Noise and solar-wind/ magnetosphere coupling studies: Data

Joseph E. Borovsky*

Center for Space Plasma Physics, Space Science Institute, Boulder, CO, United States

Using artificial data sets it was earlier demonstrated that noise in solar-wind variables alters the functional form of best-fit solar-wind driver functions (coupling functions) of geomagnetic activity. Using real solar-wind data that noise effect is further explored here with an aim at obtaining better best-fit formulas by removing noise in the real solar-wind data. Trends in the changes to best-fit solar-wind formulas are examined when Gaussian random noise is added to the solar-wind variables in a controlled fashion. Extrapolating those trends backward toward lower noise makes predictions for improved solar-wind driver formulas. Some of the error (noise) in solar-wind data comes from using distant L1 monitors for measuring the solar wind at Earth. An attempt is made to confirm the improvements in the solar-wind driver formulas by comparing results of best-fit formulas using L1 spacecraft measurements with best-fit formulas obtained from near-Earth spacecraft measurements from the IMP-8 spacecraft. However, testing this methodology fails owing to observed large variations in the best-fit-formula parameters from year-to-year and spacecraft-to-spacecraft, with these variations probably overwhelming the noise-correction variations. As an alternative to adding Gaussian random noise to the solar-wind variables, replacing a fraction of the values of the variables with other values was explored, yielding essentially the same noise trends as adding Gaussian noise.

KEYWORDS

magnetosphere, solar wind, geomagnetic activity, geomagnetic indices, solar wind magnetosphere coupling, space weather borovsky: noise solar-wind/magnetosphere coupling

1 Introduction

Correlative-type data studies comparing the behavior of the magnetosphere-ionosphere system to the behavior of the solar wind are performed 1) to determine the solar-wind variables that control magnetospheric activity and 2) to determine or confirm the physics of solar-wind/magnetosphere coupling. Often a goal is to find the most accurate solar-wind driver function (coupling function) to describe magnetospheric activity in terms of solar-wind parameters [e.g., Newell et al., 2007; 2008; Borovsky, 2014; McPherron et al., 2015; Lockwood and McWilliams, 2021]. There is hope that this best driver function describes the physics of solar-wind/magnetosphere coupling, but see the

discussion in Borovsky (2021). It is known that noise in the data reduces the quality of a fit of one data set to another data set (Spearman, 1904; Bock and Petersen, 1975; Liu, 1988; Hutcheon et al., 2010). This is true for the magnetospheric and solar-wind data sets (Sivadas and Sibeck, 2022). Using artificial data sets Borovsky (2022a) demonstrated that noise (error) in the solar-wind measurements can also change the functional form of best-fit solar-wind driver functions when fitting solar-wind data to geomagnetic indices. Note that Borovsky (2022a) found that adding noise to geomagnetic indices does not change the best-fit formula for the solar-wind driving of that index, the added noise just reduces the solar-wind/magnetosphere correlation coefficient.

We know that the real solar-wind data used for such coupling studies has noise (errors), mostly owed to the use of solar-wind monitors located far from the Earth at L1 (Sandahl et al., 1996; Ashour-Abdalla et al., 2008; Walsh et al., 2019; Burkholder et al., 2020; Borovsky, 2020a, 2022b; Lockwood, 2022; Sivadas and Sibeck, 2022). The basic error is incorrect values caused by the fact that the solar wind that hits an L1 monitor is typically not the solar wind that hits the Earth (Walsh et al., 2019; Burkholder et al., 2020; Borovsky, 2020a, 2022b). The error this causes is incorrect values, not propagation-time errors. If that measurement noise could be removed, more-valuable best-fit driver functions, in the sense that they may better describe the physics of coupling, could be obtained.

In this report the effects of noise in the real solar-wind data will be explored. Controlled noise will be added to the solar-wind data and the effect of that noise on best-fit solar-wind drivers will be analyzed. To explore the effects of noise, noise will be added to the solar-wind data in two simple, controlled fashions: 1) random values extracted from Gaussian distributions will b added to all values in the solar-wind time series or 2) a fraction of the time-series values will be completely replace with other values. Arguments extrapolating backward to remove noise in the solar-wind data will be made and data from the years in which the near-Earth solar-wind monitor IMP-8 operated will be explored.
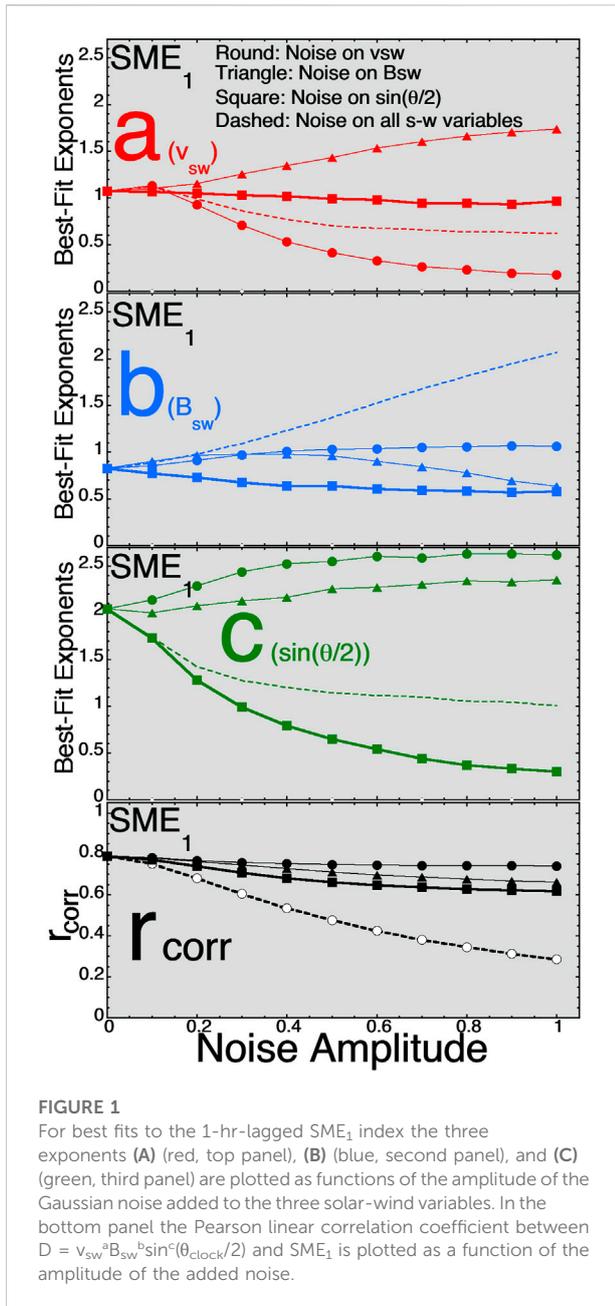
2 Data sets

A commonly used solar-wind driver function that was obtained as a best fit to geomagnetic data is the "Newell function" $v_{sw}^{4/3}B_\perp^{2/3}\sin^{8/3}(\theta_{clock}/2)$ (Newell et al., 2007), where v_{sw} is the solar-wind speed, $B_\perp$ is the magnitude of the component of the solar-wind magnetic field that is perpendicular to the Sun-Earth line, and $\theta_{clock} = \text{Arccos}(B_z/(B_y^2+B_z^2)^{1/2})$ is the (GSM) clock angle of the solar-wind magnetic field relative to the Earth's magnetic dipole. Here we will form best-fit driver functions $D = v_{sw}^a B_{sw}^b \sin^c(\theta_{clock}/2)$ where the exponents a, b, and c are optimized so that D has the largest Pearson linear correlation coefficient with various time-lagged

geomagnetic indices. For the solar-wind data 1-h averages in the OMNI2 data set (King and Papitashvili, 2005) for the years 1995–2018 are used. In making hourly averaged values of $\sin(\theta_{clock}/2)$, hourly averaged values of B_y and B_z from OMNI2 are used: Lockwood and McWilliams (2021) and Lockwood (2022) point out that creating high-time-resolution values of $\sin(\theta_{clock}/2)$ and then making an hourly average of those $\sin(\theta_{clock}/2)$ values (combine-then-average) would be a superior method. Statistically comparing the results of the two methods finds that the distribution of θ_{clock} values obtained have very similar mean values, but the calculate from high-resolution and then average yields a distribution with many fewer occurrences near 0° clock angle and many fewer occurrences near 180° clock angle. Thus for a given functional form of a driver function, the combine-then-average method yields far fewer cases of extremely weak driving and far fewer cases of extremely strong driving (as measured by the magnitude of the driver function) than does the calculating from averaged values of B_y and B_z (average-then-combine).
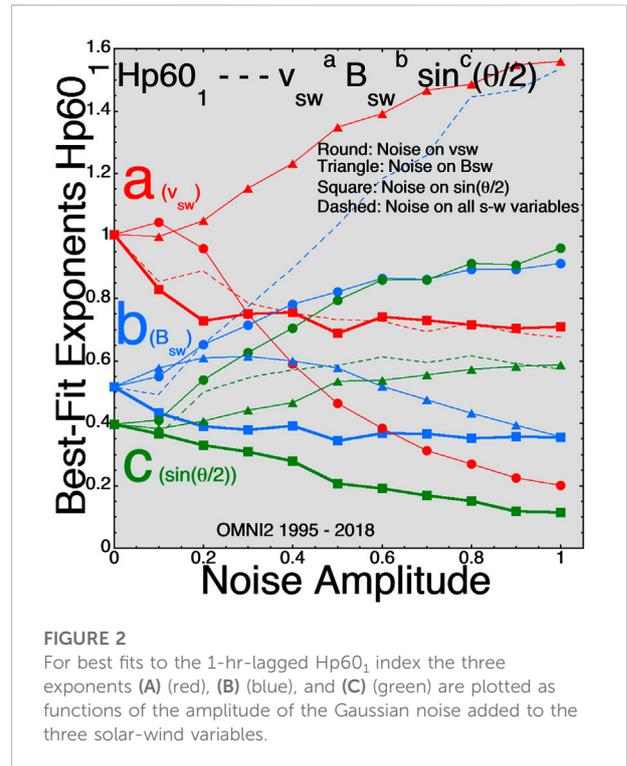
In the data analysis each of the three solar-wind variables v_{sw}, B_{sw}, and $\sin(\theta_{clock}/2)$ is "standardized" by first subtracting the mean value from every value and then by dividing every value by the standard deviation of the distribution of values. Each standardized solar-wind variable has a mean value of zero and a standard deviation of unity. Then Gaussian noise (Gaussianly distributed random numbers) is added to the standardized solar-wind variables in a number of ways: adding noise to only one solar-wind variable or adding noise to all three solar-wind variables. The amplitude of the noise is varied. When the amplitude of the noise is "1" the noise has the same standard deviation as the variable.

The years 1995–2018 are chosen for two reasons. First, we will be using at times the Hp60 index, which is only available since 1995. Hp60 is a 60-min-resolution version of the 3-hr-resolution Kp index, available at ftp://ftp.gfz-potsdam.de/pub/home/obs/Hpo. Second, as part of this project we want to compare the results from using solar-wind data at L1 with solar wind measurements closer to the Earth from the IMP-8 spacecraft (Feldman et al., 1978; Butler, 1980): the OMNI2 data beginning in 1995 is almost exclusively from L1 with the WIND spacecraft launched in 1995 and the ACE spacecraft launched in 1998.

The geomagnetic indices that will be used are the following, all at 1-h time resolution. The time lags given to each index comes from optimal lags found in prior solar-wind-coupling studies. The auroral-electrojet indices AE, AL, and AU (and their SuperMag equivalents SME, SML, and SMU) measure the peak intensity of high-latitude ionospheric current in the auroral electrojet: they are considered to be an indicator of the intensity of auroral activity (Goertz et al., 1993) and total Joule dissipation in the ionosphere (Baumjohann, 1986). They are utilized with a 1-h time lag from the OMNI solar wind data. The polar-cap index PC is a measure of the intensity of

FIGURE 1
For best fits to the 1-hr-lagged SME$_1$ index the three
exponents **(A)** (red, top panel), **(B)** (blue, second panel), and **(C)**
(green, third panel) are plotted as functions of the amplitude of the
Gaussian noise added to the three solar-wind variables. In the
bottom panel the Pearson linear correlation coefficient between
D = v$_{sw}$^{a}B$_{sw}$^{b}sin^c(θ_{clock}/2) and SME$_1$ is plotted as a function of the
amplitude of the added noise.

FIGURE 2
For best fits to the 1-hr-lagged Hp60$_1$ index the three
exponents **(A)** (red), **(B)** (blue), and **(C)** (green) are plotted as
functions of the amplitude of the Gaussian noise added to the
three solar-wind variables.

(Dessler and Parker, 1959; Sckopke, 1966; Liemohn et al., 2001),
but it is also perturbed by the Chapman-Ferraro current at the
dayside magnetosphere (Su and Konradi, 1975; Siscoe et al.,
2005) and by the cross-tail current in the magnetotail (Ohtani
et al., 2001; Borovsky and Denton, 2010). It is utilized with a 2-h
time lag from the OMNI solar wind data.

3 The data studies

In Figure 1 the 1-hr-lagged (from the solar wind)
SuperMAG auroral electrojet index SME$_1$ is studied. SME
measures high-latitude auroral-zone ionospheric currents
using 80-or-more ground-based magnetometers [the
SuperMAG network (Bergin et al., 2020)], wheras the older
AE index makes the same measurement using
~12 magnetometers. In the top three panels the values of
the three exponents a, b, and c of D = v$_{sw}$^{a}B$_{sw}$^{b}sin^c(θ_{clock}/2) are
plotted vertically as a function of the amplitude of the added
noise horizontally: a (v$_{sw}$) is plotted in red, b (B$_{sw}$) is plotted in
blue, and c [sin(θ_{clock}/2)] is plotted in green. For the solid
curves with round points noise is added only to v$_{sw}$, for the
solid curves with triangle points noise is added only to B$_{sw}$, for
the thick solid curves with square points noise is added only to
sin(θ_{clock}/2), and for the dashed curves noise is added to all
three of the solar-wind variables. In the bottom panel of
Figure 1 the Pearson linear correlation coefficient between

ionospheric current flowing across the northern polar cap: it is
related to the intensity of polar-cap anti-Sunward convection
(Stauning, 2013) and the cross-polar-cap potential (Ridley and
Kihn, 2004). It is utilized with a 0-h time lag from the OMNI
solar wind data. The Kp, Hp60, and Ap60 indices measure the
strength of mid-latitude ionospheric currents around the Earth:
they are a measure of the strength of magnetospheric convection
(Thomsen, 2004). They are utilized with a 1-h time lag from the
OMNI solar wind data. The Dst current measures an equatorial
magnetic-field perturbation: that perturbation is related to the
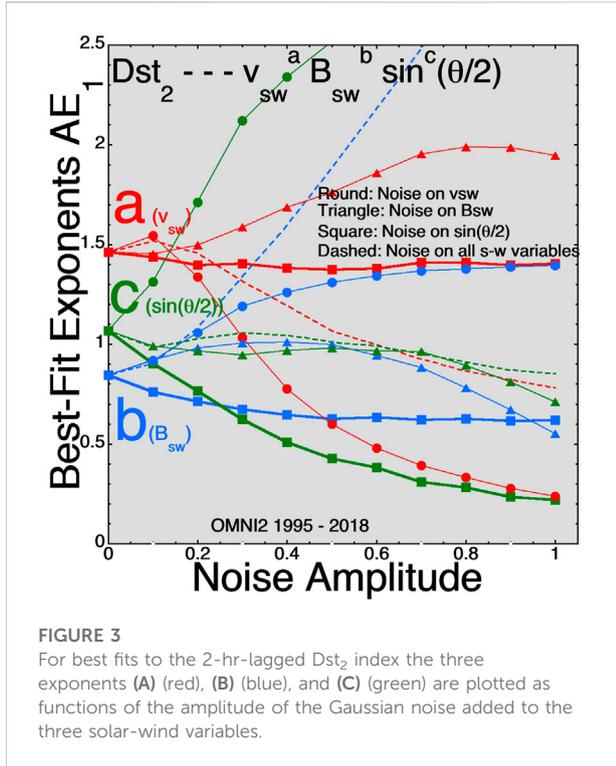plasma pressure in the dipolar portions of the magnetosphere

FIGURE 3
For best fits to the 2-hr-lagged Dst_2 index the three exponents **(A)** (red), **(B)** (blue), and **(C)** (green) are plotted as functions of the amplitude of the Gaussian noise added to the three solar-wind variables.

TABLE 1 An increase (↑) or a decrease (↓) to the three exponents a, b, and c of $D = v_{sw}{}^a B_{sw}{}^b \sin^c(\theta_{clock}/2)$ are noted as Gaussian random noise is added to the OMNI2 solar-wind variables for the years 1995−2018.

Increasing Noise on Variable	a	b	c
AE_1: Noise on v_{sw}	↓	↑	↑
AE_1: Noise on B_{sw}	↑	↑↓	↑
AE_1: Noise on $\sin(\theta/2)$	↓	↓	↓
AE_1: Noise on all 3	↓	↑	↓
SME_1: Noise on v_{sw}	↑↓	↑	↑
SME_1: Noise on B_{sw}	↑	↑↓	↑
SME_1: Noise on $\sin(\theta/2)$	↓	↓	↓
SME_1: Noise on all 3	↓	↑	↓
AL_1: Noise on v_{sw}	↓	↑	↑
AL_1: Noise on B_{sw}	↑	↑↓	—
AL_1: Noise on $\sin(\theta/2)$	↓	↓	↓
AL_1: Noise on all 3	↓	↑	↓
SML_1: Noise on v_{sw}	↑↓	↑	↑
SML_1: Noise on B_{sw}	↑	↑↓	↑
SML_1: Noise on $\sin(\theta/2)$	↓	↓	↓
SML_1: Noise on all 3	↑↓	↑	↓
AU_1: Noise on v_{sw}	↓	↑	↑
AU_1: Noise on B_{sw}	↑	↑↓	↑
AU_1: Noise on $\sin(\theta/2)$	↓	↓	↓
AU_1: Noise on all 3	↓	↑	↓
SMU_1: Noise on v_{sw}	↑↓	↑	↑
SMU_1: Noise on B_{sw}	↑	↑↓	↓↑
SMU_1: Noise on $\sin(\theta/2)$	—	↓	↓
SMU_1: Noise on all 3	↑↓	↑	↓
PCN_0: Noise on v_{sw}	↑	↑	↑
PCN_0: Noise on B_{sw}	↑	↑↓	—
PCN_0: Noise on $\sin(\theta/2)$	—	↓	↓
PCN_0: Noise on all 3	↓	↑	↓
Kp_1: Noise on v_{sw}	↑↓	↑	↑
Kp_1: Noise on B_{sw}	↑	↑↓	↑
Kp_1: Noise on $\sin(\theta/2)$	↓	↓	↓
Kp_1: Noise on all 3	↓	↑	↑
$Hp60_1$: Noise on v_{sw}	↑↓	↑	↑
$Hp60_1$: Noise on B_{sw}	↑	↑↓	↑
$Hp60_1$: Noise on $\sin(\theta/2)$	↓	↓	↓
$Hp60_1$: Noise on all 3	↓	↑	↑
$Ap60_1$: Noise on v_{sw}	↑↓	↑	↑
$Ap60_1$: Noise on B_{sw}	↑	↑↓	↓↑
$Ap60_1$: Noise on $\sin(\theta/2)$	—	↓	↓
$Ap60_1$: Noise on all 3	↓	↑	↓
Dst_2: Noise on v_{sw}	↑↓	↑	↑
Dst_2: Noise on B_{sw}	↑	↑↓	↑
Dst_2: Noise on $\sin(\theta/2)$	—	↓	↓
Dst_2: Noise on all 3	↑↓	↑	↓

D and SME_1 is plotted as a function of the amplitude of the added noise.

Note that Newell et al., 2007 used $B_\perp = (B_y{}^2 + B_z{}^2)^{1/2}$ and used multiple hours of solar-wind data in the fitting: here we are using $B_{sw} = (B_x{}^2 + B_y{}^2 + B_z{}^2)^{1/2}$ and using only a single hour of solar-wind data in the fitting.

SME is a "high-latitude" index measuring the strength of high-latitude magnetosphere-ionosphere currents. In creating plots similar to Figure 1 for other high-latitude indices (AE, AU, AL, and PCI, and the SuperMAG versions of the AL and AU indices called SML and SMU) one finds very similar trends on the plots, with the exponent c on $\sin(\theta_{clock}/2)$ being the largest of the three exponents.

Figure 2 plots the best-fit exponents a, b, and c for $D = v_{sw}{}^a B_{sw}{}^b \sin^c(\theta_{clock}/2)$ for the 1-hr-lagged $Hp60_1$ index. Hp60 is a convection-strength index (as is Kp, Ap, and MBI) (Thomsen, 2004) reacting to the strength of magnetospheric convection. These convection indices do not focus on the rapidly changing clock angle of the solar wind $\sin(\theta_{clock}/2)$ and so the c exponent for $\sin(\theta_{clock}/2)$ is small for these convective indices.

Figure 3 plots the best fit exponents a, b, and c for $D = v_{sw}{}^a B_{sw}{}^b \sin^c(\theta_{clock}/2)$ for the 2-hr-lagged Dst index. Dst measures mostly the pressure (diamagnetic currents) of plasma in the inner magnetosphere. The 2-h lag was chosen because that was found to be optimal in prior solar-wind/magnetosphere coupling studies using 1-h averaged data [e.g. Borovsky,
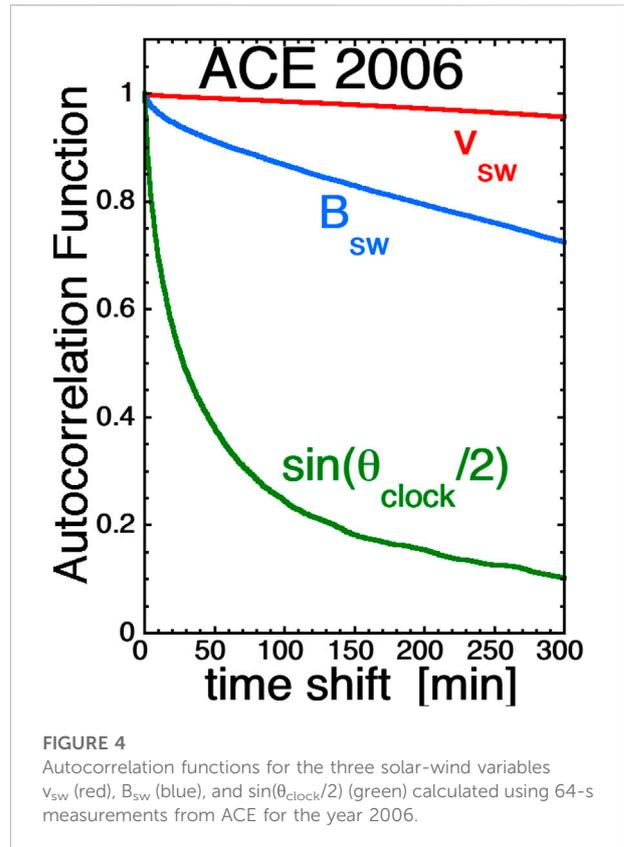
2020b]. Like the convection indices, the exponent c of $\sin(\theta_{clock}/2)$ is weak.

Examining a number of these plots for various geomagnetic indices some trends can be seen. 1) Increasing the noise on v_{sw} reduces the magnitude of the exponent a on v_{sw}. 2) Increasing the noise on $\sin(\theta_{clock}/2)$ reduces the magnitude of the exponent c on $\sin(\theta_{clock}/2)$. 3) Increasing the noise on B_{sw} at first increases the magnitude of the exponent b on B_{sw} and then as the noise amplitude becomes large it decreases the magnitude of the exponent b. 4) Increasing the noise on v_{sw} tends to raise the magnitudes of the exponents b and c on B_{sw} and $\sin(\theta_{clock}/2)$. 5) Increasing the noise on B_{sw} tends to raise the magnitudes of the exponents a and c on v_{sw} and $\sin(\theta_{clock}/2)$. 6) Increasing the noise on $\sin(\theta_{clock}/2)$ tends to lower the magnitude of all three exponents a, b, and c. Note that there are exceptions to these trends.

Examining plots similar to those of Figures 1–3 for a number of geomagnetic indices, the trend in each of the exponents a, b, and c are noted in Table 1 as noise is added to the OMNI2 solar-wind variables v_{sw}, B_{sw}, and $\sin(\theta_{clock}/2)$. An upward arrow in Table 1 indicates that the value of the exponent increases with increasing noise amplitude and a downward arrow indicates that the value of the exponent decreases with increasing noise amplitude.

4 Extrapolating backward to remove or reduce the solar-wind noise

Looking at the trends in the changes of the a, b, and c exponents of $D = v_{sw}^{a} B_{sw}^{b} \sin^{c}(\theta_{clock}/2)$ with increasing added noise, one might guess that extrapolating the curves in the plots of Figures 1–3 back to values of noise < 0 might represent the better values of the exponents a, b, and c in the presence of less solar-wind noise (less error). Figure 4 plots the temporal autocorrelation functions of the three variables v_{sw} (red curve), B_{sw} (blue curve), and $\sin(\theta_{clock}/2)$ (green curve) obtained from 64-s-resolution data from ACE SWEPAM (McComas et al., 1998) and ACE MAG (Smith et al., 1998) for the year 2006. (The year 2006 is in the late declining phase with mild high-speed streams and with a nice mix of solar-wind plasma types [cf. Xu and Borovsky, 2015].) The autocorrelation times (using the 1/e method) in Figure 4 are about 43.9 h for v_{sw}, 15.3 h for B_{sw}, and 53 min for $\sin(\theta_{clock}/2)$. The temporal autocorrelation function represents the spatial structure of the solar-wind plasma and magnetic field advected past the measuring spacecraft. The spatial structure of the variable $\sin(\theta_{clock}/2)$ in the solar wind is much smaller than the spatial structure of the variables B_{sw} and v_{sw}, hence the much shorter autocorrelation time for $\sin(\theta_{clock}/2)$ [See also Lockwood and McWilliams (2021)]. One problem with a solar-wind monitor for Earth at L1 is that the solar wind that hits the L1 monitor is not the same solar wind that hits the Earth (Sandahl et al., 1996;

FIGURE 4
Autocorrelation functions for the three solar-wind variables v_{sw} (red), B_{sw} (blue), and $\sin(\theta_{clock}/2)$ (green) calculated using 64-s measurements from ACE for the year 2006.

Ashour-Abdalla et al., 2008; Walsh et al., 2019; Burkholder et al., 2020; Borovsky, 2020a): the monitored solar-wind streamline often misses the Earth by 10s of R_E typically passing on the duskward side of the magnetosphere (Borovsky, 2022b). Contributing factors for this miss are 1) the solar-wind flow is not radial [the direction fluctuates by about ±5° plus there is a systematic dawn-dusk offset to the average flow vector (Nemecek et a., 2020a)] 2) there is a duskward aberration to the flow caused by the Earth's motion around the Sun (Fairfield, 1993), and 3) the solar-wind structure moves away from the Sun faster than the plasma wind with a velocity vector relative to the plasma that is in the Parker-spiral direction (Borovsky, 2020c; Nemecek et al., 2020b). This L1 error is much more critical for the smaller-structured $\sin(\theta_{clock}/2)$ than it is for v_{sw} or B_{sw}. Hence the thick solid curves with the square points in Figures 1–3 that add noise only to $\sin(\theta_{clock}/2)$ are the most important to think about and extrapolate backward.

For the high-latitude-current indices AE, AU, AL, SME, SMU, SML, and PCI the exponent c on $\sin(\theta_{clock}/2)$ is the largest of the three exponents and also undergoes the most-significant change (reduction) as noise is added to the $\sin(\theta_{clock}/2)$ values in the data set (cf. Figure 1). Extrapolating to negative values of noise in the thick solid curves of Figures 1–3 would increase the magnitude of the c exponent: the extrapolation would not change a (for v_{sw}) or b (for B_{sw}) significantly. For

TABLE 2 For five different geomagnetic indices the best-fit exponents a, b, and c of the solar-wind driver $D = v_{sw}{}^a B_{sw}{}^b \sin^c(\theta_{clock}/2)$ are listed for the 1995−2018 OMNI2 solar-wind data and for the OMNI2 data that utilized measurements from IMP-8. The expected change in the a, b, and c exponents if there were a reduction of noise on $\sin(\theta_{clock}/2)$ and the observed change in the exponents a, b, and c (from Table 1) are listed.

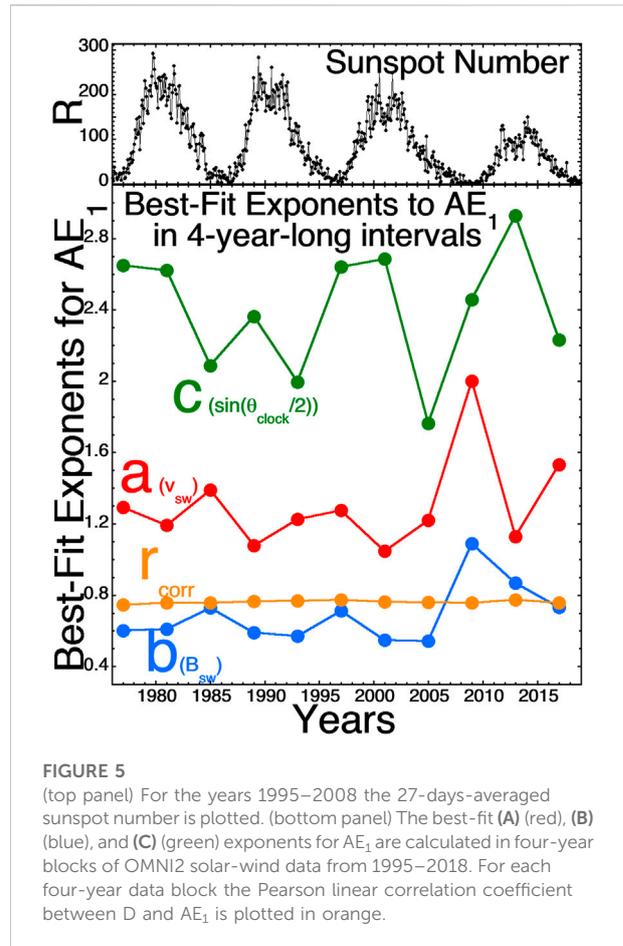|	a	b	c	r_{corr}
AE_1 1995–2018	1.071	0.711	2.025	0.772
AE_1 IMP8	1.227	0.596	2.159	0.762
AE_1 Expected	increase	increase	increase	
AE_1 Observed	increase	decrease	increase	
AL_1 1995–2018	1.206	0.703	2.406	0.749
AL_1 IMP8	1.390	0.572	2.434	0.737
AL_1 Expected	increase	increase	increase	
AL_1 Observed	increase	decrease	increase	
AU_1 1995–2018	0.779	0.706	1.310	0.678
AU_1 IMP8	0.859	0.640	1.599	0.649
AU_1 Expected	increase	increase	increase	
AU_1 Observed	increase	decrease	increase	
Kp_1 1995–2018	1.148	0.437	0.337	0.728
Kp_1 IMP8	1.198	0.441	0.368	0.714
Kp_1 Expected	increase	increase	decrease	
Kp_1 Observed	increase	none	increase	
PC_0 1995–2018	1.054	0.735	1.908	0.730
PC_0 IMP8	1.228	0.613	1.886	0.716
PC_0 Expected	none	increase	increase	
PC_0 Observed	increase	decrease	decrease	

FIGURE 5
(top panel) For the years 1995−2008 the 27-days-averaged sunspot number is plotted. (bottom panel) The best-fit (A) (red), (B) (blue), and (C) (green) exponents for AE_1 are calculated in four-year blocks of OMNI2 solar-wind data from 1995−2018. For each four-year data block the Pearson linear correlation coefficient between D and AE_1 is plotted in orange.

the magnetospheric convection index $Hp60_1$ extrapolating backward to noise values less than zero on $\sin(\theta_{clock}/2)$ in Figure 2 (thick curve with square points) would increase the a exponent (on v_{sw}) the most and also increase the b (on B_{sw}) and c [on $\sin(\theta_{clock}/2)$] exponents.

One might guess that by using the times when the nearer-to-Earth solar-wind monitor IMP-8 (Feldman et al., 1978; Butler, 1980) was used in the OMNI2 data set, these extrapolations to lower noise could be confirmed. Unfortunately, it will be seen that this is not the case. In Table 2 the best-fit values for the exponents a, b, and c for five geomagnetic indices are listed for the hours when IMP-8 solar-wind measurements were incorporated into OMNI2 (with most of the IMP-8 data incorporated in the years 1975–1994). The best-fit values of a, b, and c are also listed in Table 2 for the 1995–2018 OMNI2 data. The changes in a, b, and c predicted by extrapolation of the 1995–2018 data (cf. Table 1) are noted, and the observed changes in a, b, and c from the 1995–2018 data to the IMP-8 data are noted in Table 2. (Note that these expected changes are opposite to the arrows in Table 1: the arrows are the direction when noise is added and the "expected" change is the expectation for reduction of noise.) As can be seen in Table 2, the changes observed by using IMP-8 are in poor agreement with the

extrapolation-predicted changes from the 1995–2018 added-noise calculations: eight times there is agreement and seven times there is disagreement. Note that the change in the exponent b on B_{sw} is wrong for all five cases.

The bottom panel of Figure 5 plots the best-fit a, b, and c exponents to AE_1 in four-year blocks of OMNI2 solar-wind data from 1995–2018. As can be seen, the values of these best-fit exponents change from data subset to data subset. In the top panel of Figure 5 the 27-days-averaged sunspot number R is plotted. The temporal changes in the exponents seen in the bottom panel might be owed to solar-cycle-type variations in the solar-wind properties and/or to solar-wind spacecraft differences from year to year. As a comparison for Figure 5, Figure 6 looks at the best-fit a, b, and c exponents for AE_1 when the 1995–2018 OMNI2 data is randomized in time and binned into 10 subsets, each containing data from the full years 1995–2018. When the data is mixed in time (Figure 6) it results in much smaller variations in the best-fit exponents a, b, and c from data subset to data subset, supporting the idea that there are variances caused by spacecraft differences and by year-by-year solar-cycle differences in the wind. The variances in a, b, and c in

FIGURE 6
The best-fit **(A)** (red), **(B)** (blue), and **(C)** (green) exponents for AE_1 are plotted when the 1995−2018 OMNI2 data is randomized in time and binned into 10 subsets, each of which containing data from the full years 1995−2018. The Pearson linear correlation coefficient for each subset is plotted in orange.

FIGURE 7
For best fits to the 1-hr-lagged SME_1 index the three exponents **(A)** (red), **(B)** (blue), and **(C)** (green) are plotted in the top panel as functions of the fraction of time-series values replaced for three solar-wind variables. In the bottom panel the Pearson linear correlation coefficient between D and SME_1 is plotted as a function of the fraction replaced.
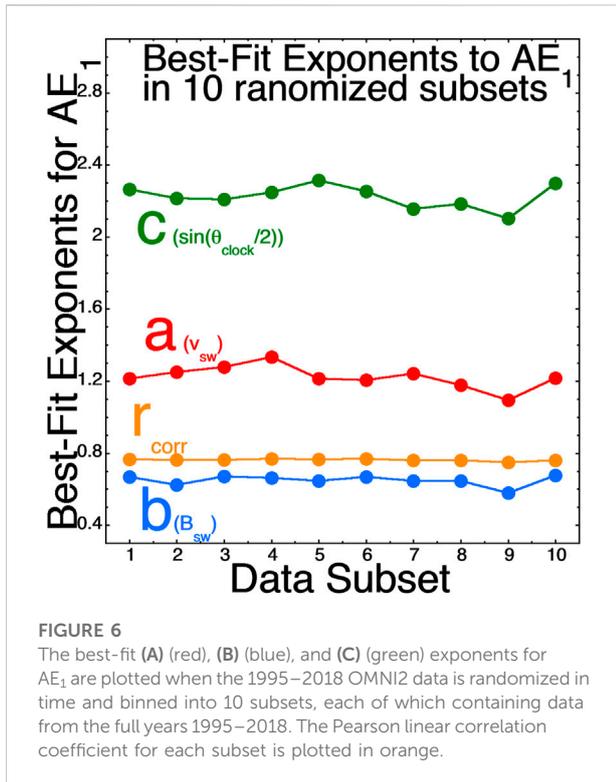
Figure 5 are large, and so the choosing of IMP-8 data in a finite set of years could result in variations of a, b, and c that are more-significant than the noise-reduction changes. Hence the hope that using IMP-8 to reduce the noise and improve the a, b, and c values is not so straightforward: moving to IMP-8 from the 1995−2018 data set does not always move the a, b, and c values in the expected direction and changing eras (years) and spacecraft results in changes to the behavior of the a, b, and c values. A question is which of these year-to-year changes in a, b, and c (e.g. Figure 5) are better and which are worse in terms of finding a driver function that describes the physics of the solar wind coupling to the Earth; as discussed in Borovsky (2021) there are math-versus-physics reasons why a driver function has a good correlation with the Earth.
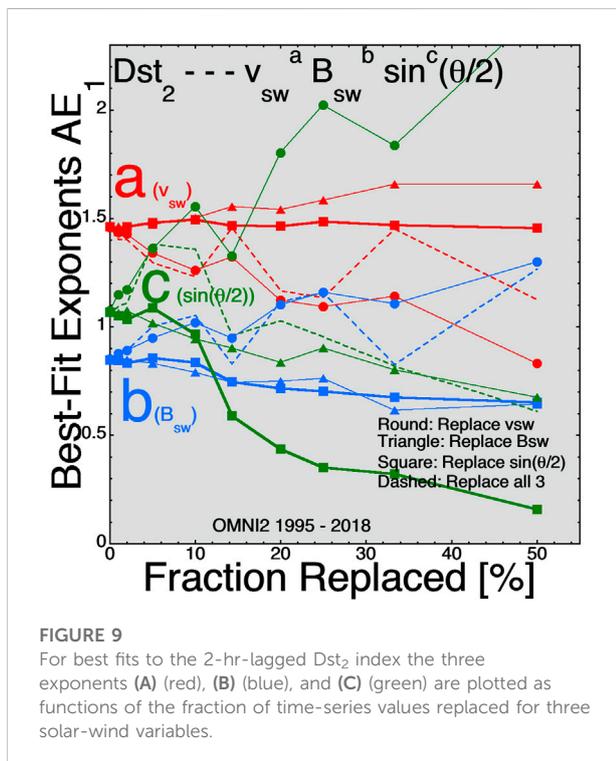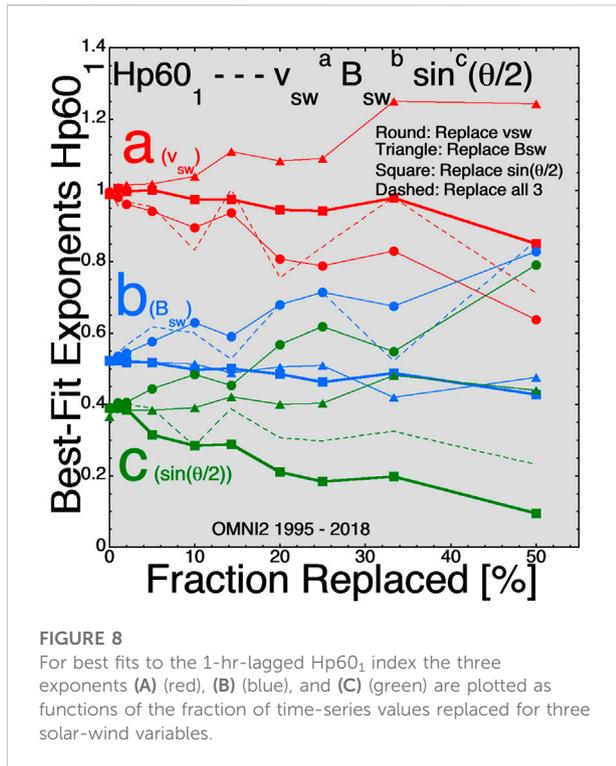
5 Discussion

This report explored expectations for solar-wind noise reduction by looking at the trends that resulted from adding Gaussian noise to solar-wind data. Extrapolating those added-noise trends backward yielded predictions for noise-reduction changes to the exponents a, b, and c of best-fit connections between the solar-wind driver function $D = v_{sw}^a B_{sw}^b \sin^c(\theta_{clock}/2)$ and various geomagnetic indices. Going from L1 solar-wind data to near-Earth IMP-8 solar-wind data, the observed changes in the exponents a, b, and c of the best-fit formula
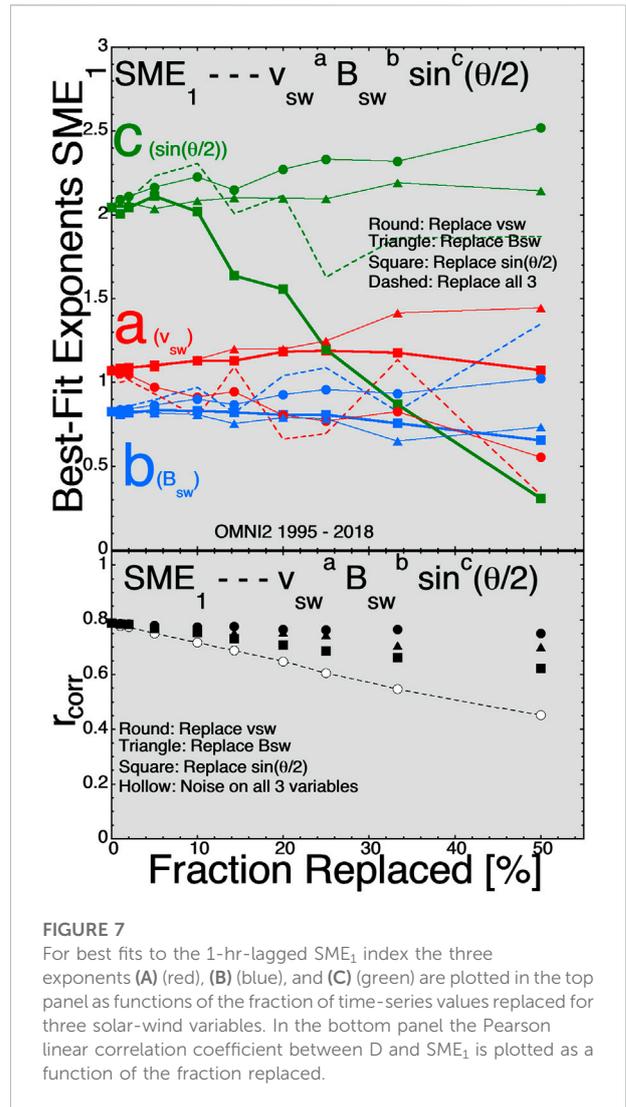
$D = v_{sw}^a B_{sw}^b \sin^c(\theta_{clock}/2)$ did not show good agreement with the expected noise-reduction changes. Looking at year-to-year changes in the a, b, and c exponents for the best-fit formula found that these year-to-year changes were large: hence different parts of the solar cycle and/or using different spacecraft to monitor the solar wind results in large changes that may overwhelm observation of the desired noise-reduction changes.

One issue with the methodology used here is the addition of Gaussian noise (random numbers with Gaussian distributions) to the OMNI2 solar-wind variables. The main error with solar-wind monitors at L1 is that the solar wind that hits a monitoring spacecraft at L1 is not the solar wind that hits the Earth. Hence the real error in the solar-wind data for Earth is not random noise on all solar-wind values, rather that some of the values are completely wrong. Adding noise, for instance, by replacing a fraction of the $\sin(\theta_{clock}/2)$

FIGURE 8
For best fits to the 1-hr-lagged $Hp60_1$ index the three exponents **(A)** (red), **(B)** (blue), and **(C)** (green) are plotted as functions of the fraction of time-series values replaced for three solar-wind variables.

FIGURE 9
For best fits to the 2-hr-lagged Dst_2 index the three exponents **(A)** (red), **(B)** (blue), and **(C)** (green) are plotted as functions of the fraction of time-series values replaced for three solar-wind variables.

values in the time series with completely different values of $\sin(\theta_{clock}/2)$ may be a more-realistic way to emulate the L1 errors in the solar-wind measurements. This is explored

in Figures 7–9 where the best-fit exponents a, b, and c are plotted as a function of the fraction of the solar-wind time-series values that are replaced with different values. Comparing Figure 7 with Figure 1 (fits to SuperMAG SME_1), Figure 8 with Figure 2 (Fits to $Hp60_1$), and Figure 9 with Figure 3 (fits to Dst_2), it is seen that the trends in the changes of the exponents a, b, and c with increasing noise (either increasing the amplitude of added Gaussian noise or increasing the fraction of time-series values replaced) are quite similar. Between these three pairs of plots there are no contradictions where one method leads to an increase in an exponent and the other leads to a decrease in that exponent. There are a few cases where the Gaussian-noise method leads to an increase or a decrease of an exponent while the replacement method shows no change in that exponent.

For 1-hour-averaged data, the L1 error most likely does not result in the whole hour being incorrect, but just part of the hour being incorrect. A hybrid methodology to add noise might try adding random noise (for the fraction of the hour being wrong) to only some of the hourly solar-wind data points. Since the two methods 1) adding noise to all solar-wind data points and 2) replacing a fraction of the data points both yielded similar results, one might guess that adding noise to a fraction of the data points will also yield similar results.

The analysis of this study using IMP-8 data found difficulty in demonstrating that near-Earth monitoring yields more-accurate solar-wind measurements for Earth, although there have been several studies estimating of the solar-wind errors between L1 and Earth (Crooker et al., 1982; Ridley, 2000; Weimer et al., 2002; Mailyan et al., 2008; Case and Wild, 2012). That is not to say that this study does not recommend fielding near-Earth solar-wind monitors in future. It is an intellectual certainty that better solar-wind-monitor measurements with truer values of what is hitting the Earth will result in better future data studies of solar-wind/magnetosphere interactions. A study to optimize monitors for solar-wind/magnetosphere interaction science is needed. One possibility is multiple spacecraft in IMP-type circular orbits (r ~ 30 R_E) (Feldman et al., 1978; Butler, 1980) where at least one spacecraft would always be in the near-Earth upstream solar wind.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://omniweb.gsfc.nasa.gov.

Author contributions

The author confirms being the sole contributor of this work and has approved it for publication.

Funding

Acknowledgments

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Ashour-Abdalla, M., Walker, R. J., Peroomian, V., and El-Alaoui, M. (2008). On the importance of accurate solar wind measurements for studying magnetospheric dynamics. *J. Geophys. Res.* 113, A08204. doi:10.1029/2007ja012785

Baumjohann, W. (1986). Merits and limitations of the use of geomagnetic indices in solar wind-magnetosphere coupling studies, in *Solar wind-magnetosphere coupling*, Y. Kamide and J. A. Slavin (eds.), p. 3, Terra Scientific, Tokyo.

Bergin, A., Chapman, S. C., and Gjerloev, J. W. (2020). A E, D_{ST}, and their SuperMAG counterparts: The effect of improved spatial resolution in geomagnetic indices. *J. Geophys. Res. Space Phys.* 125, e2020JA027828. doi:10.1029/2020ja027828

Bock, R. D., and Petersen, A. C. (1975). A multivariate correction for attenuation. *Biometrika* 62, 673–678. doi:10.1093/biomet/62.3.673

Borovsky, J. E. (2020b). A survey of geomagnetic and plasma time lags in the solar-wind-driven magnetosphere of Earth. *J. Atmos. Sol. Terr. Phys.* 208, 105376. doi:10.1016/j.jastp.2020.105376

Borovsky, J. E. (2014). Canonical correlation analysis of the combined solar-wind and geomagnetic-index data sets. *J. Geophys. Res. Space Phys.* 119, 5364–5381. doi:10.1002/2013ja019607

Borovsky, J. E., and Denton, M. H. (2010). The magnetic field at geosynchronous orbit during high-speed-stream-driven storms: Connections to the solar wind, the plasma sheet, and the outer electron radiation belt. *J. Geophys. Res.* 2010, 115A08217. doi:10. 1029/2009JA015116

Borovsky, J. E. (2022a). Noise, regression dilution bias, and solar-wind/ magnetosphere coupling studies. *Front. Astron. Space Sci.* 9, 867282. doi:10. 3389/fspas.2022.867282

Borovsky, J. E. (2020c). On the motion of the heliospheric magnetic structure through the solar wind plasma. *J. Geophys. Res. Space Phys.* 125, e2019JA027377. doi:10.1029/2019JA027377

Borovsky, J. E. (2021). Perspective: Is our understanding of solar-wind/ magnetosphere coupling satisfactory? *Front. Astron. Space Sci.* 8, 634073. doi:10. 3389/fspas.2021.634073

Borovsky, J. E. (2022b). The triple dusk-dawn aberration of the solar wind at Earth. *Front. Astron. Space Sci.* 9, 917163. doi:10.3389/fspas.2022.917163

Borovsky, J. E. (2020a). What magnetospheric and ionospheric researchers should know about the solar wind. *J. Atmos. Sol. Terr. Phys.* 204, 105271. doi:10.1016/j.jastp.2020.105271

Burkholder, B. L., Nykyri, K., and Ma, X. (2020). Use of the L1 constellation as a multispacecraft solar wind monitor. *JGR. Space Phys.* 125, e2020JA027978. doi:10. 1029/2020ja027978

Butler, P. (1980). "Interplanetary Monitoring Platform engineering history and acievements," in *Nasa TM-80758* (Greenbelt Maryland USA: Goddard Space Flight Center).

Case, N. A., and Wild, J. A. (2012). A statistical comparison of solar wind propagation delays derived from multispacecraft techniques. *J. Geophys. Res.* 117, A02101. doi:10.1029/2011ja016946

Crooker, N. U., Siscoe, G. L., Russell, C. T., and Smith, E. J. (1982). Factors controlling degree of correlation between ISEE 1 and ISEE 3 interplanetary magnetic field measurements. *J. Geophys. Res.* 87, 2224–2230. doi:10.1029/ ja087ia04p02224

Dessler, A. J., and Parker, E. N. (1959). Hydromagnetic theory of geomagnetic storms. *J. Geophys. Res.* 64, 2239–2252. doi:10.1029/jz064i012p02239

Fairfield, D. H. (1993). Solar wind control of the distant magnetotail: Isee 3. *J. Geophys. Res.* 98, 21265–21276. doi:10.1029/93ja01847

Feldman, W. C., Asbridge, J. R., Bame, S. J., and Gosling, J. T. (1978). Long-term variations of selected solar wind properties: Imp 6, 7, and 8 results. *J. Geophys. Res.* 83, 2177. doi:10.1029/ja083ia05p02177

Goertz, C. K., Shan, L.-H., and Smith, R. A. (1993). Prediction of geomagnetic activity. *J. Geophys. Res.* 98, 7673–7684. doi:10.1029/92ja01193

Hutcheon, J. A., Chiolero, A., and Hanley, J. A. (2010). Random measurement error and regression dilution bias. *BMJ* 340, c2289–1406. doi:10.1136/bmj.c2289

King, J. H., and Papitashvili, N. E. (2005). Solar wind spatial scales in and comparisons of hourly Wind and ACE plasma and magnetic field data. *J. Geophys. Res.* 110, A02104. doi:10.1029/2004ja010649

Liemohn, M. W., Kozyra, J. U., Thomsen, M. F., Roeder, J. L., Lu, G., Borovsky, J. E., et al. (2001). Dominant role of the asymmetric ring current in producing the stormtime Dst. *J. Geophys. Res.* 106, 10883–10904. doi:10.1029/ 2000ja000326

Liu, K. (1988). Measurement error and its impact on partial correlation and multiple linear regression analysis. *Am. J. Epidemiol.* 127, 864–874. doi:10.1093/ oxfordjournals.aje.a114870

Lockwood, M., and McWilliams, K. A. (2021). On optimum solar wind-magnetosphere coupling functions for transpolar voltage and planetary geomagnetic activity. *JGR. Space Phys.* 126, e2021JA029946. doi:10.1029/ 2021ja029946

Lockwood, M. (2022). Solar wind--magnetosphere coupling functions: Pitfalls, limitations, and applications. *Space weather.* 20, e2121SW002989. doi:10.1029/ 2021sw002989

Mailyan, B., Munteanu, C., and Haaland, S. (2008). What is the best method to calculate the solar wind propagation delay? *Ann. Geophys.* 26, 2383–2394. doi:10. 5194/angeo-26-2383-2008

McComas, D. J., Blame, S. J., Barker, P., Feldman, W. C., Phillips, J. L., Riley, P., et al. (1998). Solar wind electron proton alpha monitor (SWEPAM) for the advanced composition explorer. *Space Sci. Rev.* 86, 563.

McPherron, R. L., Hsu, T.-S., and Chu, X. (2015). An optimum solar wind coupling function for the AL index. *JGR. Space Phys.* 120, 2494–2515. doi:10.1002/ 2014ja020619

Nemecek, Z., Durovcova, T., Safrankova, J., Nemec, F., Matteini, L., Stansby, D., et al. (2020b). What is the solar wind frame of reference? *Astrophys. J.* 889, 163. doi:10.3847/1538-4357/ab65f7

Nemecek, Z., Durovcova, T., Safrankova, J., Richardson, J. D., Simunek, J., and Stevens, M. L. (2020a). Non)radial solar wind propagation through the heliosphere. *Astrophys. J.* 897, L39. doi:10.3847/2041-8213/ab9ff7

Newell, P. T., Sotirelis, T., Liou, K., Meng, C. I., and Rich, F. J. (2007). A nearly universal solar wind-magnetosphere coupling function inferred from 10 magnetospheric state variables. *J. Geophys. Res.* 112, A01206. doi:10.1029/2006ja012015

Ohtani, S., Nose, M., Rostoker, G., Singer, H., Lui, A. T. Y., and Kakamura, M. (2001). Storm-substorm relationship: Contribution of the tail current to Dst. *J. Geophys. Res.* 106, 21199–21209. doi:10.1029/2000ja000400

Ridley, A. J. (2000). Estimations of the uncertainty in timing the relationship between magnetospheric and solar wind processes. *J. Atmos. Sol. Terr. Phys.* 62, 757–771. doi:10.1016/s1364-6826(00)00057-2

Ridley, A. J., and Kihn, E. A. (2004). Polar cap index comparisons with AMIE cross polar cap potential, electric field, and polar cap area. *Geophys. Res. Lett.* 31, L07801. doi:10.1029/2003gl019113

Sandahl, I., Lundstedt, H., Koskinen, H., and Glassmeir, K.-H. (1996). On the need for solar wind monitoring close to the magnetosphere. *Asp. Conf. Ser.* 95, 300.

Sckopke, N. (1966). A general relation between the energy of trapped particles and the disturbance field near the Earth. *J. Geophys. Res.* 71, 3125–3130. doi:10.1029/jz071i013p03125

Siscoe, G. L., McPherron, R. L., and Jordanova, V. K. (2005). Diminished contribution of ram pressure to Dst during magnetic storms. *J. Geophys. Res.* 110, A12227. doi:10.1029/2005ja011120

Sivadas, N., and Sibeck, D. G. (2022). Regression bias in using solar wind measurements. *Front. Astron. Space Sci.* 9, 924976.

Smith, C. W., Acuna, M. H., Burlaga, L. F., L'Heureux, J., Ness, N. F., and Scheifele, J. (1998). The ACE magnetic fields experiment space. *Space Sci. Rev.* 86, 613–632. doi:10.1023/a:1005092216668

Spearman, C. (1904). The proof and measurement of association between two things. *Am. J. Psychol.* 15, 72. doi:10.2307/1412159

Stauning, P. (2013). The polar cap index: A critical review of methods and a new approach. *J. Geophys. Res. Space Phys.* 118, 5021–5038. doi:10.1002/jgra.50462

Su, S.-Y., and Konradi, A. (1975). Magnetic field depression at the Earth's surface calculated from the relationship between the size of the magnetosphere and the Dst values. *J. Geophys. Res.* 80, 195–199. doi:10.1029/ja080i001p00195

Thomsen, M. F. (2004). Why Kp is such a good measure of magnetospheric convection. *Space weather.* 2, S11004. doi:10.1029/2004sw000089

Walsh, B. M., Bhakyapaibul, T., and Zou, Y. (2019). Quantifying the uncertainty of using solar wind measurements for geospace inputs. *JGR. Space Phys.* 124, 3291–3302. doi:10.1029/2019ja026507

Weimer, D. R., Ober, D. M., Maynard, N. C., Burke, W. J., Collier, M. R., McComas, D. J., et al. (2002). Variable time delays in the propagation of the interplanetary magnetic field. *J. Geophys. Res.* 107, SMP 29-3 1–SMP 29-15. doi:10.1029/2001ja009102

Xu, F., and Borovsky, J. E. (2015). A new four-plasma categorization scheme for the solar wind. *J. Geophys. Res. Space Phys.* 120, 70–100. doi:10.1002/2014ja020412