt diffusion were not included in Baker's analysis but are within the range of timescales we study here.

2 Data used

We have selected the related planetary geomagnetic activity indices *aa* and *am* to correlate with the solar wind coupling functions. Both indices are available continuously since the International Polar Year (IPY) in 1957, and the *aa* index is available continuously since 1868. Also described is the available solar wind data, which is only available from spacecraft located outside the magnetopause and thus since the beginning of the space age. This data must be treated with caution as it is discontinuous and subject to some intercalibration issues. We demonstrate, and show how to mitigate, the large errors that result from naïve handling of the solar wind data.

2.1 Geophysical indices

The *am* index is a planetary geophysical activity range-based index constructed using the K data from a number of mid-latitude magnetometer stations. Mayaud (1980), the originator of the am index, describes it as "the average 3-h range observed, in each hemisphere, within a band close to a 50° corrected geomagnetic latitude". In practice the index is constructed from the K index values of a number of longitudinally-separated geomagnetic stations which are not perfectly located at 50° geomagnetic latitude. The K indices are derived from the difference between the maximum and minimum value of the horizontal field (the range) in each 3-h interval. A simple latitudinal correction is applied to the K value of each station and these corrected indices are then grouped into longitudinal sectors. This grouping allows for small differences in the K scalings at observatories within the group and reduces the effect of changes in station site within each group. In the Northern Hemisphere 5 groups, approximately equally spaced in latitude are averaged to form the *an* index. In the Southern Hemisphere the large proportion of ocean at 50° geomagnetic latitude means that only 3 groups contribute to the *as* index and coverage of a large portion of the southern Pacific is not possible. The two indices, *an* and *as*, are then averaged together to form the overall *am* planetary index.

The *am* index is available continuously from 1957, the first International Polar Year, at 3-hourly resolution. The fact that the index is constructed from data from a large number of longitudinally-separated magnetometer stations makes it relatively immune to seasonal and diurnal effects, such as those due to changes in ionospheric conductivity, which do not originate in the solar wind or from its coupling to the terrestrial magnetosphere.

The *aa* index is constructed in the same way as the *am* index but uses only two roughly antipodal sites, a Northern Hemisphere site in southern England and a Southern Hemisphere site in south-eastern Australia. (The position of each site has been moved a number of times, with periods of intercalibration between new and old site, during the interval for which the index is available.) Although the use of two sites introduces some minor seasonal and more important diurnal effects in the index, it has the principle advantage of being one of the longest-running continuous geophysical data sets, extending back to 1868. Since it is not possible to remove all diurnal variations from the index, its originator Mayaud (1972) advised caution in using it at its highest resolution and suggested that appropriate 24 h, or longer, interval averages should be used.

2.2 OMNI 2 data set

The OMNI 2 data set (King and Papitashvili, 2005) is produced at hourly resolution from solar wind data collected by

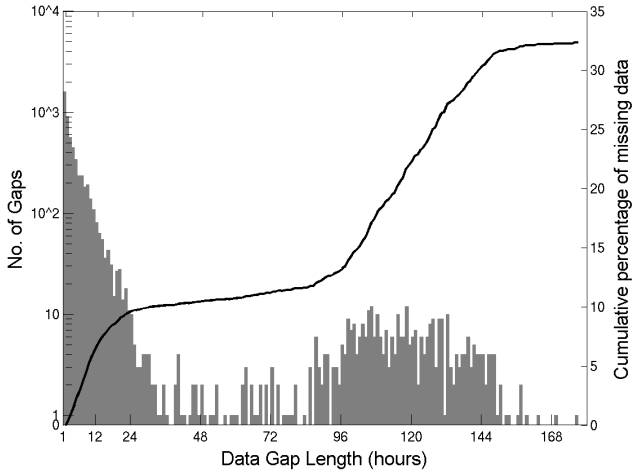

Fig. 1. Frequency of occurrence of data gaps, and the percentage of data lost from a notional ideal continuous data set for the OMNI 2 data after 1 January 1974. The grey histogram is the frequency at which data gaps of a particular length occur, while the solid black line is the cumulative percentage of missing data that the data gaps represent; for example data gaps of length less than 24 h represent approximately 10% loss from a notional continuous data set and data gaps of less than 176 h (i.e. all data gaps) represent a 33% loss.

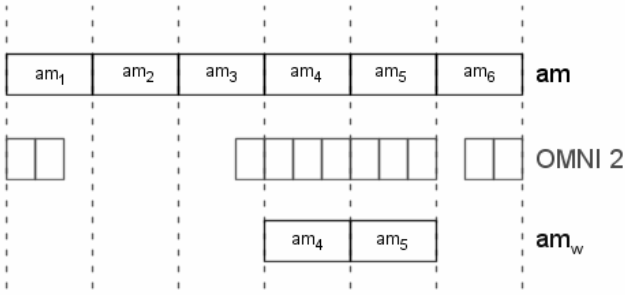

Fig. 2. Illustration of the formation of am_w from *am* and the availability of OMNI 2 hourly means. Values for each 3-h period are taken from *am*, while the presence or absence of a record is determined by whether a matching 3 h of OMNI 2 data is available.

spacecraft in geocentric orbit and in orbit around the L1 Sun-Earth Lagrange point. Over the interval of the OMNI 2 data set, since the first record available taken in 1963, data has been collected from 15 geocentric satellites and 3 upstream spacecraft. The data set is comprised of a large number of parameters, though in this study we largely restrict attention to the number density, N_{sw}, bulk flow speed, v_{sw}, and interplanetary magnetic field (IMF) strength, $|\boldsymbol{B}|$. Data from each spacecraft is lagged at a higher temporal resolution (1–5 min), assuming planar structures propagating radially away from the Sun and orientated along the ideal Parker spiral, and then averaged in "Earth time". Each hourly average point in each solar wind parameter may itself have been created from a variable number of data points depending on the spacecraft data available, with a requirement that only a single sample be available to define an hourly average.

A key part of compiling the OMNI 2 data set is the intercalibration of the various instruments used. The original compilers (Couzens and King, 1986; King, 1977) noted large uncertainties in this respect for the earliest (pre-1974) data. Recently Rouillard and Lockwood (2004) showed that the IMF data from the OMNI data set that had been high-passed filtered to remove the solar cycle variation showed a strong 1.68 year variation, which was highly anti-correlated with observed cosmic ray fluxes that had been similarly filtered. This correlation was found for all the filtered data, including that from before 1974. The unfiltered data was also highly anti-correlated with the same regression slope and correlation coefficient, but this was only true for post-1974 data. The inference is that there were calibration drifts and discontinuities in the earliest IMF data. Here we only use data including and after 1974 to avoid any such problems.

2.2.1 Gaps in the OMNI 2 data set – frequency and distribution

The OMNI 2 data set is not continuous, as demonstrated by Fig. 1. Data gaps with a length of 1-h, the temporal resolution of the OMNI 2 data set, are the most frequent, with the frequency of data gaps declining logarithmically as their length increases to about 24 h. Data gaps of 24 h in length or less account for approximately 10% of the total data that would exist for a continuous hourly-resolution data set covering the same period, as shown by the solid line on Fig. 1. Data gaps of length between 24 h and 96 h are infrequent and account for approximately a further 3% reduction from ideal continuous data. A large number of data gaps of length between 96 h and 144 h make up the majority of "missing" data however, accounting for a 20% reduction from ideal continuous data. The longest data gap present in the OMNI 2 data set after 1974 has a duration of 176 h. In total, approximately 33% of data is unavailable between 1974 and 2003, as compared to an ideal continuous solar wind data set.

The existence of these data gaps will introduce sampling errors and biases in any study based on the OMNI 2 data set. To investigate the statistical effects of these data gaps we define a new index, am_w, based on the *am* index and the availability of matching OMNI 2 data. Each *am* data point is three hours in extent, starting on hour boundaries wholly divisible by 3 (i.e. 00:00, 03:00, 06:00, 09:00... UT), and each OMNI 2 data point is one hour long starting on the hour (i.e. 00:00, 01:00, 02:00, 03:00... UT). A point is included in the am_w index, taken directly from the am index, if there are three matching OMNI 2 data points covering the same time period. The am_w index is therefore a discontinuous index with values, where present, identical to those of *am*. The absence of values in the am_w index is controlled by gaps in the OMNI 2 data set. This is illustrated in Fig. 2.

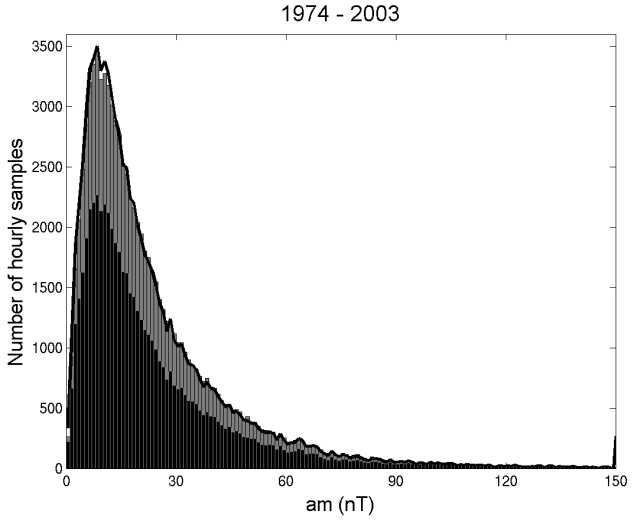

Fig. 3. Distributions of hourly values of the geomagnetic *am* (grey histogram) and am_w (black histogram) indices for 1974–2003. The final bin is for all samples with *am* or $am_w > 150$ nT. The solid black line is the distribution of the am_w values scaled by the overall ratio between all *am* and am_w sample numbers.

In Fig. 3 we plot the distributions of values of the *am* and am_w indices, as grey and black histograms, respectively. The two distributions appear extremely similar in form, indicating that the gaps in am_w are distributed randomly with respect to *am*. However, in the next section we study the effects of averaging the indices over a variety of timescales and find nevertheless significant effects.

2.2.2 Gaps in the OMNI 2 data set – effects on temporal averages

It is clear that if we compare *am* and am_w then, except for the data gaps, these two indices are identical. However, in order to correlate these indices with coupling functions at a variety of timescales, averaging will have to be performed. It is important to understand how the presence of these gaps will affect the average coupling functions constructed from the discontinuous OMNI 2 data set. The am_w index has been constructed to have the same discontinuities and thus comparison with the continuous *am* data set gives insight into the effect of data gaps on the averages.

The *am* index has no missing data and is straightforward to temporally average. We simply start at a fixed date, 1 January 1974, and take averages from non-overlapping equal intervals which are a multiple of 3-h long. (Remember that the resolution of the *am* index is 3-h.) This is illustrated for averaging intervals of 6 and 15 h in Fig. 4.

The situation with the discontinuous am_w index is more complicated. We prepare it in the same way as the *am* index, taking averages from the non-overlapping equal intervals which are a multiple of 3-h in length. However, as illustrated in Fig. 4, this means that the number of data points averaged may vary, in fact in some cases no data will be available and no average can be formed. The figure illustrates how *am* and am_w averages for the same temporal averaging interval, for example 6 and 15 h in Fig. 4, will no longer be identical.

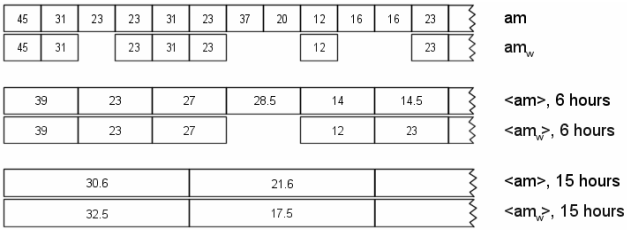

Fig. 4. Illustration of the construction of averages from identical non-overlapping periods of *am* and am_w of increasing duration. Illustrative values of *am* and am_w are given and the consequent averages of *am* and am_w for 6- and 15-h averaging intervals.

In order to evaluate the effect of data gaps, we here divide *am* and am_w into identical non-overlapping periods of increasing duration, from 3 h to 365 days (note that each period must be a whole multiple of 3-h), and average within each period for each index. Since *am* is continuous the same amount of data is averaged to form each bin of the same temporal duration. The discontinuous index am_w on the other hand, means that, in general, it will have a different amount of data averaged to form each bin of the same duration. In some cases there will be no am_w data to average for a period and the bin will be empty.

We can now examine the standard distribution of the ratio between am_w to *am* for these timescales to see how much averaged am_w deviates from the corresponding averaged *am* over the range of averaging timescales studied. (If an averaged am_w period contains no am_w data we discard both the averaged am_w and *am* for that period and it does not enter the set used to construct the standard deviation.)

The result of this evaluation is given in Fig. 5, which shows the standard deviation σ in the distribution of averaged am_w as a ratio of the corresponding *am* average, as a function of the averaging timescale. The difference between *am* and am_w is zero, by definition, at timescales of 3 hours and low at timescales of 1-year. There is a significant difference between the average values of the two indices at timescales of approximately 1-week, the difference being a maximum at 4.5 days. These differences are entirely due to the existence of the data gaps in the OMNI 2 data set, since these gaps are the only source of difference between *am* and am_w.

To develop deeper understanding of the reasons for this timescale-dependent variation in the difference between am_w and *am*, in Fig. 6 we present (in black) the ratio of am_w to *am* between 1974 and 2003 for three different timescales alongside (in red) the "coverage" of am_w (and hence the OMNI 2 data set). Coverage is here defined as the ratio of points in

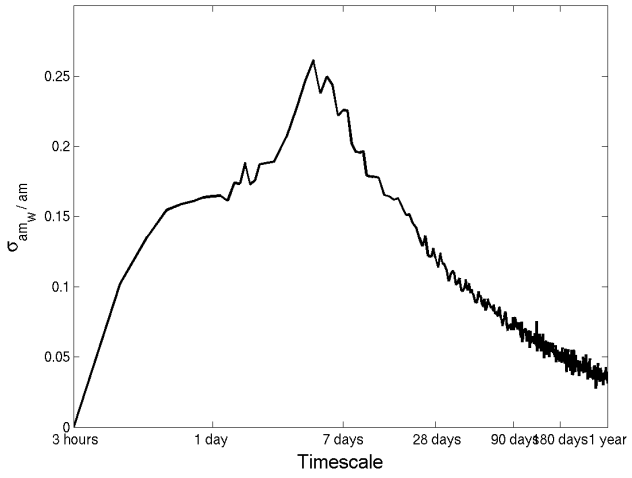

Fig. 5. Standard deviation of the ratio of am_w to am as a function of averaging timescale between 3 hours and 1 year. Periods for which there is no am_w value are discarded before computing the standard deviation.

the am_w averaging period to those in the am period, i.e. if the number of am_w points matches the number of am points in an averaging period then the coverage ratio is 1 whereas if there are half as many am_w points compared to am points in a period then the coverage ratio is 0.5. It can be seen, as expected, that the ratio of am_w to am only deviates from 1 when the coverage <1. However, the effect of less than full coverage is more significant at shorter timescales. This is because at longer timescales, as shown by Fig. 5, am_w and am tend toward long-term averages which can be approximated with fewer data points. Note that even in the period after 1995, in which the ACE satellite provides almost continuous reporting of the solar wind parameters, those data gaps that do exist can have a significant effect at short timescales.

3 Coupling functions

Solar wind parameters used as, or used to construct, the coupling functions in this study are: $|\boldsymbol{B}|$ – the magnitude of the interplanetary magnetic field (IMF), B_s – the southward component of the IMF in the GSM frame, v_{sw} – the solar wind speed, m_{sw} – the mean ion mass and N_{sw} – the solar wind number density. Additionally we study various combinations including the coupling functions of Vasyliunas et al. (1982) derived through dimensional analysis, and here labelled P_α, and the ε parameter (Perreault and Akasofu, 1978; Akasofu, 1979, 1981; Koskinen and Taskanen, 2002). The former is the only coupling function with allowance for variability in M_E – the magnetic moment of the Earth.

The first attempt to study the correlation between a geomagnetic index and one of the solar wind parameters measured by spacecraft was made by Snyder et al. (1963) using data obtained from the Mariner 2 spacecraft. They found a positive correlation between the K_p index and the velocity of the solar wind. Later work by Hirshberg and Colburn (1969) established a connection between the southward component of the interplanetary magnetic field (IMF) and geomagnetic activity. Arnoldy (1971) introduced a half-wave rectified parameter B_S, with the definition B_S=B_z for B_z<0 and B_S = 0 for B_z>0, and found a linear relationship between B_S and the geomagnetic index AE. Because a relationship was also established between the solar wind velocity, v_{sw}, and geomagnetic disturbances a number of authors (Garrett et al., 1974; Murayama and Hakamada, 1975; Burton et al., 1975) established improved correlations using B_S and v_{sw} in combination.

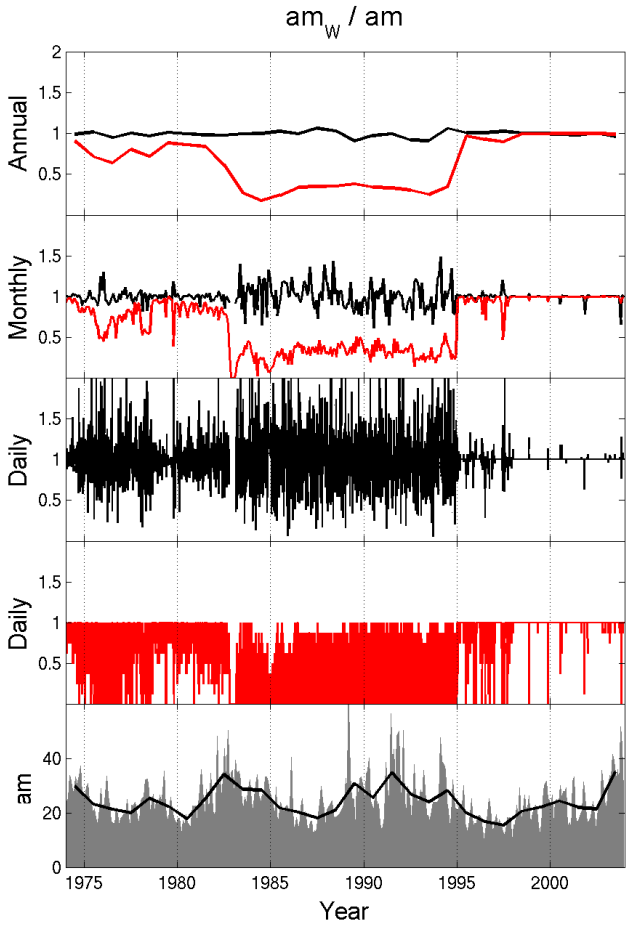

Fig. 6. Ratio of am_w to am at a variety of timescales. In the upper four panels the solid black line is the ratio of am_w to am, the solid red line is the data coverage: we display these on the same panel for annual and monthly timescales, but the high-frequency variability of both on daily timescales required us to use separate panels (3 and 4). In the fifth, bottom panel, the solid black line is the average value of am at annual timescales and the grey histogram the average value of am at monthly timescales. There is no apparent connection between am magnitude and the am_w to am ratio.

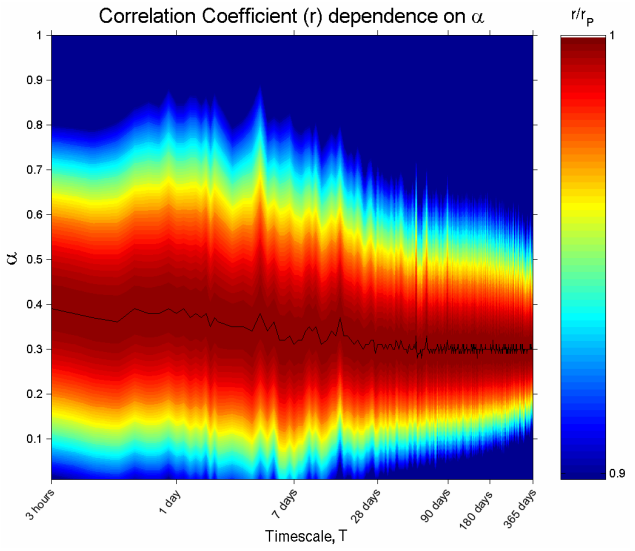

Fig. 7. Colour contours of the correlation coefficient r, between P_α and the *am* geomagnetic index, as a ratio of its peak value at that T, r_p, as a function of the averaging timescale T (on a logarithmic scale) and the value of the coupling exponent α.

Burton et al. used the coupling function $B_S v_{sw}$ (which equals the half-wave rectified dawn-to-dusk component of the interplanetary electric field) to predict the time profile of the D_{st} index and obtained good agreement between the predicted and observed values. Murayama and Hakamada used the empirical formula $B_S v_{sw}^2$ to establish a correlation with the *AE* index. Garrett et al. selected the formula $B_S v_{sw}+k_1\sigma\, v_{sw}+k_2$ to correlate with the *Ap* and *AE* indices, where k_1 and k_2 are constants and σ^2 is the total variance of the IMF. They found this produced a correlation equal to that of $B_S v_{sw}^2$ and preferred the latter because of the clearer physical meaning of $B_S v_{sw}$. Svalgaard (1977) was the first to incorporate the solar wind dynamic pressure P_{sw} in his correlative study with the *am* index, using a coupling function of the form $|\boldsymbol{B}| v_{sw}(N_{sw} v_{sw}^2)$.

Subsequent studies focused more on theoretical derivations of expected power transfer to the magnetosphere. Prime amongst these were the studies of Perreault and Akasofu (1978) and Vasyliunas et al. (1982). The coupling function, P_α, developed by Vasyliunas et al using dimensional analysis, is a physics-based estimate of the power extracted from the solar wind. P_α, and the product of three terms:

$$P_\alpha = \left(\pi\, l_0^2\right) \times \left(m_{sw} N_{sw} v_{sw}^3 \Big/ 2\right) \times (t_r) \tag{1}$$

The first term in brackets on the right-hand side is the area (a circle of radius l_0) that the magnetosphere presents to the solar wind flow. The second term in brackets is the flux of the kinetic energy density in the solar wind flow. The third term is the "transfer function", t_r, which is the fraction of the power incident on the magnetosphere that is extracted.

A hemispherical shape for the dayside magnetosphere is assumed, for which l_0 is the stand-off distance of the nose of the magnetosphere and can be computed from the pressure balance between the magnetic pressure of the terrestrial field and the solar wind dynamic pressure P_{sw} $\left(=m_{sw} N_{sw} v_{sw}^2\right)$ (Schield, 1969). This yields a value of l_0 proportional to $\left(M_E^2 \big/ P_{sw}\mu_0\right)^{1/6}$, i.e.

$$l_0 = k\left(M_E^2 \Big/ P_{sw}\mu_0\right)^{1/6} \tag{2}$$

The dimensionless form of the transfer function adopted by Vasyliunas et al., has a $\sin^4(\theta/2)$ dependence on the IMF clock angle θ (in the GSM reference frame):

$$t_r = k_1\left(1/M_A^2\right)^{\alpha} \sin^4\left(\theta/2\right) \tag{3}$$

where k_1 is a dimensionless constant, M_A is the solar wind Alfvén Mach number (equal to $\sqrt{\mu_0 P_{sw}}\big/|\boldsymbol{B}|$) and α is called the coupling exponent. Aoki (2005) found that the $|\boldsymbol{B}|\sin^4(\theta/2)$ function does not correlate as highly as B_S with geomagnetic activity, but the former has the advantage of being continuous in slope. In the theory of Vasyliunas et al., the transfer function must be dimensionless and we note that Aoki did not include a term of the form $|\boldsymbol{B}|^{2\alpha}\sin^4(\theta/2)$, as actually used by Vasyliunas et al, in his study.

Substituting Eqs. (2) and (3) into (1), we get

$$P_\alpha = \left(k\pi \Big/ 2\mu_0^{(1/3+\alpha)}\right)\, m_{sw}^{(2/3-\alpha)} M_E^{2/3} N_{sw}^{(2/3-\alpha)} \times v_{sw}^{(7/3-\alpha)}\, |\boldsymbol{B}|^{2\alpha} \sin^4\left(\theta/2\right) \tag{4}$$

We here fix the value of α at 0.3, ensuring that P_α has no more free parameters than any of the other coupling functions. Figure 7 analyses the dependence of the correlation coefficient r, on the timescale T and the value of α. (We explain our choice to plot timescale T logarithmically in Sect. 4). The value of r, as a ratio of its peak value at that T, r_p, is contoured as a function of T and α. Note that the correlation is only a weak function of α for any one T with values of r/r_p exceeding 0.9 for much of the phase space shown. The black line is for $r=r_p$ and it can be seen that the optimum α is 0.3 for all T exceeding 28 days. At lower T, there is a slight rise in the optimum α, such that it is near 0.4 at T=3 h. We can compare this to previous estimates: Murayama (1982) found α=0.4 for T near 1 day, Bargatze et al. (1986) found α=0.5 for T<1 h and Stamper et al. (1999) found α=0.38 for T=1 year. We note the differences between all of these results and α=0.3 are not significant and that the earlier studies had smaller (with much less continuous data) datasets available and any differences are almost exclusively due to this.

Physically, Vasyliunas et al. stress that α is an empirical fit parameter that is constrained by dimensional analysis. As discussed below, α=1, with a fixed l_0 value, reduces P_α to the epsilon parameter. Vasyliunas et al. point out that α=1 yields a P_α dependence on $|\boldsymbol{B}|^2$ and α=0.5 yields a linear dependence on $|\boldsymbol{B}|$. It is useful to note that Eq. (4) also shows

that α=0 would mean that there was no dependence on $|\boldsymbol{B}|$ and that α=2/3 would mean there was no dependence on solar wind density, $m_{sw}N_{sw}$ (and P_α would vary as $v_{sw}|\boldsymbol{B}|^{4/3}$: i.e. the compression effect on the magnetospheric cross sectional area would happen to counter-balance exactly any rise in solar wind kinetic energy density) and that α=7/6 would mean there was no dependence on solar wind speed, v_{sw} (and P_α would vary as $|\boldsymbol{B}|^{7/3}(m_{sw}N_{sw})^{-1/3}$). A value of α=0.33 yields a P_α that varies as $(m_{sw}N_{sw})^{1/3}v_{sw}^{5/3}\ |\boldsymbol{B}|^{2/3}$ and thus increases with all these solar wind parameters.

The epsilon factor described by Perreault and Akasofu (1978), on the other hand uses the Poynting vector in the solar wind $\boldsymbol{S}=\boldsymbol{E}\times\boldsymbol{B}/\mu_0$. Given $\boldsymbol{E}=-\boldsymbol{v}_{sw}\times\boldsymbol{B}$ this yields a magnitude of the solar wind Poynting vector of $S=v|\boldsymbol{B}|^2$ and

$$\varepsilon = \frac{4\pi\, l_0^2}{\mu_0} v_{sw}|\boldsymbol{B}|^2 \sin^4(\theta/2) \qquad (5)$$

where the effective magnetosphere radius l_0 is here fixed at $7\,R_E$. Physically, the problem with ε is that the energy brought by the solar wind to the magnetosphere is not in the form of Poynting vector but rather in the form of particle kinetic energy which is converted to Poynting vector by current density $\boldsymbol{J}$ in the bow shock, magnetosheath and magnetopause where $\boldsymbol{J}.\boldsymbol{E}>0$ (Cowley, 1991; Lockwood, 2004) according to Poynting's theorem. Kan and Akasofu (1982) showed that ε does reduce to P_α if α=1 and l_0 is constant; however, this is not the optimum α and hence although ε is quite similar to P_α it does not have a firm theoretical basis and is not expected to perform as well as P_α. We note that the ε parameter remains in frequent use, for example Alex et al. (2006); Wu and Lepping (2005); Partamies et al. (2003); Tanskanen et al. (2002), as a direct proxy for input power to the magnetospheric system. We note too that it is often used at timescales of between 1 min and 1 h, which is outside the scope of this study, but as shown in the following discussion ε is an inferior proxy to P_α of geomagnetic activity at all timescales greater than 3-h.

4 Correlation at a variety of timescales

In Fig. 8 we present a plot of the correlation coefficients at zero lag between the *am* index and a number of the previously defined coupling functions. The upper and lower panels show the same results; in the upper panel the timescale is plotted linearly while in the lower it is plotted logarithmically. It may be seen that for all coupling functions, apart possibly from P_{sw}, the overall trend is for a steady decline in correlation as the length of the averaging interval drops from 365 days towards 90 days. As the averaging interval shortens further, the correlations decline much more rapidly, with a rapid drop and then recovery in correlation apparent between 7 days and 3 h. These trends are more easily seen in the logarithmic plot and, since similar trends are present in all correlations between the solar wind coupling functions

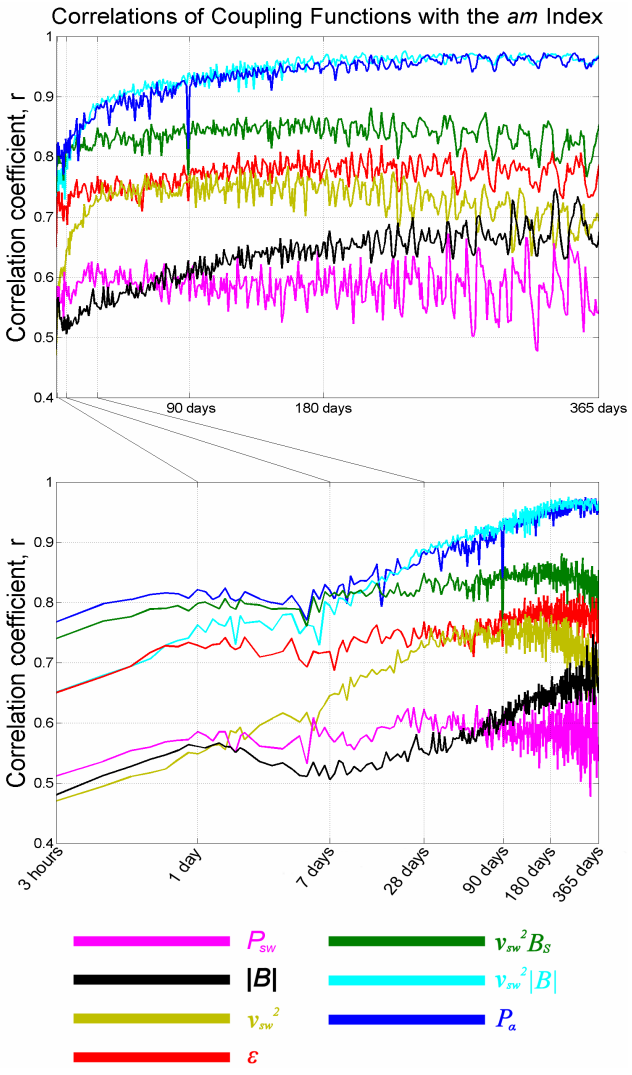

Fig. 8. Correlation between the *am* index and a number of coupling functions. Upper and lower graphs are identical, other than that the upper graph displays timescale linearly, the lower graph displays it logarithmically. The coloured lines give the results for: dark blue P_α, light blue $v_{sw}|\boldsymbol{B}|$, green $v_{sw}^2 B_S$, red ε, olive v_{sw}^2, magenta P_{sw} and black $|\boldsymbol{B}|$.

and geomagnetic indices in this paper, we choose to present all further graphs with averaging timescales plotted logarithmically.

There is considerable variability overlaying these trends. The variability increases at longer averaging intervals, which indicates that it is connected to the decreasing number of data points in the correlations as the averaging interval lengthens. (We have a finite data period to divide). Additionally, variability from the trend is greater at lower correlation coefficients, so that P_{sw} shows considerably more variability than P_α.

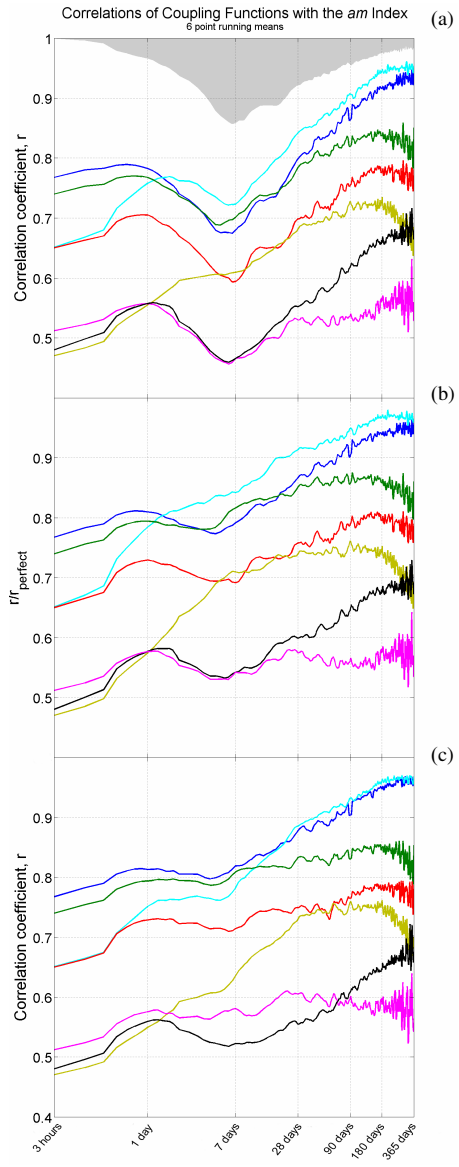

Fig. 9. Correlations between the *am* index and a number of coupling functions illustrating the effect of data gaps. Correlations are evaluated every 3 h for averaging periods of 3 h to 2 days, every 12 h for averaging periods from 2 days to 10 days and every 24 h for averaging periods from 10 days to 365 days. Point-to-point variability has been reduced using a 6-point smoothing in timescale. From upper to lower:

1. Correlation functions are identical to those shown in Fig. 7, aside from the mentioned smoothing (i.e. for coincident OMNI 2 data to all *am* data). The lines use the same colour coding as Fig. 7. The grey area is r_{perfect} and cannot be exceeded by even a perfect coupling function because of data gaps.
2. Correlation functions in **(a)** divided by r_{perfect}.
3. Correlation coefficients between the OMNI data and the am_w index.

We expect correlation coefficients to be high at longer averaging timescales as both the *am* index and the solar wind parameters from which the coupling functions are computed will tend towards their long-term averages. Evidence for this can be seen in Fig. 6 where the coverage for the monthly and annual timescales are not greatly different, but the deviation from unity of the am_w to *am* ratio is significantly greater at the shorter averaging interval. We then expect correlation coefficients to decrease as timescales shorten since we expect there to be a storage-and-release component to the energy entering the magnetospheric system which none of the studied coupling functions account for. Additionally all solar wind measurements are point measurements, often on solar wind streamlines that do not impinge on the magnetosphere, and spatial structure in the solar wind means they may differ somewhat from the average solar wind parameters at the magnetosphere. As we reduce the averaging interval, individual turbulent events and spatial structures will become relatively more important and so these differences are more significant at shorter timescales.

We emphasise here that no pairwise removal of missing data has been conducted, that the averages constructed from the continuous *am* data set are being correlated with averages constructed from the coupling functions of the discontinuous solar wind data. This appears to be how previous studies have been conducted and so we include these results for comparative purposes.

Minima in the correlation coefficients of the solar wind coupling functions occur at averaging intervals of between 5 and 6 days. If the solar wind data were continuous these minima in the coupling functions would reflect a geophysical process, for example storm timescales are of a similar magnitude. However, if we refer to Fig. 5, we see that the minima coincide with the maximum in the standard deviation of the ratio between am_w and *am*. This indicates that the minimum is at least partially due to sampling issues in the data set rather than any physical process.

The data gaps are an additional source of decorrelation. This is clear if we consider the correlation between am_w and *am*. Without any temporal averaging these two indices are identical except for the gaps in am_w and, if a pairwise removal of missing data points is conducted, must have a correlation coefficient of 1. If we conduct a temporal averaging as described previously, and illustrated in Fig. 4, the correlation is immediately reduced from unity. Correlation studies make the implicit assumption that there exists a linear function relating the two parameters being correlated. If this is true then no other data series with the same gaps as are present in the OMNI 2 data set can produce a correlation with *am* better than that for am_w.

To develop an understanding of how the missing data are affecting the correlation of the coupling functions at all timescales, we examine the effects of using am_w instead of *am* in Fig. 9. The uppermost panel of Fig. 9 simply repeats the lower panel of Fig. 8 for comparative purposes. In this

case however we apply a 6 point running mean in order to emphasise the trend in the correlation coefficient at different timescales rather than the point-to-point variability on top of those trends that can be seen in Fig. 8.

In this uppermost panel the bottom of the shaded region is the correlation coefficient between am_w and am at the relevant timescale, discussed above. We label this correlation r_{perfect}, as it is the best correlation possible between am and another variable with the same data gaps as the OMNI 2 data set. The only way it would be possible for the correlation of any coupling function to extend into the shaded area would be if the data gaps were not random in their effect. Figure 3 shows that they are random with respect to am and so we can regard r_{perfect} as the maximum r possible at that T. The closer to the r_{perfect} line that a correlation reaches, the nearer to "perfect" (given the effect of data gaps) it really is. Note how closely the lowest point of r_{perfect} matches the minima in the correlations for the coupling functions. Since all deviations from a correlation of 1 between am and am_w are due to the data gaps in OMNI 2, this is strong evidence that this is also the source of the minima in the coupling functions.

Panel (b) of Fig. 9 shows the coupling functions from panel (a) divided by r_{perfect} as a simple way of allowing for the effects of the gaps in the OMNI 2 data set. Finally in panel (c) we conduct the correlation analysis using am_w instead of am. This means that matching gaps are present in both the coupling functions and geomagnetic index data, hence a pairwise removal of missing data. This is the correct way to deal with missing data and produces a set of coupling function correlations in line with those seen in panel (b).

For the majority of the coupling functions, again excluding $|\boldsymbol{B}|$, we see that the minimum at 5 days is greatly reduced in relative importance in Figs. 9b and c where allowance is made for the data gaps. Note that many coupling functions still have a weak minimum correlation coefficient at around 7 days in Figs. 9b and c. Given that these plots have made allowance for data gaps this could be a reflection of energy storage and release on storm timescales or of effects from sector structure. We demonstrate that the true reason is the effects of gaps in the coverage of the underlying data sets.

However, even with the rigorous use of only pairwise data, we may not be comparing like with like as we average intervals. Each averaging bin will potentially contain a different number of data points depending on the presence of data gaps within it. The coverage parameter we defined previously allows us to control this variation in the number of average data points. In Fig. 10 we show the result of setting thresholds on the coverage required before including a bin in our correlation analysis and Fig. 11 gives the significance levels of these correlations. Requiring a coverage of greater than 25% is sufficient to produce a notable improvement in correlation coefficients between timescales of 1 day and 1 week. Requiring stricter coverage conditions does further improve coupling coefficients but the changes are largely marginal after this initial improvement. Note too that the significance of

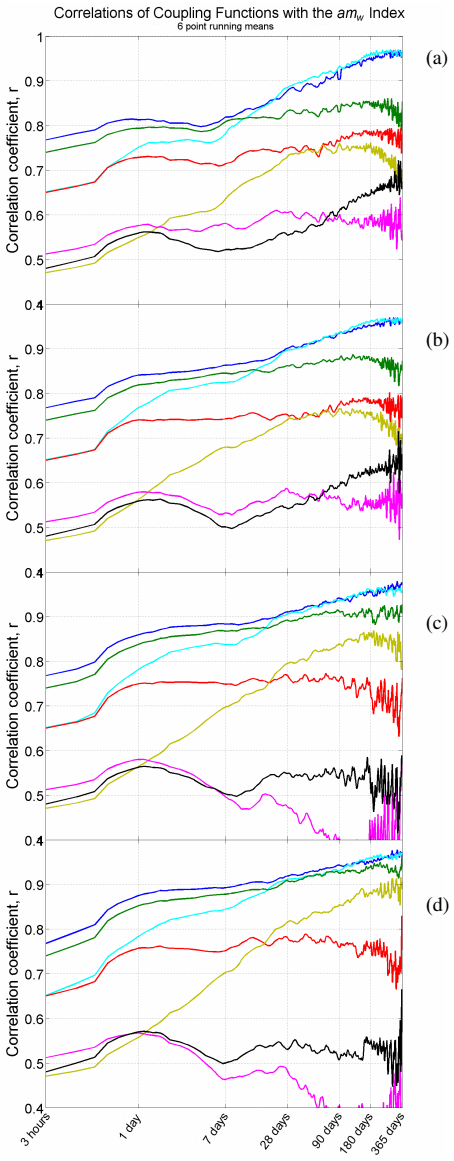

Fig. 10. Correlations between the am_w index and a number of coupling functions, illustrating the importance of considering how much data contributes to each data point in the correlation. The corresponding significance levels are shown in Fig. 10.

1. Identical to Fig. 9c, included for easier comparison. (Coincident OMNI data to coincident am_w data.) Each point need only have a single 3-hourly OMNI and am data point.
2. Coincident OMNI and am_w data, each point must contain at least 25% of its period in data.
3. Coincident OMNI and am_w data, each point must contain at least 50% of its period in data.
4. Coincident OMNI and am_w data, each point must contain at least 75% of its period in data. See significance levels of $|\boldsymbol{B}|$ and P_{sw} in panel **(d)**.

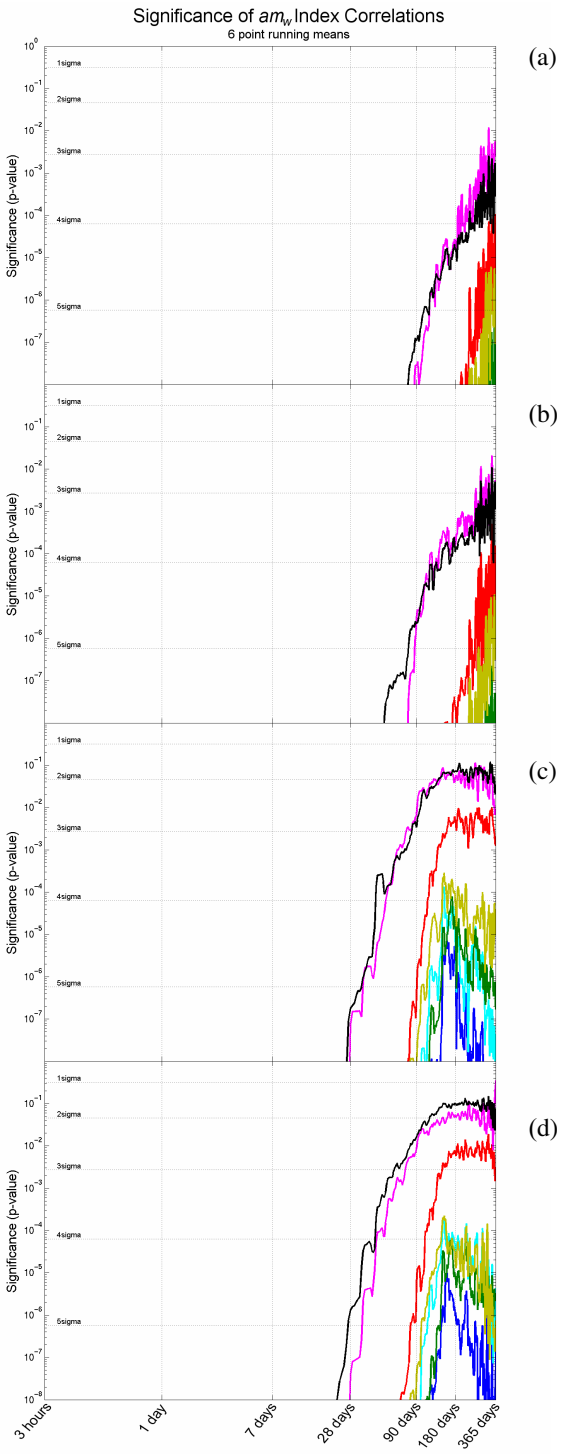

Fig. 11. Significance levels (p-values) of correlations from Fig. 10. The scale chosen means that values are only shown where they are below the 6σ level ($p > \sim 10^{-8}$).

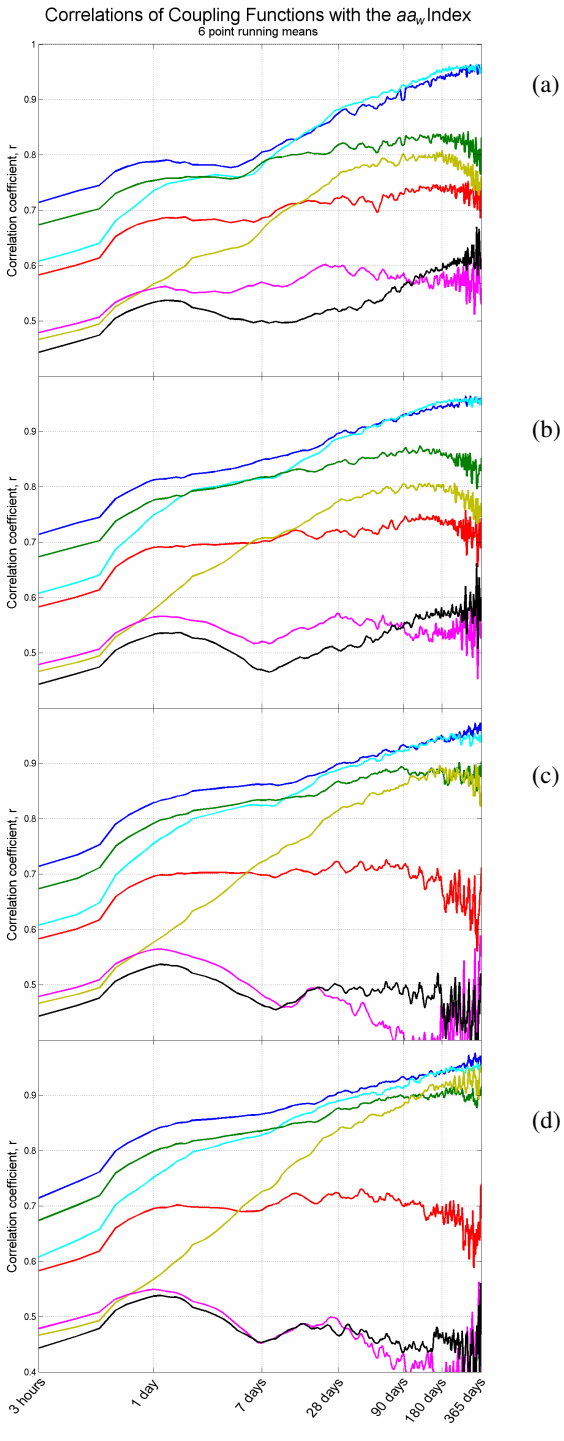

Fig. 12. The same as Fig. 10, but for the aa_w rather than the am_w index.

the less well correlated coupling functions begins to collapse at the highest coverage levels as the number of points in the correlation drops.

At a coverage of greater than 25% P_α has the highest or joint highest correlation coefficient at all timescales. Its correlation coefficient is greater than 0.9 at timescales longer than 28 days and remains better than 0.8 at all timescales of over a day. At a coverage of greater than 75% then the

correlation exceeds 0.85 at all timescales greater than 1 day. Although not shown, under a coverage criteria of 100% the correlation coefficients of the P_α coupling function exceed 0.9 at all timescales between 2 and 28 days. (Lack of data points means that results at timescales longer than 28 days are no longer significant at $p<0.05$, i.e. at the 2σ level)

Note that in Figs. 10b to d the minimum at $\sim$7 days has disappeared for all but the worst-performing coupling functions. We conclude that there is no evidence here for energy storage and release on storm timescales when data gaps are fully accounted for.

Almost all the coupling functions improve in their correlation with am_w as we make the coverage criteria stricter, the exceptions being $|\boldsymbol{B}|$ and P_{sw}. The correlation of $|\boldsymbol{B}|$ is largely independent of coverage, while that of P_{sw} is actually decreased by stricter coverage criteria. If we examine Fig. 11, which shows the significance levels (p-values) of the correlations of the coupling functions including the effect of self-correlation, then it becomes clear why these two coupling functions' correlations are exceptional. As the coverage criteria is made stricter the significance level of all coupling function correlations falls. P_{sw} begins to fail a significance test of $p<0.05$ at timescales longer than 180 days for coverage $>50\%$. The situation for $|\boldsymbol{B}|$ is somewhat better but requiring a coverage of 75% still means that it fails to be significant at timescales greater than $\sim$260 days and may be unreliable at timescales shorter than that.

Due to the large number of data samples available, all coupling function correlations are significant at greater than the 5-σ level at timescales shorter than 28 days at all coverage level requirements shown (up to 75%). The significance of all coupling function correlations decrease as timescale lengthens since the number of data samples is reduced. Similarly tighter coverage criteria reduce the number of data points at each timescale, reducing the correlation. However even for a coverage threshold of 75% and a timescale of 365 days, correlations are significant at greater than the 4-σ level for all coupling functions, except P_{sw}, $|\boldsymbol{B}|$, ε and v_{sw}^2. The low correlations at longer timescale, with reduced numbers of samples due to stricter coverage criteria cause the significance of the correlations for (in order) P_{sw}, $|\boldsymbol{B}|$, ε and v_{sw}^2 to become considerably lower than for P_α and $v_{sw}^2\,|\boldsymbol{B}|$ which remain at better than the 5-sigma level, even for 75% coverage and a timescale of one year.

We also note that increasingly strict coverage criteria have a very large effect on those coupling functions most dependent of v_{sw}, the velocity of the solar wind. v_{sw}^2 and $v_{sw}^2 B_s$ both show significant improvements in their correlations at longer timescales as coverage criteria are made stricter. At 365 days this means that between coverage $>0\%$ and coverage $>75\%$ there is an improvement of 0.2 in the correlation coefficient for v_{sw}^2. Although v_{sw} is an important component of P_α, $v_{sw}^2\,|\boldsymbol{B}|$ and ε, these parameters do not appear to be affected in the same way. The correlations of these three coupling functions appear to be largely independent of coverage at timescales longer than 90 days.

Finally, we can prepare an aa_w index in exactly the same way as the am_w index, again removing data points for which there are not 3-h of matching OMNI 2 data. Using this with the same coverage requirements as previously described results in Fig. 12. The results here are very similar to those from Fig. 10 (and thus for am_w), with the possible exception of the correlation coefficient of v_{sw}^2 which seems to be improved even more strongly at longer timescales as the coverage criteria is made stricter. The matching significance levels for Fig. 12 are not given as they are almost identical to those given in Fig. 11. At long timescales the correlations are almost exactly as for the am_w index and, as expected, the only differences arise at timescales near one day for which the correlations with the aa_w index are all slightly lower than the corresponding correlations with am_w. We conclude that *aa* is as good a proxy of energy input into the magnetosphere as the more extensive *am* index on annual timescales and is only marginally inferior at daily timescales.

5 Conclusions

We have clearly demonstrated the importance of correctly dealing with the presence of data gaps in the existing solar wind data set when comparing solar wind magnetosphere coupling functions. These data gaps can have an important influence on correlations in a way that depends on timescale and which may be mistaken for physical effects. We note in particular that after correcting for the presence of these data gaps that we are left with no evidence of storage-release affecting the solar wind magnetosphere correlations at storm timescales.

At all timescales, and with all coverage criteria, P_α consistently provides the best correlation with geomagnetic indices. $v_{sw}^2\,|\boldsymbol{B}|$ is almost identical in performance at timescales longer than 1 month but significantly less good for timescales shorter than a week. This reflects the fact that $\sin^4\left(\theta/2\right)$, B_s and $|\boldsymbol{B}|$ all tend towards constants at longer timescales. We emphasise again that P_α performs significantly better than ε at all timescales and has a firmer theoretical basis.

Acknowledgements. The OMNI data were obtained from the GSFC/SPDF OMNIWeb interface at http://omniweb.gsfc.nasa.gov. The am and aa data were obtained from the UKSSDC at http://www.ukssdc.ac.uk. We thank R. Stamper, R. Henwood and M. Wild for their help in the provision and processing of these data.

Topical Editor I. A. Daglis thanks H. Koskinen and another referee for their help in evaluating this paper.

References

Akasofu, S.-I.: Energy coupling between the solar wind and the magnetosphere, Space Sci. Rev., 28, 121–190, 1981.

Akasofu, S.-I.: Interplanetary energy flux associated with magnetospheric substorms, Planet. Space Sci., 27, 425–431, 1979.

Alex, S., Mukherjee, S., and Lakhina, G. S.: Geomagnetic signatures during the intense geomagnetic storms of 29 October and 20 November, J. Atmos. Solar-Terr. Phys., 68(7), 769–780, 2006.

Aoki, T.: On the validity of Akasofu's ε parameter and of the Vasyliunas et al. general formula for the rate of solar-wind magnetosphere energy input, Earth Planets Space, 57, 131–137, 2005.

Arnoldy, R. L.: Signature in the interplanetary medium for substorms, J. Geophys. Res., 76, 5189–5201, 1971.

Bargatze, L. F., McPherron, R. L., and Baker, D. N.: Solar wind – magnetosphere energy input functions, in: Solar Wind-Magnetosphere Coupling, edited by: Kamide, Y. and Slavin, J. A., pp 93–100, Terrapub/Reidel, Tokyo, 1986.

Baker, D. N., Hones Jr., E. W., Payne, J. B., and Feldman, W. C.: A high-time resolution study of interplanetary parameter correlations with AE, Geophys. Res. Lett., 8, 179–182, 1981.

Baker, D. N.: Statistical analyses in the study of solar wind-magnetosphere coupling, in: Solar Wind-Magnetosphere Coupling, edited by: Kamide, Y. and Slavin, J. A., pp 17–38, Terrapub/Reidel, Tokyo, 1986.

Burton, R. K., McPherron, R. L., and Russell, C. T.: An empirical relationship between interplanetary conditions and Dst, J. Geophys. Res., 80, 4204–4214, 1975.

Couzens, D. A. and King, J. H.: Interplanetary Medium Data Book - Supplement 3, Nat. Space Sci. Data Cent., Goddard Space Flight Cent., Greenbelt, M, USA, 1986.

Crooker, N. U., Feynman, J., and Gosling, J. T.: On the high correlation between long-term averages of solar wind speed and geomagnetic activity, J. Geophys. Res., 83, 1933–1937, 1977.

Garrett, H. B., Dessler, A. J., and Hill, T. W.: Influence of solar wind variability on geomagnetic activity, J. Geophys. Res., 79, 4603–4610, 1974.

Gonzalez, W. D.: A unified view of solar wind-magnetosphere coupling functions, Planet. Space Sci., 38, 627–631, 1990.

Hirshberg, J. and Colburn, D. S.: Interplanetary field and geomagnetic variations – A unified view, Planet. Space Sci., 17, 1183–1206, 1969.

Kan, J. R. and Akasofu, S.-I.: Dynamo process governing solar wind magnetosphere coupling, Planet. Space Sci., 30, 367–370, 1982.

King, J. H. and Papitashvili, N. E.: Solar wind spatial scales in and comparisons of hourly Wind and ACE plasma and magnetic field data, J. Geophys. Res., 110(A2), A02104, doi:10.1029/2004JA010649, 2005.

King, J. H.: Interplanetary medium data book, National Space Science Data Center preprint NSSDC/WDC-A-R&S 77-04, and supplements 1–5, 1977.

Koskinen, H. E. J. and Tanskanen, E. I.: Magnetosphere energy budget and the epsilon parameter, J. Geophys. Res., 107(A11), 1415–1424, doi:10.1029/2002JA009283, 2002.

Lockwood, M.: Solar output and their effects, in: The Sun, Solar Analogs and the Climate, edited by: Haigh, J. D., Lockwood, M., and Giampapa, M. S., Springer, ISBN: 3-540-23856-5, 2004.

Mayaud, P. N.: Derivation, Meaning, and Use of Geomagnetic Indices, Geophysical Monograph Series, 1980.

Mayaud, P. N.: The aa indices: A 100-year series characterising the magnetic activity, J. Geophys. Res., 77, 6870–6874, 1972.

Meng, C.-I., Tsurutani, B., Kawasaki, K., and Akasofu, S.-I.: Cross-correlation analysis of the AE index and the interplanetary magnetic field B_Z component, J. Geophys. Res., 78, 617–629, 1973.

Murayama, T. and Hakamada, K.: Effects of solar wind parameters on the development of magnetospheric substorms, Planet. Space Sci., 23, 75–91, 1975.

Murayama, T.: Coupling function between solar wind parameters and geomagnetic indices, Rev. Geophys., 20, 623–629, 1982.

Partamies, N., Amm, O., Kauristie, K., Pulkkinen, T. I., and Tanskanen, E.: A pseudo-breakup observation: Localized current wedge across the postmidnight auroral oval, J. Geophys. Res., 108(A1), 1020, doi:10.1029/2002JA009276, 2002.

Perreault, P. and Akasofu, S.-I.: A study of geomagnetic storms, Geophys. J. R. Astron. Soc., 54, 547–573, 1978.

Rouillard, A. and Lockwood, M.: Oscillations in the open solar magnetic flux with a period of 1.68 years: imprint on galactic cosmic rays and implications for heliospheric shielding, Ann. Geophys., 22, 4381–4395, 2004, http://www.ann-geophys.net/22/4381/2004/.

Schield, M. A.: Pressure balance between the solar wind and magnetosphere, J. Geophys. Res., 74, 1275–1286, 1969.

Snyder, C. W., Neugebauer, M., and Rao, U. R.: The solar wind velocity and its correlation with cosmic ray variations and with solar and geomagnetic activity, J. Geophys. Res., 68, 6361–6370, 1963.

Stamper, R., Lockwood, M., Wild, M. N., and Clark, T. D. G.: Solar causes of the long-term increase in geomagnetic activity, J. Geophys. Res., 104(A12), 28 325–28 342, doi:10.1029/1999JA900311, 1999.

Svalgaard, L.: Geomagnetic activity: Dependence on solar wind parameters, in: Coronal Holes and High Speed Wind Streams, p. 371, Colorado Assoc. Univ. Press, Boulder, 1977.

Tanskanen, E., Koskinen, H. E. J., Pulkkinen, T. I., Slavin, J. A., and Ogilvie, K.: Dissipation to the joule heating: Isolated and stormtime substorms, Adv. Space Res., 30, 2305–2311, 2002.

Vasyliunas, V. M., Kan, J. R., Siscoe, G. L., and Akasofu, S.-I.: Scaling relations governing magnetospheric energy transfer, Planet. Space Sci., 30, 359–365, 1982.

Wu, C.-C. and Lepping, R. P.: Relationships for predicting magnetic cloud-related geomagnetic storm intensity, J. Atmos. Solar-Terr. Phys., 67, 283–291, 2005.