A 27 day persistence model of near-Earth solar wind conditions: a long lead-time forecast and a benchmark for dynamical models

Article

Published Version

SPACE WEATHER, VOL. 11, 225–236, doi:10.1002/swe.20040, 2013

A 27 day persistence model of near-Earth solar wind conditions: A long lead-time forecast and a benchmark for dynamical models

M. J. Owens,[1] R. Challen,[2] J. Methven,[2] E. Henley,[3] and D. R. Jackson[3]

Received 8 February 2013; revised 11 March 2013; accepted 14 March 2013; published 2 May 2013.

[1] Geomagnetic activity has long been known to exhibit approximately 27 day periodicity, resulting from solar wind structures repeating each solar rotation. Thus a very simple near-Earth solar wind forecast is 27 day persistence, wherein the near-Earth solar wind conditions today are assumed to be identical to those 27 days previously. Effective use of such a persistence model as a forecast tool, however, requires the performance and uncertainty to be fully characterized. The first half of this study determines which solar wind parameters can be reliably forecast by persistence and how the forecast skill varies with the solar cycle. The second half of the study shows how persistence can provide a useful benchmark for more sophisticated forecast schemes, namely physics-based numerical models. Point-by-point assessment methods, such as correlation and mean-square error, find persistence skill comparable to numerical models during solar minimum, despite the 27 day lead time of persistence forecasts, versus 2–5 days for numerical schemes. At solar maximum, however, the dynamic nature of the corona means 27 day persistence is no longer a good approximation and skill scores suggest persistence is out-performed by numerical models for almost all solar wind parameters. But point-by-point assessment techniques are not always a reliable indicator of usefulness as a forecast tool. An event-based assessment method, which focusses key solar wind structures, finds persistence to be the most valuable forecast throughout the solar cycle. This reiterates the fact that the means of assessing the "best" forecast model must be specifically tailored to its intended use.

Citation: Owens, M. J., R. Challen, J. Methven, E. Henley, and D. R. Jackson (2013), A 27 day persistence model of near-Earth solar wind conditions: A long lead-time forecast and a benchmark for dynamical models, *Space Weather*, *11*, 225–236, doi:10.1002/swe.20040.

1. Introduction

[2] Space weather poses a threat to a range of ground- and space-based technologies [*Feynman and Gabriel*, 2000; *Hapgood*, 2011]. In order to mitigate such space-weather effects, reliable forecasts of the near-Earth solar wind conditions are required, preferably with lead times significantly longer than the 20–40 minutes provided by spacecraft observations at the first Lagrange point, L1. The current state-of-the-art numerical forecast models use the observed photospheric magnetic field to estimate the near-Sun solar wind conditions and propagate structures to near-Earth space [e.g., *Owens et al.*, 2008, and references therein]. The majority of large geomagnetic storms are driven by fast coronal mass ejections (CMEs) [e.g., *Tsurutani et al.*, 1988], which are inherently episodic in nature. As pre-eruption forecasting of CMEs is currently not possible, forecasts of CME-driven storms are also limited to lead times of 2–4 days, a CME's typical Sun-to-Earth travel time [*Gopalswamy et al.*, 2001]. The ambient solar wind, however, can still be geoeffective in its own right, particularly in regions of fast and slow solar wind interaction and subsequent compression [e.g., *Richardson et al.*, 2002; *Tsurutani et al.*, 2006, and references therein]. At solar minimum and during the declining phase of the solar cycle, the ambient solar wind evolves sufficiently slowly over a solar rotation that solar wind interaction regions approximately corotate with the Sun [corotating interaction regions, CIRs. e.g., *Pizzo*, 1991; *Gosling and Pizzo*, 1999] and hence the pattern of the near-Earth solar wind approximately

[1]Space Environment Physics Group, Department of Meteorology, University of Reading, Reading, UK.
[2]Department of Meteorology, University of Reading, Reading, UK.
[3]Met Office, Exeter, Devon, UK.

Corresponding author: M. J. Owens, Space Environment Physics Group, Department of Meteorology, University of Reading, Earley Gate, PO Box 243, Reading RG6 6BB, UK. (m.j.owens@reading.ac.uk)

1542-7390/13/10.1002/swe.20040

repeats every synodic solar rotation (T_{SYN} = 27.27 days). While this recurrent form of geomagnetic activity is generally weaker than the CME-initiated activity, it can still affect technology through significant energization of the terrestrial radiation belts [*McPherron and Weygand*, 2006] and effects on the ionosphere [*Lei et al.*, 2008].

[3] While geomagnetic activity [e.g., *Chree and Stagg*, 1928; *Bartels*, 1932, 1934] and solar wind structure [e.g., *Sargent*, 1985; *Diego et al.*, 2010] have long been known to exhibit approximately 27 day periodicity, the effective use of a 27 day persistence model as a forecast tool, however, requires the performance and uncertainty of the persistence model to be fully characterized. Another important application of a solar wind persistence model is as a reference forecast for more sophisticated techniques [e.g., *Siscoe et al.*, 2004]. Indeed, in the development of meteorological forecasting, models were generally regarded as having real forecast skill once they "beat" a daily persistence forecast, i.e., a simple forecast that the weather tomorrow will be the same as the weather today [e.g., *Siscoe*, 2007]. Similarly, for radiation belt models, 27 day persistence is used as a benchmark against which physical and empirical models are tested for skill [*Baker et al.*, 1990].

[4] In this paper we first develop the persistence model by examining autocorrelations in the time series of solar wind observations (section 2). Then we test the 27 day prediction forecast of various near-Earth solar wind parameters as a function of solar cycle (section 3), before using the persistence model to provide a bench mark for more sophisticated dynamical models (section 4).

2. Autocorrelation in the Solar Wind Time Series

[5] This study uses the 1 hour OMNI data set of near-Earth solar wind conditions from the Space Physics Data Facility [*King and Papitashvili*, 2005]. The data are in Geocentric Solar Ecliptic (GSE) coordinates, with the x-direction pointing towards the Sun, the y-direction in the ecliptic plane against the direction of Earth's orbital motion and the z-direction (approximately parallel to the solar rotation axis) completing the right-hand set.

[6] We begin by considering the data set in its entirety, from 1963 to 2012, some 432,097 data points. Figure 1 shows the autocorrelation (r_A) in the OMNI time series for increasing time delay and for different solar wind parameters. For the radial solar wind speed (V_X, thick black line), radial magnetic field component (B_X, green), proton density (N_P, cyan) and proton temperature (T, blue), there is clear predictability in the data at not only the solar synodic rotation period, $T_{SYN} \approx 27.27$ days, but also at the second and third harmonics, albeit with decreasing amplitude. Given the large number of data points, these correlations are all highly statistically significant. The alternating inward and outward structure of the ecliptic magnetic field means B_X is in anticorrelation for delays approximately equal to $T_{SYN}/2$. B_Y, not shown, has a very similar behavior to B_X, owing to the predominance of the Parker-spiral magnetic field [*Parker*, 1958]. The out-of-ecliptic magnetic

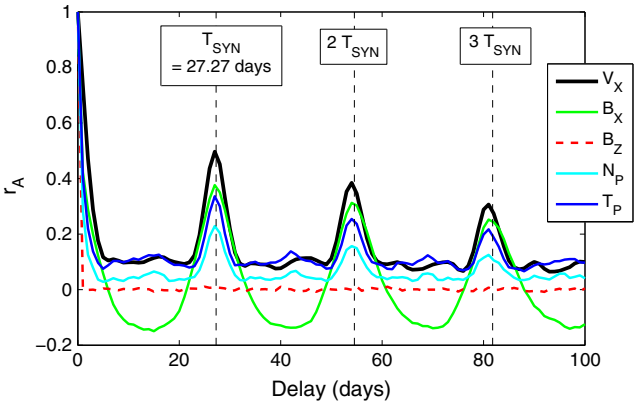

Figure 1. Autocorrelation for various solar wind parameters at increasing autocorrelation delay using OMNI data of near-Earth observations at 1 hour resolution over the period 1963–2012. With the exception of B_Z, all parameters show significant correlation around the solar synodic rotation period, T_{SYN}, of 27.27 days. Multiples of T_{SYN} show autocorrelation peaks, although the amplitude decreases with increasing delay.

field component (B_Z, dashed red line), arguably the most important parameter for determining the geoeffectiveness of the solar wind [*Dungey*, 1961], however, shows near-zero autocorrelation at all time delays. Note that the large sample size means that even marginally non-zero autocorrelations, such as the observed value $r_A(B_Z) = 0.007$ at a delay of T_{SYN}, are highly statistically significant. However, this value is so small as to be of no practical use for forecasting, as discussed further below.).

[7] The next step is to quantify how the autocorrelation changes over the solar cycle. The top and middle panel of Figure 2 show V_X and B_X autocorrelation as a function of time and autocorrelation delay. The OMNI data are split into yearly blocks of approximately 8760 data points. Only 1975 to 2011 are shown, as there is insufficient data coverage to separate data into yearly chunks before this period. The bottom panel shows sunspot number, R, for reference. Both parameters show autocorrelation around the 27 day delay throughout the solar cycle. For V_X, however, autocorrelation is stronger close to solar minimum and during the declining phase of the solar cycle. This effect is less pronounced for B_X. There are also differences between solar cycles, with the minimum and rise phase of cycle 23 (approximately 1996–2000) showing less 27 day autocorrelation than the equivalent period of cycle 22 (approximately 1986–1989). The generally low autocorrelation at solar maximum is likely the result of the solar corona (and resulting solar wind properties at 1 AU) being at its most dynamic and evolving significantly over a solar rotation. Similar plots for proton density and temperature (not shown) are very similar. B_Z, however, displays no

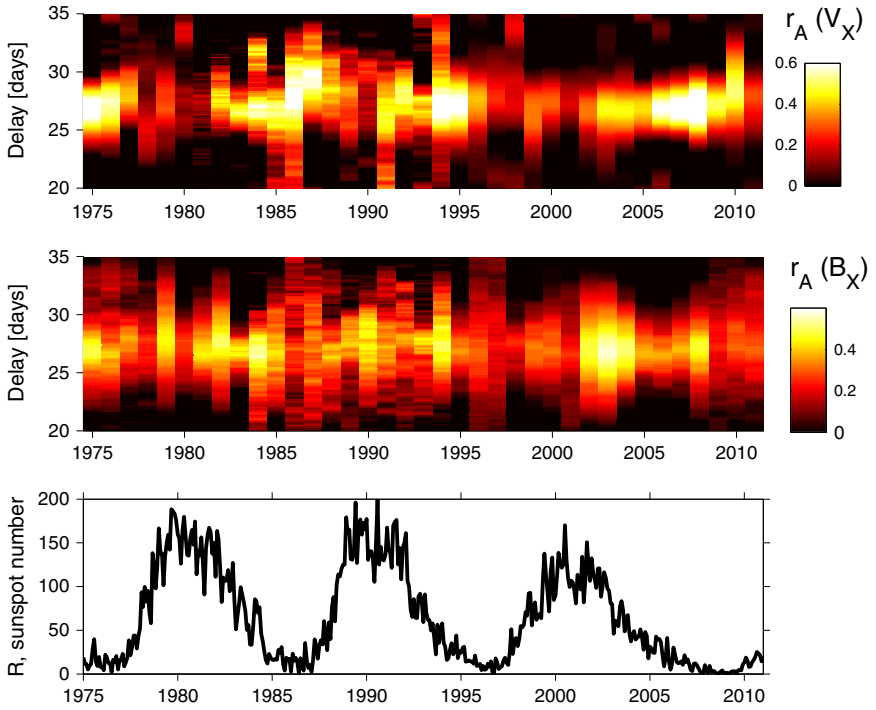

Figure 2. The top and middle panels show the V_X and B_X autocorrelation, respectively, as a function of time through solar cycle and delay (in 1 hour intervals). The bottom panel shows the sunspot number, R, for reference. For V_X (and to a lesser extent B_X), the autocorrelation around 27 days delay is stronger at solar minimum and during the declining phase of the solar cycle than at solar maximum. There does not seem to be an obvious solar-cycle variation in the delay which results in maximum correlation.

apparent autocorrelation at any delay or at any point in the solar cycle.

[8] The B_X and V_X autocorrelation peaks centered around 27 days in Figure 2 are reasonably broad, suggesting the optimum delay for a persistence forecast may change with time. Using the yearly V_X data from 1975–2011, we find the median value of the maximum autocorrelation delay is 27.145 days, slightly faster than the solar synodic rotation period (27.27 days). As hourly data are used throughout this study, we use delay closest to the nearest whole hour, 27.125 days, in the remainder of this study. Note that *Mursula and Zieger* [1996] performed a spectral analysis of solar wind speed time series and found the 13.5 periodicity to be stronger than the 27.27 day periodicity. The 13.5 day B_X autocorrelation trough in Figure 1 is much weaker than the 27 day peak, suggesting that this $T_{SYN}/2$ structuring is not always present or else shifts in phase, making it less useful than T_{SYN} persistence as the basis of a simple forecast.

3. Persistence as a Forecast Tool

3.1. Quantifying Forecast Skill

[9] A widely used technique for testing the quality of a forecast is to calculate the mean square error (MSE) between the observation time series and the model time series. If the observed parameter values are denoted by $X_{OBS}(t)$ and the model parameter values by $X_{MOD}(t)$, then the MSE is given by

$$MSE = \frac{1}{N}\sum_{t=1}^{N}[X_{OBS}(t) - X_{MOD}(t)]^2 \quad (1)$$

where N is the number of observations.

[10] Obviously, lower MSE is generally indicative of better model performance. However, raw MSE does not differentiate between periods when the model is performing badly and periods which are inherently difficult to predict, such as times of particularly enhanced variability. Thus it helps to estimate the "skill" of a model by comparing with another "baseline," or "reference," forecast method [e.g., *Roebber*, 1998; *Siscoe et al.*, 2004; *Spence et al.*, 2004]. The skill of the model being tested is then defined as follows:

$$Skill = 100\left(1 - \frac{MSE_{MOD}}{MSE_{REF}}\right) \quad (2)$$

where MSE_{MOD} is the mean square error between the model and observations and MSE_{REF} is the mean square error between the reference model and observations.

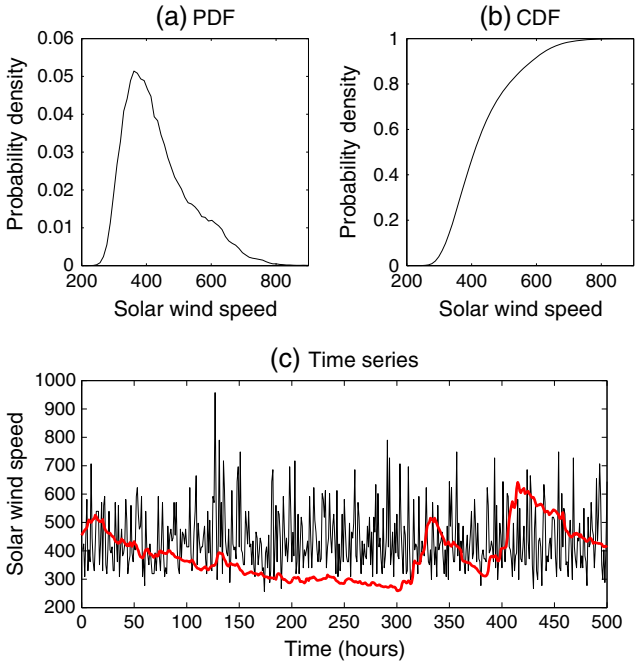

Figure 3. (a) Probability distribution functions for V_X using 1 hour OMNI data over the period 1963–2012. (b) The associated cumulative distribution function (CDF). (c) A V_X time series created from the CDF using a random number generator (black). This is the random reference model against which skill is calculated in the remainder of this study. For comparison, a 500 hour interval of the observed solar wind speed is shown in red.

The range of skill is from -∞ to +100. A skill score of 100 would indicate a perfect forecast. A negative skill score would imply that the model being tested performs worse than the reference model. While skill could in principle be computed on a point-by-point basis then averaged up to the required time resolution, this is undesirable, as the reference model can sometimes make a perfect forecast for a single point, resulting in a $-\infty$ in the time series. Instead, the MSE is averaged over the required interval (e.g., for hourly data, a month or a year), then the skill is computed from this.

[11] Persistence models actually provide a very useful reference forecast against which more complex forecasting schemes can be tested, as outlined in section 4. However, testing the skill of a persistence model itself requires another reference model, discussed in the next section.

3.2. Generating a Reference Forecast

[12] The simplest reference against which the persistence forecast could be tested is climatology, in which the reference forecast for a given parameter is simply the average or median value of that parameter. A reference forecast which is completely unchanging with time, however, is not desirable for the purpose of computing skill scores [e.g., *Murphy*, 1993]. Thus, we also use the climatological variability, creating a reference model has the same bulk statistical properties as the observations, but which is random in time, with zero autocorrelation on all time scales when a long enough interval is considered.

[13] This random reference model is generated in the following manner. First, a probability distribution function (PDF) is created for each solar wind variable using the 1 hour OMNI data for the whole 1963–2012 interval. Figure 3a shows the PDF for V_X, which typically varies between approximately 250 and 850 kilometers per second, with a median solar wind speed of 413 kilometers per second, a typical value for the slow solar wind, and a slightly higher mean value around 436 kilometers per second, owing to fast solar wind, encountered relatively infrequently in the ecliptic plane. The standard deviation of the time series is 103 kilometers per second. The PDF is then converted to a cumulative distribution function (CDF). Figure 3b shows the V_X CDF. We use a random number generator to create a series of numbers between 0 and 1 with uniform probability. These are converted to corresponding solar wind speeds using the CDF. An example time series of 500 points is shown as the black line in Figure 3c. This particular 500 point series has a median value of 414 kilometers per second and a mean of 440 kilometers per second, with a standard deviation of 108 kilometers per second. Thus, the statistical

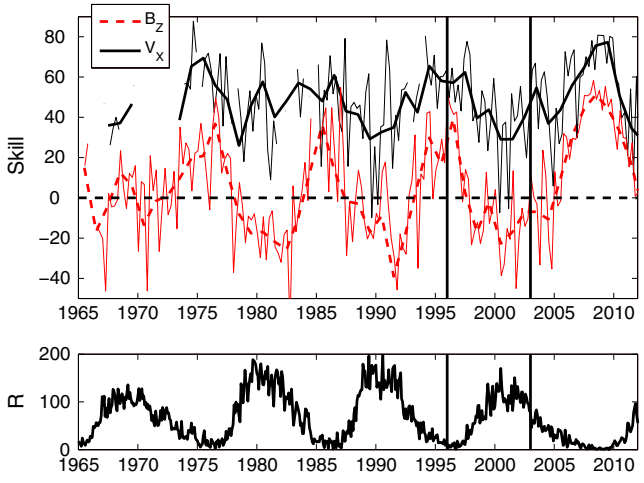

Figure 4. The top panel shows the skill of the persistence forecast relative to the random reference model for V_X (black) and B_Z (red dashed line) as a function of time. Thick (thin) lines show 1 year (3 month) averages. Mean-square error is computed from hourly data, then averaged to the required resolution before computing the skill score. The bottom panel shows sunspot number, for reference. There are clear solar cycle trends in the skill of the persistence forecast. Solid vertical lines indicate the period over which we have numerical model data, analyzed in section 4.

Table 1. The Skill of the Persistence Forecast Relative to the Random Reference Model for Various Solar Wind Parameters Over the Period 1963 to 2012[a]

	Skill (Persistence Model)		
	All Data	Solar Min (R<50)	Solar Max (R>50)
B_X	37 ± 12	45 ± 11	39 ± 11
B_Z	5 ± 22	22 ± 17	–10 ± 12
$\|B_Z\|$	12 ± 20	26 ± 19	–1 ± 10
V_X	48 ± 13	58 ± 11	40 ± 9
N_P	22 ± 19	31 ± 18	16 ± 18
T_P	33 ± 18	45 ± 16	24 ± 15
E_Y	8 ± 28	27 ± 23	–8 ± 21

[a]Data have been split into solar minimum and solar maximum conditions using a sunspot number threshold of 50.

properties are very similar to those observed; however, the time series lacks the autocorrelation which results from coherent large-scale solar wind structures. The random reference model contains a single 1 hour interval of ∼ 950 kilometers per second solar wind surrounded by ∼ 350 kilometers per second intervals, whereas high speed streams typically last many hours to days in the ecliptic, as is illustrated by the red line in Figure 3c, which shows a 500 hour interval of the observed solar wind speed from January 2011.

3.3. Persistence Skill

[14] The skill in the persistence forecast is computed by comparing MSE_{MOD}, the mean-square error in the 27.125 day persistence forecast from the observed solar wind parameters, with MSE_{REF}, the equivalent error in the random reference model. The skill score is computed from mean MSE values, determined by averaging hourly MSE over 3 month or 1 year intervals. The top panel of Figure 4 shows the persistence forecast skill score for V_X (black) and B_Z (red dashed line), with yearly means shown as thick lines and 3 month means as thin lines. The bottom panel shows sunspot number: There are clear solar cycle variations in the persistence skill for both parameters. Maximum forecast skill generally occurs around solar minimum, as might be expected owing to the more steady-state conditions in the solar corona at this time: both V_X and B_Z have higher skill than the random reference model at this time. The positive B_Z skill of the persistence model, despite near-zero autocorrelation in B_Z throughout the solar cycle, is likely a result of a limitation of the random reference model: By using the CDF over the entire OMNI dataset, the random reference model slightly overestimates the B_Z variability at solar minimum, while the persistence model predicts the correct variability, although not necessarily the correct B_Z structures. Thus, the

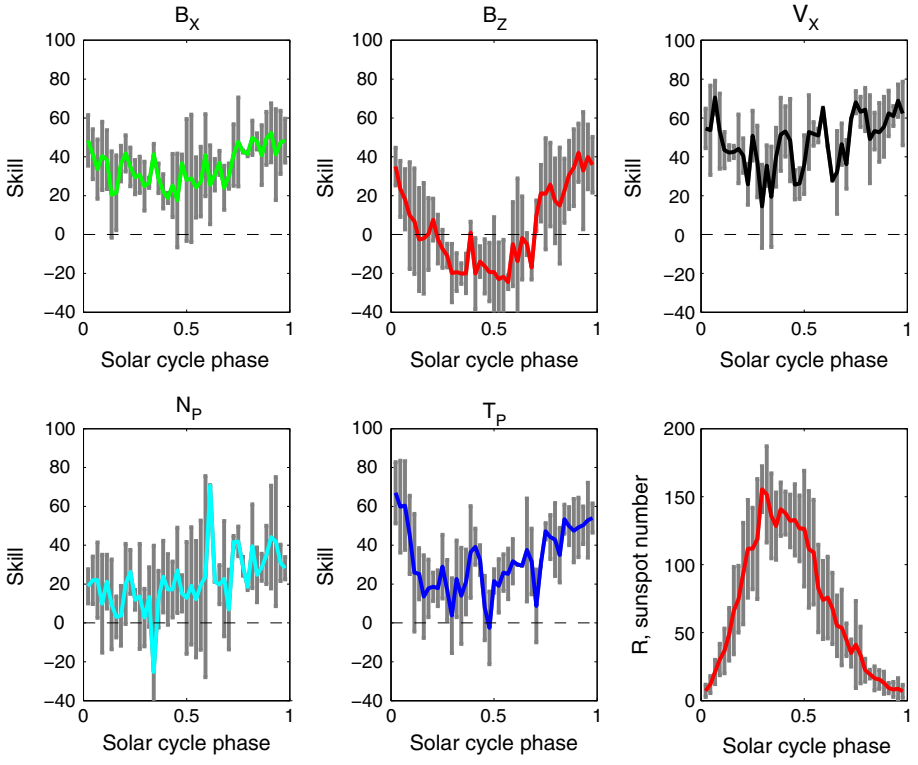

Figure 5. A super-posed epoch analysis of the 4.5 solar cycles of OMNI data. Persistence forecast skill, relative to the random reference model, is shown for various solar wind parameters as a function of solar cycle phase. Grey bars show one standard error on the mean.

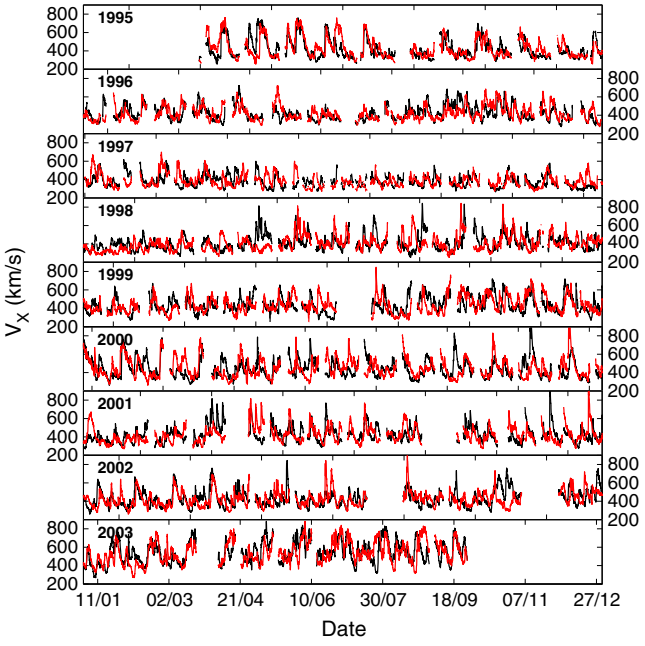

Figure 6. Time series of observed (black) and persistence model (red) solar wind speed.

persistence model achieves skill in predicting B_Z not because of autocorrelation, but because it correctly captures the variability in B_Z over the solar cycle. This issue is discussed further in section 4 and highlights the importance of using an appropriate reference model to benchmark forecasts.

[15] At solar maximum, the skill of the V_X persistence forecast relative to the reference model is still much greater than zero. As shown in Figure 2, the solar wind stream structure at solar maximum still exhibits some 27 day recurrence. For B_Z, however, the persistence model has a lower skill than the random reference model, again the result of neither model correctly forecasting the B_Z structures, but the random reference model underestimating the variability and thus producing a larger MSE.

[16] Table 1 summarizes the skill of the persistence model relative to the reference model, for a range of solar wind parameters. In addition to those shown in Figure 4, we consider $|B_Z|$ and the dawn-to-dusk electric field, $E_Y = V_X B_Z$, as this approximates the coupling efficiency between the heliospheric and magnetospheric magnetic fields [*Dungey*, 1961]. In order to determine the solar cycle effect, a sunspot number (R) threshold is applied. $R = 50$ approximately bisects the data set into solar maximum ($R > 50$) and solar minimum ($R \leq 0$), and shows that persistence skill for all solar wind parameters at solar minimum is significantly higher than at solar maximum.

[17] Of course, a $R = 50$ cut off is entirely arbitrary and cycle-to-cycle variations in peak sunspot magnitudes means it applies at different phases of the solar cycle. To better investigate the solar cycle-dependence of persistence skill, we use the solar minimum times defined by *Owens et al.* [2011] to perform a super-posed epoch analysis (also known as a "composite analysis") of persistence skill as a function of solar cycle phase. Phase is defined as 0 and 1 at successive solar minima, and solar maximum corresponds to approximately 0.4, as shown in the bottom-right panel of Figure 5. The other panels show the results for the solar wind parameters considered in Figure 1. Grey bars show one standard error on the mean. All parameters show some solar cycle variation. N_P shows low values and the least variation, possibly because density enhancements at the interfaces between fast and slow solar wind streams, which are common throughout the solar cycle, tend to be short in duration and thus even very small timing offsets are likely to produce large errors. There remains significant skill in persistence forecasts of B_X and V_X even at solar maximum. This is not true of B_Z, where the skill is negative at solar maximum. From these plots, we conclude that a persistence forecast of the solar wind is of most use during solar minimum and during the declining phase of the solar cycle. *Diego et al.* [2010] formed similar conclusions from an analysis of periodicities in geomagnetic time series.

3.4. High Speed Streams

[18] Computing mean-square error-based skill scores is a useful means of assessing a model's performance. In many cases, however, it does not directly capture a model's usefulness as a forecast tool. One approach to this problem is to identify key "events" of interest in the data and ask how well the model does at reproducing the characteristics of those events. For the solar wind, the key events for geoeffectiveness are high speed solar wind streams and southward turnings of the heliospheric magnetic field. As the latter are predominantly associated with transient events, we instead look at high speed streams.

[19] *Owens et al.* [2005] defined a high speed enhancement (HSE) as an interval of solar wind lasting at least 2 days, where the solar wind speed is always at least

Table 2. A Comparison of High Speed Enhancements (HSEs) in the Observed and Persistence Model Solar Wind Speed Time Series Over the 1963–2012 Period[a]

	High Speed Enhancements (Persistence Model)		
	All Data	Solar Min ($R < 50$)	Solar Max ($R > 50$)
Hits	788	412	352
Misses	208	97	104
False alarms	223	111	105
Threat score	0.646	0.665	0.627
Bias	1.02	1.03	1.00
$< \lvert\Delta T\rvert >$ (days)	1.59	1.50	1.69
$< \Delta T >$ (days)	–0.12	0.07	–0.28
$< \lvert\Delta V\rvert >$ (km/s)	91.2	80.3	101.4
$< \Delta V >$ (km/s)	0.39	0.47	–0.10

[a]"Hits" are HSEs present in both data sets, "misses" appear only in observations, while "false alarms" appear only in the model data. For the hits, ΔT is the observed minus model HSE arrival time, while ΔV is the observed minus model peak speed associated with the HSE

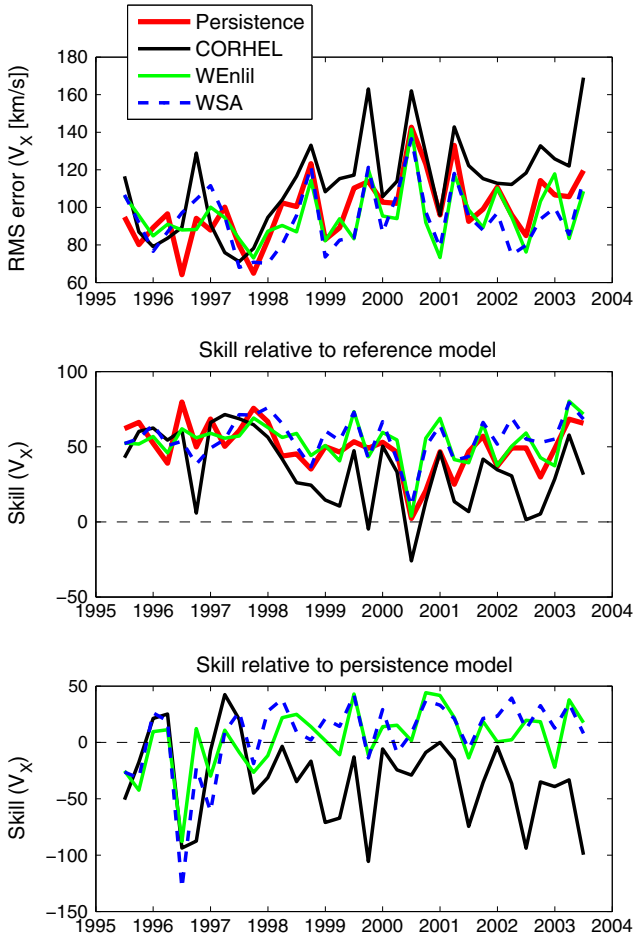

Figure 7. The skill of various model predictions of solar wind speed, relative to the reference model, as a function of time. Solid black and green lines show CORHEL and WEnlil, respectively, while the dashed blue line shows WSA. The persistence model is shown as a thick red line.

100 kilometers per second above that before the event. Multiple events within 2 days of each other are considered to be part of the same HSE. The 2 day requirement means that many fast CMEs, which typically have durations of order 1 day, are excluded from the analysis, which is designed to focus on the ambient solar wind structure. *Owens et al.* [2005] also outlined a framework for associating and comparing the HSEs in spacecraft and model time series so as to compute the number of correctly predicted events ("hits"), as well as the number of HSEs present in the observations but not in the model ("misses") and vice versa ("false alarms"). Note data gaps in the time series affect the number of missed and false alarm HSEs recorded: to avoid contamination by data gaps, those occurring within 2 days of a data gap are removed from the study.

[20] Applying the HSE selection criteria to the full OMNI data set, from 1965 to present, identifies 1030 HSEs. Cross-referencing the OMNI and persistence model HSE lists gives 788 correct predictions (hits), with an average peak speed of 592 kilometers per second. 208 observed HSEs were missed by the persistence model. These had a lower average peak speed of 543 kilometers per second. Conversely, there were 223 false alarms with an average peak speed of 545 kilometers per second. (Note that if a missed or false alarm HSE falls within 2 days of a data gap, it can be removed from the study.) The threat score (TS, or "critical success index") encapsulates a model's ability to forecast key events. Thus, in this instance,

$$TS = \frac{N_{HIT}}{N_{HIT} + N_{MISS} + N_{FALSE}} \quad (3)$$

where N_{HIT}, N_{MISS} and N_{FALSE} are the number of hit, miss and false alarm events, respectively. For the persistence model, $TS = 0.646$. Thus, at a very basic level, we can say that a persistence forecast gets the solar wind stream structure right more often than wrong. This figure can only really be put into context by comparison with TS values from other models, as in section 4.

[21] We can also compute the bias to under (< 1) or over-predicting (> 1) HSEs:

$$Bias = \frac{N_{HIT} + N_{FALSE}}{N_{HIT} + N_{MISS}} \quad (4)$$

[22] This parameter is less a metric of the model's forecast ability and more a diagnostic of how the model can be improved. For the persistence model, the bias is 1.02. Again, this value will be put into better context in section 4, but it essentially means there is almost no systematic bias to over/under-estimation of HSEs. This should be expected of a 27 day persistence model: A transient HSE, such as a flow resulting from a fast CME, would be missed by the persistence forecast as there was no equivalent flow 27 days previously. However, the following solar rotation, the persistence forecast assumes this HSE will repeat and so creates a false alarm. (A clear example of this behavior is shown in December 2001 in Figure 6. A very fast CME creates a 900 kilometers per second HSE, which is missed by the persistence forecast, but which then generates a false alarm 27 days later.) If transient features can be differentiated from corotating features, such as via anomalously low temperatures [*Cane and Richardson*, 2003], then it may be possible to reduce the number of false alarms by removing these transients from the persistence forecast for the following rotation. This would increase the model's TS, but also introduce a significant bias, as the number of missed (unforeseen) HSEs will remain unchanged.

[23] The average absolute error in the HSE arrival time, $< |\Delta T| >$, was 1.59 days, while the average timing error, $< \Delta T >$, was near zero at –0.12 days, suggesting no systematic offset in the predicted HSE arrival times. Looking at the maximum solar wind speed in the HSEs,

Table 3. Properties of the Observed and Model Time Series Over the 1995–2003 Period[a]

	B_X(nT)	$\|B_X\|$(nT)	B_Z(nT)	$\|B_Z\|$(nT)	$\|V_X\|$(km/s)	n_P(cm^{-3})	T_P(10^{4}K)
Observations							
< X >	–0.06	3.13	0.007	1.98	432.2	6.90	9.91
STD	3.76	2.09	2.99	2.23	98.8	5.38	8.79
Persistence							
< X >	–0.03	3.13	0.007	2.00	431.7	6.93	9.81
STD	3.77	2.09	3.02	2.26	98.5	5.42	8.72
RMS error	3.22	2.12	2.95	1.94	75.4	4.38	6.84
r_L	0.36	0.11	0.008	0.06	0.47	0.20	0.30
Skill (ref)	41.9	12.6	–1.02	2.21	52.9	27.2	41.4
CORHEL							
< X >	–0.12	2.20	–0.008	0.17	389.1	10.1	3.14
STD	2.50	1.20	0.22	0.14	101.0	6.87	1.79
RMS error	2.72	1.88	1.98	1.83	89.9	6.14	7.17
r_L	0.41	0.08	0.02	0.03	0.42	0.12	0.30
Skill (ref)	58.5	31.4	54.4	13.1	33.0	–42.6	35.7
Skill (pers)	28.6	21.6	54.9	11.2	–42.2	–95.9	–9.72
WEnlil							
< X >	–0.07	2.15	–0.002	0.16	430.8	7.68	3.68
STD	2.26	0.68	0.19	0.10	74.2	3.70	1.23
RMS error	2.62	1.78	1.98	1.84	75.7	4.35	6.83
r_L	0.45	0.03	0.02	0.02	0.40	0.12	0.23
Skill (ref)	61.4	37.8	54.4	12.54	52.4	28.5	41.8
Skill (pers)	33.6	28.9	54.9	10.6	–0.94	1.80	0.60
WSA							
< X >	–0.18	3.43	0	0	409.8	9.02	6.29
STD	3.49	0.68	0	0	84.3	1.70	4.03
RMS error	2.64	1.69	1.98	1.98	71.2	4.49	5.66
r_L	0.53	0.08	N/A	N/A	0.51	0.20	0.37
Skill (ref)	60.9	44.3	54.4	–1.68	58.0	23.6	60.0
Skill (pers)	32.8	36.2	54.9	–3.98	10.8	–4.9	31.7

[a]Skill scores are relative to the random reference model or persistence model, as indicated. We also list the root mean-square (RMS) error and the linear cross correlation coefficient (r_L) relative to the observations.

the average absolute error, < $|\Delta V|$ >, is 90.2 kilometers per second, while the average error, < ΔV >, is 0.39 kilometers per second, again suggesting no systematic offset. These values are shown in Table 2.

[24] Table 2 also shows how the properties of HSEs vary with solar cycle, splitting solar minimum and maximum using the R = 50 threshold. (Note this division fragments the time series so that more HSEs can become associated with data gaps.) The persistence model performs slightly better during solar minimum than solar maximum, although the difference is not as apparent as with the skill scores in the previous section. Both the HSE timing error and peak speed error increase slightly from minimum to maximum, although again, no systematic offset exists.

4. An Initial Benchmarking of Dynamical Models

[25] Aside from providing a long lead-time forecast of solar wind conditions, the most obvious use for a persistence forecast is to provide a suitable benchmark for the more advanced, physics-based, numerical solar wind models. In this section we make a comparison between the persistence model and three numerical forecast schemes for which reasonably long periods of data are available. These data sets are the same as used in [*Owens et al.*, 2008], allowing easy comparison of the results. It should be noted that the models, explained further below, have undergone significant development from the early versions used here [e.g., *McGregor et al.*, 2008; *McGregor et al.*, 2011] and so the following should not be considered a proper benchmarking exercise. Instead, the aim here is to illustrate the process and its associated pitfalls, rather than to determine the "best" forecast model (indeed, we conclude that the "best" model depends strongly on the specific requirements of the forecaster). These validation techniques could equally be applied to other solar wind forecasting techniques [e.g., *Manchester et al.*, 2004; *Detman et al.*, 2006].

[26] All three dynamical models used in this study are driven by observations of the photospheric magnetic field.

[27] 1. Wang-Sheely-Arge (WSA) uses a potential-field model of the corona [e.g., *Schatten et al.*, 1969; *Altschuler and Newkirk*, 1969] to derive solar wind speed from empirical relations to coronal magnetic structure, which is then ballistically propagated to Earth, accounting for stream interactions [*Arge and Pizzo*, 2000; *Arge et al.*, 2003].

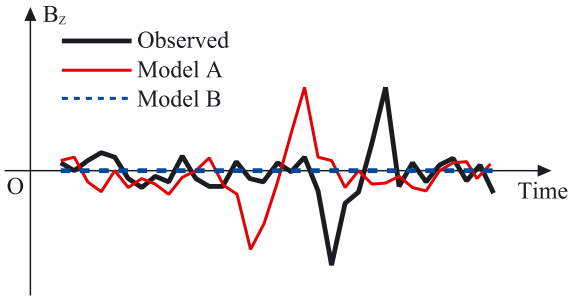

Figure 8. A sketch of an observed B_Z event, which involves a short-duration bipolar flip from negative to positive values (thick black line). Model A (red line) captures the correct magnitude and form of the B_Z event, but the timing is wrong. Model B (blue dashed line) predicts zero B_Z throughout the entire interval. Point-by-point analyses, such as mean square error or correlation, will find Model B to have more skill than Model A. A forecaster, however, may find more value in Model A, if the magnitude of the timing error is acceptable. In this instance an event-based skill score may be more appropriate.

[28] 2. WSA-Enlil (WEnlil) [e.g., *McGregor et al.*, 2011] uses the same method to characterize the solar wind conditions at the top of the corona, but the solar wind propagation to Earth is handled by the numerical magnetohydrodynamic (MHD) "Enlil" model [*Odstrcil*, 2003, and references therein]. Finally,

[29] 3. CORHEL [*Odstrcil et al.*, 2004] comprises the MAS MHD model of the corona [*Linker et al.*, 1999; *Mikic et al.*, 1999; *Riley et al.*, 2001] coupled to the Enlil solar wind model.

[30] In forecasting situations, these models use photospheric magnetic fields updated on a daily basis as the Sun rotates and new Carrington longitudes become visible on the solar disc observed from Earth. The lead time for such forecasts would be the solar wind travel time of 2–5 days. However, the "hindcasts" used in this study are based upon complete "Carrington maps" of the photospheric field, which take a full 27.27 days to compile. Consequently, the lead time is difficult to define, but is probably best considered to be 2–4 days, but with the caveat that the forecasts are between 0 and 27 days "old."

4.1. Skill Scores

[31] Figure 6 shows the observed (black) and persistence model (red) solar wind speed in yearly stack plots, which can be compared to equivalent plots for the three dynamical models, in *Owens et al.* [2008, Figures 2–4] . As in *Owens et al.* [2008], the interval under consideration spans the period 1995–2003, shown as the vertical black lines in Figure 4. All data are at 1 hour resolution and data gaps in any of the model or observed time series are removed from all time series. The agreement between the persistence model and observed solar wind speed is very good for 1995 and the first half of 1996, but gradually decreases through solar maximum, as expected.

[32] Figure 7 shows 3 month averages of the skill of the various solar wind speed predictions relative to the reference model. Solid black and green lines show CORHEL and WEnlil, respectively, while the dashed blue line shows WSA. The persistence model is shown as a thick red line. It provides one of the better predictions until 1998, when the dynamic corona means the skill of the 27 day advance prediction drops slightly below that of the 3 day advance WSA and WEnlil predictions.

[33] Table 3 summarizes the properties of the observed and model solar wind parameters over the 1995–2003 period. In addition to skill scores, it quotes the root mean-squared (RMS) error between modeled and observed time series, as well as their (unlagged) linear cross correlation, r_L. It raises a number of points about the value of MSE- (or correlation-) based skill scores. The most striking example is for B_Z. The persistence time series is simply a time-lagged version of the observed time series, so obviously, it has near-identical statistical properties, in terms of the mean and standard deviation (minor differences arise as the lagged time series covers an interval which differs from the observed time series by twice the lag time). The WSA prediction for B_Z, on the other hand, is exactly zero at all times, so the mean and standard deviation are exactly zero (and calculating the cross correlation is meaningless). The mean observed value of B_Z is indeed very close to zero, but the short-lived excursions to significant negative or positive values are critical to space weather forecasting. However, the nature of the B_Z fluctuations mean that the skill of the WSA B_Z prediction is higher than that of the reference and persistence models. This is explained schematically in Figure 8: A point-by-point analysis such as MSE or cross correlation penalizes a mistimed event twice, whereas a forecast predicting no event only gets penalized once. Clearly, there would be little value to a space-weather forecaster in the WSA model of B_Z, despite its apparent "skill." Whether the persistence model would have any value to a forecaster cannot really be determined by such point-by-point analysis.

[34] A similar situation, although less extreme, exists for solar wind speed predictions. The WSA model has the highest skill over the whole interval, yet it underestimates

Table 4. A Comparison of High Speed Enhancements (HSEs) in the Observed and Model Solar Wind Speed Time Series Over the 1995–2003 Period

	Persistence	WSA	CORHEL	WEnlil
Hits	209	163	146	110
Misses	61	105	122	156
False alarms	71	31	34	20
Threat score	0.613	0.545	0.483	0.385
Bias	1.04	0.72	0.67	0.49
$< \lvert\Delta T\rvert >$ (days)	1.71	1.73	1.75	2.14
$< \Delta T >$ (days)	–0.17	0.26	0.45	0.75
$< \lvert\Delta V\rvert >$ (km/s)	91.6	88.1	94.7	100.1
$< \Delta V >$ (km/s)	1.04	58.1	52.8	76.4

the variability in V_X by roughly 15%, suggesting a "flatter" solar wind speed time series than that observed (although some of this reduced variability will also be due to the WSA time series not including transient HSEs). The HSE analysis outlined in section 3.4 provides an alternative method for assessing the forecast value of a model of solar wind speed. The results are outlined in the next section. If B_Z "events," rather than V_X events, are the critical feature needed by a forecaster, then an equivalent scheme which identifies and compares the requisite B_Z structures should be employed.

4.2. High Speed Streams

[35] Applying the HSE criteria outlined in section 3.4 to the 1995–2003 interval, 276 HSEs were identified in the OMNI data, with an average peak speed of 576 kilometers per second. The persistence model correctly identified 209 of these streams, while 61 were missed (the remainder were associated with data gaps). This is significantly higher than any of the dynamical models, which miss almost twice as many events. However, it should also be noted that the persistence model generates 71 false alarms, twice as many as the dynamical models. These findings are in agreement with Table 3, which shows that the variability of the solar wind speed for the persistence model is significantly higher than that of the dynamical models. A significant fraction of these false alarms are likely to be due to long-duration fast CMEs (or multiple interacting CMEs), which can pass through the 2 day filter and generate false alarms when they do not recur on the following solar rotation.

[36] The magnitude of the average HSE timing error is comparable across all models, although the dynamical models do seem to exhibit a systematic bias for late arrival of HSEs. This is in agreement with the finding that the dynamical models underestimate both the average solar wind speed and the peak speed within HSEs, although it could also result from the details of the coupling between the coronal and solar wind portions of the models.

5. Discussion and Conclusions

[37] Geomagnetic activity has long been known to exhibit approximately 27 day periodicity, resulting from solar wind structures recurring each solar rotation. In this paper, we have investigated how well a 27 day persistence forecast of the solar wind works in practice, illustrating how this could be used as a tool for benchmarking more sophisticated models and determining the circumstances under which persistence forecasts may be useful in their own right. It is worth remembering that persistence gives a forecast with a 27 day lead time, much longer than the dynamical models, which typically have lead times around 2–4 days. So even when dynamical models perform better, a persistence forecast may still have value in situations where long-lead times are essential.

[38] Using the entirety of the OMNI data set of near-Earth solar wind properties from 1963 to 2012, we found the autocorrelation peaked at 27.125 days, slightly shorter than the 27.27 day synodic rotation period of the Sun. This was used as the basis for the persistence model in the remainder of the study. By comparing the persistence forecast with that of a simple random reference model (which has the same statistical properties of the observed time series, but lacks the observed autocorrelation), persistence was found to have some degree of skill for some parameters. However, when the time series were separated into solar minimum and maximum periods, persistence was found to have significant skill for all parameters at solar minimum, including B_Z, the out-of-ecliptic magnetic field component. Skills reduced for all parameters during solar maximum and there was no skill in B_Z at this time. As B_Z partly determines the coupling between the terrestrial and solar wind magnetic fields, this limits its usefulness as a space-weather forecasting tool during times of high solar activity. Further work is needed to see if the apparent skill in B_Z at solar minimum is real, or due to the random reference model overestimating variability in that portion of the cycle.

[39] Conversely, the solar wind speed, the second factor in terrestrial-solar wind coupling, is well forecast throughout the solar cycle: skill relative to the random reference model always remains positive. Furthermore, identifying key events within the time series, namely high speed streams (HSEs), shows that persistence correctly predicts nearly 80% of observed HSEs, but also that 23% of its predicted HSEs are false alarms. A significant fraction of both the missed and false alarm events are likely to be the result of transient HSEs resulting from high-speed CMEs. It may be possible to reduce the false alarm rate by differentiating between transient and corotating events, then removing the former from the persistence forecast. This will be investigated in future work. Note, however, that this will not reduce the number of missed HSEs.

[40] Looking to the near future, a persistence model may be extremely useful in the coming solar cycle. The most recent, extended, solar minimum has shown very recurrent stream structure. If the downward trend in sunspot number is part of a long-term decline in solar activity, as has been speculated [*Barnard et al.*, 2011; *Owens et al.*, 2011], the persistence forecast could be more accurate in the coming cycles than it has been during the space-age to date.

[41] Having characterized the general performance of the persistence model, the next step was to use the persistence model as a benchmark for assessing the performance of more sophisticated physics-based models throughout most of the last solar cycle (although results here are not definitive, as we did not use the latest versions of the physical models). Looking first at point-by-point comparisons, such as skill score based upon mean square error (MSE), the persistence model performs at least as well as the dynamical models during solar minimum, despite the 27 day lead time to forecasts. As solar activity picks up towards solar maximum, the corona becomes increasingly dynamic and persistence becomes a less

reliable approximation for near-Earth solar wind conditions. Consequently, dynamical models have higher skill for almost all solar wind parameters at solar maximum. However, there is evidence that this is at least partly the result of dynamical models generally underestimating the variability in the solar wind, which can not only reduce MSE but also reduce the usefulness of a forecast. In contrast with the point-by-point findings, an event-based assessment of solar wind structures suggests that persistence may be the most valuable forecast throughout the solar cycle. This reiterates the fact that the means of assessing the "best" forecast model must be specifically tailored to its intended use.

[42] **Acknowledgments.** We are grateful to the Space Physics Data Facility (SPDF) and National Space Science Data Center (NSSDC) for OMNI, and to Nick Arge, Dusan Odstrcil and Pete Riley for model data.

References

Altschuler, M. A., and G. Newkirk Jr. (1969), Magnetic fields and the structure of the solar corona, *Sol. Phys.*, *9*, 131–149.

Arge, C. N., and V. J. Pizzo (2000), Improvement in the prediction of solar wind conditions using near-real time solar magnetic field updates, *J. Geophys. Res.*, *105*, 10,465.

Arge, C. N., D. Odstrcil, V. J. Pizzo, and L. R. Mayer (2003), Improved method for specifying solar wind speed near the Sun, in *Proc. of the Tenth Internat. Solar Wind Confer.,* vol. 679, 190–193.

Baker, D. N., R. L. McPherron, T. E. Cayton, and R. W. Klebesadel (1990), Linear prediction filter analysis of relativistic electron properties at 6.6 R_E, *J. Geophys. Res.*, *95*, 15,133–15,140, doi: 10.1029/JA095iA09p15133.

Barnard, L., M. Lockwood, M. A. Hapgood, M. J. Owens, C. J. Davis, and F. Steinhilber (2011), Predicting space climate change, *Geophys. Res. Lett.*, *381*, L16103, doi:10.1029/2011GL048489.

Bartels, J. (1932), Terrestrial-magnetic activity and its relations to solar phenomena, *Terr. Magn. Atmos. Electr. (J. Geophys. Res.)*, *37*, 1, doi: 10.1029/TE037i001p00001.

Bartels, J. (1934), Twenty-seven day recurrences in terrestrial-magnetic and solar activity, 1923–1933, *Terr. Magn. Atmos. Electr. (J. Geophys. Res.)*, *39*, 201, doi:10.1029/TE039i003p00201.

Cane, H. V., and I. G. Richardson (2003), Interplanetary coronal mass ejections in the near-Earth solar wind during 1996–2002, *J. Geophys. Res.*, *108*, 1156, doi:10.1029/2002JA009817.

Chree, C., and J. M. Stagg (1928), Recurrence phenomena in terrestrial magnetism, *Phil. Trans. R. Soc. A*, *227*, 21–62, doi: 10.1098/rsta.1928.0002.

Detman, T., Z. Smith, M. Dryer, C. D. Fry, C. N. Arge, and V. Pizzo (2006), A hybrid heliospheric modeling system: Background solar wind, *J. Geophys. Res.*, *111*, A07102, doi:10.1029/2005JA011430.

Diego, P., M. Storini, and M. Laurenza (2010), Persistence in recurrent geomagnetic activity and its connection with space climate, *J. Geophys. Res.*, *115*, A06103, doi:10.1029/2009JA014716.

Dungey, J. W. (1961), Interplanetary magnetic field and the auroral zones, *Phys. Rev. Lett.*, *6*, 47.

Feynman, J., and S. B. Gabriel (2000), On space weather consequences and predictions, *J. Geophys. Res.*, *105*(10), 543.

Gopalswamy, N., A. Lara, S. Yashiro, M. L. Kaiser, and R. A. Howard (2001), Predicting the 1-AU arrival times of coronal mass ejections, *J. Geophys. Res.*, *106*, 29, 207.

Gosling, J. T., and V. J. Pizzo (1999), Formation and evolution of corotating interaction regions and their three dimensional structure, *Space Sci. Rev.*, *89*, 21–52, doi:10.1023/A:1005291711900.

Hapgood, M. A. (2011), Towards a scientific understanding of the risk from extreme space weather, *Adv. Space Res.*, *47*, 2059–2072.

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

Lei, J., J. P. Thayer, J. M. Forbes, Q. Wu, C. She, W. Wan, and W. Wang (2008), Ionosphere response to solar wind high-speed streams, *Geophys. Res. Lett.*, *35*, L19105, doi:10.1029/2008GL035208.

Linker, J., et al. (1999), Magnetohydrodynamic modeling of the solar corona during whole sun month, *J. Geophys. Res.*, *104*, 9809–9830.

Manchester, W. B., IV, T. I. Gombosi, I. Roussev, A. Ridley, D. L. De Zeeuw, I. V. Sokolov, K. G. Powell, and G. Toth (2004), Modeling a space weather event from the Sun to the Earth: CME generation and interplanetary propagation, *J. Geophys. Res.*, *109*, A02,107, doi: 10.1029/2003JA010150.

McGregor, S., W. J. Hughes, C. N. Arge, and M. J. Owens (2008), Analysis of the magnetic field discontinuity at the PFSS and Schatten current sheet interface in the WSA model, *J. Geophys. Res.*, *113*, A08112, doi:10.1029/2007JA012330.

McGregor, S. L., W. J. Hughes, C. N. Arge, M. J. Owens, and D. Odstrcil (2011), The distribution of solar wind speeds during solar minimum: Calibration for numerical solar wind modeling constraints on the source of the slow solar wind, *J. Geophys. Res.*, *116*, A03101, doi:10.1029/2010JA015881.

McPherron, R. L., and J. Weygand (2006), The solar wind and geomagnetic activity as a function of time relative to corotating interaction regions, in *Recurrent Magnetic Storms: Corotating Solar Wind, Washington DC American Geophysical Union Geophysical Monograph Series*, edited by R. McPherron, W. Gonzalez, G. Lu, H. A. José, and S. Natchimuthukonar Gopalswamy, vol. 167, p. 125.

Mikic, Z., J. A. Linker, D. D. Schnack, R. Lionello, and A. Tarditi (1999), Magnetohydrodynamic modeling of the global solar corona, *Phys. Plasma*, *6*, 2217.

Murphy, A. H. (1993), What is a good forecast? An essay on the nature of goodness in weather forecasting, *Wea. Forecasting*, *8*, 281–293.

Mursula, K., and B. Zieger (1996), The 13.5-day periodicity in the Sun, solar wind, and geomagnetic activity: The last three solar cycles, *J. Geophys. Res.*, *101*, 27,077–27,090, doi:10.1029/96JA02470.

Odstrcil, D. (2003), Modeling 3-D solar wind structures, *Adv. Space Res.*, *32*, 497–506.

Odstrcil, D., V. Pizzo, J. A. Linker, P. Riley, R. Lionello, and Z. Mikic (2004), Initial coupling of coronal and heliospheric numerical magnetohydrodynamic codes, *J. Atmos. Solar Terr. Phys*, *66*, 1311–1320.

Owens, M. J., C. N. Arge, H. E. Spence, and A. Pembroke (2005), An event-based approach to validating solar wind speed predictions: High speed enhancements in the Wang-Sheeley-Arge model, *J. Geophys. Res.*, *110*, A12105, doi:10.1029/2005JA011343.

Owens, M. J., H. E. Spence, S. McGregor, W. J. Hughes, J. M. Quinn, C. N. Arge, P. Riley, J. Linker, and D. Odstrcil (2008), Metrics for solar wind prediction models: Comparison of empirical, hybrid and physics-based schemes with 8-years of l1 observations, *Space Wea. J.*, *6*, S08001, doi:10.1029/2007SW000380.

Owens, M. J., M. Lockwood, C. J. Davis, and L. Barnard (2011), Solar cycle 24: Implications for energetic particles and long-term space climate change, *Geophys. Res. Lett.*, *38*, L19106, doi: 10.1029/2011GL049328.

Parker, E. N. (1958), Dynamics of the interplanetary gas and magnetic fields, *Astrophys. J.*, *128*, 664–676.

Pizzo, V. J. (1991), The evolution of corotating stream fronts near the ecliptic plane in the inner solar system. II - Three-dimensional tilted-dipole fronts, *J. Geophys. Res.*, *96*, 5405–5420, doi: 10.1029/91JA00155.

Richardson, I. G., H. V. Cane, and E. W. Cliver (2002), Sources of geomagnetic activity during nearly three solar cycles (1972–2000), *J. Geophys. Res.*, *107*, doi:10.1029/2001JA000504.

Riley, P., J. A. Linker, and Z. Mikic (2001), An empirically-driven global MHD model of the solar corona and inner heliosphere, *J. Geophys. Res.*, *106*, 15,889–15,902.

Roebber, P. J. (1998), The regime dependence of degree day forecast technique, skill, and value, *Weather and Forecast.*, *13*, 783–794, doi: 10.1175/1520-0434(1998)013<0783:TRDODD>2.0.CO;2.

Sargent, H. H., III (1985), Recurrent geomagnetic activity—Evidence for long-lived stability in solar wind structure, *J. Geophys. Res.*, *90*, 1425–1428, doi:10.1029/JA090iA02p01425.

Schatten, K. H., J. M. Wilcox, and N. F. Ness (1969), A model of interplanetary and coronal magnetic fields, *Sol. Phys.*, *9*, 442–455.

Siscoe, G. (2007), Space weather forecasting historically viewed through the lens of meteorology, in *Space Weather—Physics and*

Effects, edited by V. Bothmer, and I. A. Daglis, p. 5. doi:10.1007/978-3-540-34578-7-2.

Siscoe, G., D. Baker, R. Weigel, J. Hughes, and H. Spence (2004), Roles of empirical modeling within CISM, *J. Atmos. Solar Terr. Phys.*, *66*, 1469–1480.

Spence, H., D. Baker, A. Burns, T. Guild, C.-L. Huang, G. Siscoe, and R. Weigel (2004), Center for integrated space weather modeling metrics plan and initial model validation results, *J. Atmos. Solar Terr. Phys.*, *66*, 1491–1498.

Tsurutani, B. T., W. D. Gonzalez, F. Tang, S. I. Akasofu, and E. J. Smith (1988), Origin of the interplanetary southward magnetic fields responsible for the major magnetic storms near solar maximum (1978–1979), *J. Geophys. Res.*, *93*, 8519.

Tsurutani, B. T., R. L. McPherron, W. D. Gonzalez, G. Lu, J. H. A. Sobral, and N. Gopalswamy (2006), Introduction to special section on corotating solar wind streams and recurrent geomagnetic activity, *J. Geophys. Res.*, *111*, A07S00, doi:10.1029/2006JA011745.