SPACE WEATHER, VOL. 7, S06004, doi:10.1029/2009SW000463, 2009

Validation of community models: Identifying events in space weather model timelines

Peter MacNeice[1]

Received 13 January 2009; revised 22 April 2009; accepted 24 April 2009; published 25 June 2009.

[1] I develop and document a set of procedures which test the quality of predictions of solar wind speed and polarity of the interplanetary magnetic field (IMF) made by coupled models of the ambient solar corona and heliosphere. The Wang-Sheeley-Arge (WSA) model is used to illustrate the application of these validation procedures. I present an algorithm which detects transitions of the solar wind from slow to high speed. I also present an algorithm which processes the measured polarity of the outward directed component of the IMF. This removes high-frequency variations to expose the longer-scale changes that reflect IMF sector changes. I apply these algorithms to WSA model predictions made using a small set of photospheric synoptic magnetograms obtained by the Global Oscillation Network Group as input to the model. The results of this preliminary validation of the WSA model (version 1.6) are summarized.

Citation: MacNeice, P. (2009), Validation of community models: Identifying events in space weather model timelines, *Space Weather*, *7*, S06004, doi:10.1029/2009SW000463.

1. Introduction

[2] Forecast models are frequently required to predict the time evolution of specific physical parameters. Typically, the model operators are searching for a specific type of signal in the predicted timeline. For example, forecasters need to know when the solar wind will experience an abrupt transition from slow wind to fast wind, since these events are frequently associated with enhanced geomagnetic activity [*McPherron et al.*, 2004]. However, to make use of these predictions, the forecasters must know what confidence they can place in the model forecast. This requires an analysis of the models performance as measured against archived data. Under ideal circumstances this analysis should be based on the largest possible archived data set, which means the process must be automated. This presents a challenge to develop an algorithm which is capable of identifying the appropriate pattern in the data and corresponding signal in the model output.

[3] In this paper I consider the problem of identifying high-speed enhancements (HSEs, i.e., abrupt transitions from slow to fast solar wind), and crossings of the Heliospheric Current Sheet (HCS) in the Operating Mission as Nodes on the Internet (OMNI) data set, for comparison with predictions made with the Wang-Sheeley-Arge (WSA) model [*Arge and Pizzo*, 2000]. I will outline some of the issues faced in developing this specific feature detection algorithm, and discuss its performance. The algorithm I develop is being used to analyze the performance of the WSA model as it is run in quasi-operational mode at the Community Coordinated Modeling Center (CCMC). It is also being used to analyze the performance of the next generation of MHD based forecast models, also in use at the CCMC. Because it forms the foundation of so much model validation, it is essential that the algorithm be presented, and its design justified, in detail, and in the peer reviewed literature. Our intent is to achieve proper disclosure of our algorithms which I will use extensively in future validation of this class of forecast model. Our intent is not to advocate their adoption as a community standard or to apply them in a general way beyond this very specific application.

2. Data Preparation

[4] To test and refine our algorithms, I applied them to observations of the solar wind speed and the polarity of the radial component of the interplanetary magnetic field (IMF) at the L1 point. I considered the time interval covering Carrington rotations (CR) 1650 through to 2074 (i.e., from January 1977 to August 2008). I chose this interval because the WSA model requires synoptic line of sight magnetogram data as input, and the available archive of usable full rotation synoptic maps begins with CR1650.

[5] The L1 measurements of solar wind speed and IMF were obtained in the form of hourly averages from the

[1]NASA Goddard Space Flight Center, Greenbelt, Maryland, USA.

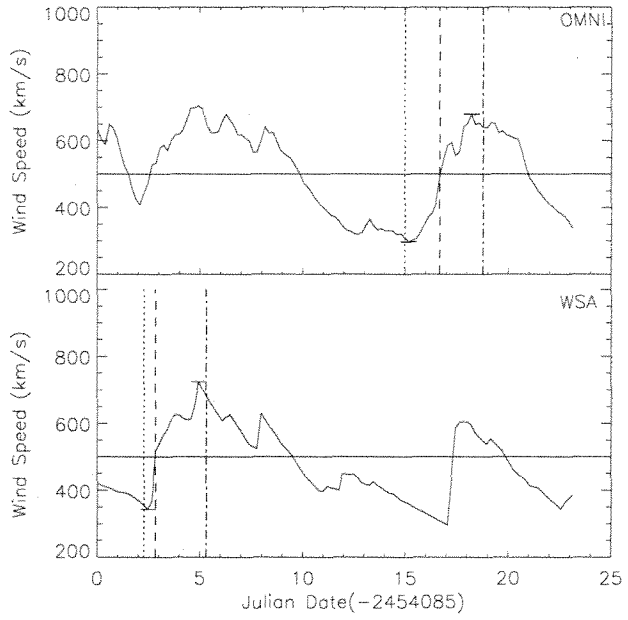

Figure 1. HSE detection for CR2051 in the WSA model output and the OMNI solar wind speed data, using the detection algorithm of *Owens et al.* [2005]. (top) The OMNI wind speeds and (bottom) the WSA forecast, using a GONG synoptic magnetogram. The three vertical lines, dotted, dashed, and dotted-dashed, associated with each HSE indicate the start time, time of peak speed gradient, and the end time of the rapid rise, as identified by the detection algorithm, respectively. The short horizontal bars mark each HSEs minimum and maximum speed.

OMNI 2 database [*King and Papitashvili*, 2005]. The OMNI database provides a nearly continuous timeline of the near-Earth solar wind state by combining measurements from multiple spacecraft. For the time frame of interest to us, the observations were taken from the Interplanetary Monitoring Platform 8 (IMP8), WIND and Advanced Composition Explorer (ACE) spacecraft. For consistency I rebin the OMNI wind speed measurements to match the temporal sampling of the WSA results. The OMNI data is provided in the form of 1 h averages. Our rebining does a time integration of these OMNI 1 h averages over each WSA time bin, using linear interpolation to reconstruct the profile between the OMNI data points.

[6] I also rebin the IMF polarity observations to match the WSA temporal sampling, but before I can do this I filter it with a low-pass filter. This filtering, and its rationale are discussed in detail in section 4.

[7] The WSA (version 1.6) model runs which I have used in this study, begin by interpolating the magnetograms onto a spherical grid with 2.5° resolution from latitude −90° to +90° and longitude 0 to 360°. The Sun's rotation sweeps through 2.5° in 4.5 h (i.e., 27.27 × 2.5/360 days), and so the WSA timelines for wind speed and IMF are reported with this time resolution.

[8] To illustrate the use of our algorithms I present a limited validation of the WSA model using the Global Oscillation Network Group (GONG) archive of synoptic magnetograms. The GONG archive begins with CR2047. I analyze model runs for rotations 2047 through 2074. A comprehensive validation using synoptic magnetograms from GONG, the National Solar Observatory and Mount Wilson Observatory, and considering all the factors which influence the WSA model's performance will be presented by P. MacNeice (Validation of community models: Part 2. Development of a baseline using the Wang-Sheeley-Arge model, manuscript in preparation, 2009).

[9] Although our goal is to characterize the quality of forecasts made by the models, our validation does not test the models exactly as they are used in day to day operations. The difference here is that I use archived synoptic magnetograms that have been constructed from uniformly time averaged full disk magnetograms. The daily synoptic magnetograms that are available in a true realtime forecasting environment cannot be uniformly time averaged in this way. Therefore our measures of the model's forecast quality should be considered as an upper bound on the forecast quality that is possible in a realtime environment.

3. HSE Detection

3.1. Basic HSE Algorithm

[10] To define our basic algorithm I follow the simple approach defined by *Owens et al.* [2005]. Their intent is to define a HSE to be a period in which a speed gradient threshold (they choose 50 km/day) is exceeded for a minimum duration (they choose 2 days). They add some criteria for concatenating HSEs which are sufficiently close in time, and for resolving ambiguity in the association of events in the model and data timelines. I began by testing it when applied to the OMNI data from Carrington rotations 1650 through 2074.

[11] This is a simple definition and is very easy to apply. Its simplicity is a strength, but it is also its weakness. I found that it frequently fails to identify good HSE candidates, and can produce a positive identification for questionable features. As is evident in some of the examples below, there are occasions in which the model timeline appears visually to be a good match to the data, and yet, because it narrowly fails to exceed a threshold value it is not recorded as a prediction hit. Figure 1 shows a good example of these failings.

[12] Figure 1 shows the WSA and OMNI solar wind speeds for Carrington rotation 2051. In this case I have used the *Owens et al.* [2005] algorithm to identify HSEs. Figure 1 (top) is the OMNI observations and Figure 1 (bottom) is the WSA model output based on a synoptic magnetogram from the GONG network. The WSA forecast shows two extended periods of high-speed wind, the first starting at day 3 and the second at 17. The vertical

lines indicate HSEs which were identified by the algorithm. In this case the *Owens et al.* [2005] algorithm spots the first event in the WSA forecast but misses the second because its duration was slightly less than 2 days. When applied to the OMNI data (Figure 1 (top)), the *Owens et al.* [2005] algorithm misses the first event, again because the duration criteria marginally fails. It does identify the second event. The result is that for this rotation, the model would have been judged to have made one false prediction and missed one event. In reality, as I can see from a visual inspection of the timelines, the model and data are in surprisingly good agreement for almost the entire period.

[13] In fairness, I should point out that the CR2051 example illustrates the *Owens et al.* [2005] algorithm performing at it's weakest. There are some rotations for which it works better. However this case shows clearly a need to improve upon their basic definition.

[14] Since HSEs are expected to have additional signatures (such as density enhancements and magnetic field compression coincident with the start of the rise in wind speed [*Tsurutani et al.*, 2006]), it could be argued that these properties should also be incorporated in the HSE detection algorithm. Under ideal conditions this would be desirable. However, as the detection algorithm is extended to include more observable wind parameters, the fraction of the the overall time interval in which some part of the required data is missing because of data gaps, grows rapidly. Since the relatively simple speed gradient approach seems to work well, I have chosen not to implement these more complex strategies.

3.2. Revising the Basic HSE Algorithm

[15] It is clear from the previous section that the simple definition of *Owens et al.* [2005] does not reliably match events in the data record with similar events in the model output. There are two principal failings. The first, and most important, is that it excludes too many candidate events because the durations of steep speed gradient are too short. The second is that slight differences in onset time can lead to failures in achieving an association between an event in the model and its corresponding event in the data. Some adjustment is required in the way events are recognized, and then associated with the appropriate event in the other timeline.

[16] I have developed a modification of the *Owens et al.* [2005] detection algorithm and supplemented it with some additional checks. These modifications are not arbitrary. In designing them I have been guided by the extensive study, by *Zhang et al.* [2006], of solar wind conditions that are observed to be associated with geomagnetic storms.

[17] The first and most important modification I introduced was to define a "candidate HSE" as any region of solar wind, 1 day or longer, in which the net solar wind speed increase is 50 km/s or more. This adjustment achieves the goal of accepting HSEs with the same speed gradient as the *Owens et al.* [2005] algorithm, while permitting a sharper rise time. The *Owens et al.* [2005] algorithm can discard these if the pre-HSE wind speed was elevated, or if the period immediately following the HSE rise shows a temporary decline in wind speed.

[18] Missing too many events with rise times which are too short is a significant shortcoming because observations indicate that HSEs that are associated with magnetic storms, typically have short rise times. *Zhang et al.* [2006] studied 549 storms for multiple periods during times of solar minimum and maximum. Regardless of the phase of solar cycle, the average wind speed rise was of order 100 km/s and occurred in a period of less than one day. Their results also underscore the important role that coincident spikes in southward IMF have on the likelihood of occurrence of a magnetic storm.

[19] Because this adjustment allows for shorter and sharper disturbances I supplement it with some additional criteria to make sure that the event is not an isolated spike in wind speed. First, I insist that the HSE must genuinely transition from slow wind to fast wind. In the large study of *Zhang et al.* [2006], except in the case of intense storms occurring at solar minimum, the averaged speed profiles in the solar wind at the time of the storm showed a rise from about 450 km/s to peak speeds of between 500 and 550 km/s. Hence I choose 500 km/s as the speed separating these two regimes. Therefore, a HSE candidate must begin from a speed below 500 km/s and must have a maximum speed in excess of 500 km/s.

[20] As done by *Owens et al.* [2005] I identify the nominal HSE time as the time of peak gradient during the rise phase of the HSE. I search back in time from this nominal time for the start of the HSE rise. This will be a local minimum in the wind speed. However simply searching for the first preceding local minimum can be misleading. Many HSEs have a rise phase which is largely monotonic, but includes isolated single point minima somewhere in the middle of the rise phase. These isolated minima, if taken as the HSE start time, could lead the algorithm to consider a single HSE as two distinct HSEs, or even worse, might cause us to reject the HSE completely because the different parts individually fail an amplitude test which their combination would easily pass. Therefore, in searching for the preceding minimum, I test the 2 day interval preceding the time of maximum speed gradient. If I find a speed in that interval which is more than 50 km/s less than our initial speed minimum, I choose the time of minimum speed during that 2 day window as the "true" HSE start time.

[21] I look forward in time from the nominal HSE time to determine the time of the peak speed. I compute the duration of the HSE speed rise and reject any events whose duration is so short that it is not adequately resolved. The model's effective temporal resolution is 4.3 h. By choosing a minimum HSE duration of 12 h, I minimize the possibility that noisy data will trigger a spurious HSE detection. This minimum allowed duration is significantly shorter that the 2 day minimum duration

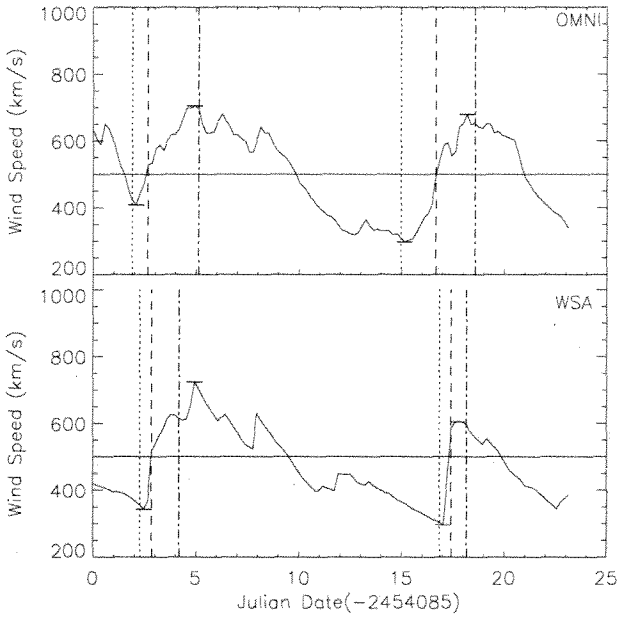

Figure 2. HSE detection for CR2051 in the WSA model output and the OMNI solar wind speed data, using our new detection algorithm.

used in the *Owens et al.* [2005] algorithm. Finally, I require that the amplitude of the HSE be at least 200 km/s.

[22] To be more specific, our algorithm is implemented as follows: (1) Mark all time points, t, for which $v(t) - v(t - 1\ \text{day}) > 50$ km/s. (2) Eliminate any isolated single time points which were marked. (3) Number each contiguous block of marked points as a distinct HSE. (4) Combine HSEs separated by less than 0.75 days. (5) Reject any HSEs with minimum speed greater than 500 km/s or maximum speed less than 500 km/s. (6) Reject any HSEs where the duration of acceptable speed gradient is less than 0.5 days. (7) Determine the HSE amplitude and reject HSEs with an amplitude less than 200 km/s. (I search for the HSE's minimum velocity by locating the minimum velocity in the 2 days preceding the nominal HSE time. If at that time the speed gradient is positive I step back in time until I encounter a zero gradient. The speed at this time (designated T_{start}) is then the minimum velocity at the start of the HSE. I search forward in time from the nominal HSE time in a similar way to find the peak speed of the HSE. Its time is designated T_{end}. The amplitude is the difference between the maximum and minimum speeds.)

[23] The event detections by the revised algorithm for the CR2051 data set is shown in Figure 2. Both events are now detected in both the model and observed timelines, providing a much more accurate assessment of the model's quality in this case.

[24] Once an event has been identified in either of the timelines, I search for a corresponding event in the other timeline. I define a "hit" as a HSE, predicted by WSA, which has a corresponding HSE in the OMNI timeline for which the intervals defined by the HSE start and end times, T_{start} and T_{end}, overlap. An event in the WSA timeline with no match in the OMNI data is called a "false positive." An event in the OMNI timeline with no corresponding event in the WSA timeline is called a "miss."

[25] I need to give special consideration to HSEs which occur near the start or end of each Carrington rotation, because our study is using archived full rotation synoptic magnetograms rather than the daily updated magnetograms that would be available to a forecaster in a realistic forecasting environment. The issue is discussed in detail in Appendix A. To avoid this complication I do not include unmatched HSEs which occur within 2 days of the start (or end) of each rotation, since these may have a matching HSE during the previous (or next) rotation.

3.3. Demonstrating the Algorithm's Application

[26] A comprehensive validation of the WSA model will be published by P. MacNeice (manuscript in preparation, 2009) of this series, which will use almost all the available archived synoptic magnetograms, and will consider all the factors which affect the model performance. Here I illustrate the use of our revised algorithm by summarizing the results of a validation, limited to just the archive of GONG synoptic magnetograms, which includes CR2047 through 2074.

[27] For this set of Carrington rotations, the *Owens et al.* [2005] algorithm reported 11 hits, 21 misses and 5 false positives. By comparison, our revised algorithm recorded 22 hits, 32 misses and 14 false positives. Of these, 12 false positives are unambiguous. I reject two of the false positives which missed possible matches in the OMNI data because the predicted event occurred too close to the start or end of the Carrington rotation.

[28] Stated in a different way, of the 54 events which our algorithm detected in the OMNI data, 41% were predicted by WSA. Of the 36 events predicted 33% were unambiguous false positives. For the *Owens et al.* [2005] algorithm, of the 32 events detected in the OMNI data, only 33% were predicted, and of the 16 events predicted, 33% were false positives. Our algorithm suggests that WSA is slightly better as a HSE forecaster than is reported by the *Owens et al.* [2005] algorithm.

[29] Figure 3 shows the distribution of timing error for each category of HSE. For misses and false positives I compute the time to the closest possible match. Of the 32 total misses, only 6 have potential forecast HSEs within a 5 day time error. For false positives only 6 out of 14 are within a 5 day timing error of an observed HSE. This indicates that when the forecast is in error, it is not simply due to a timing error, but frequently due to an absence of a credible matching candidate HSE. This can be because of a complete absence of any suitable matching HSE candidate or because a possible match fails the amplitude or rise time tests.

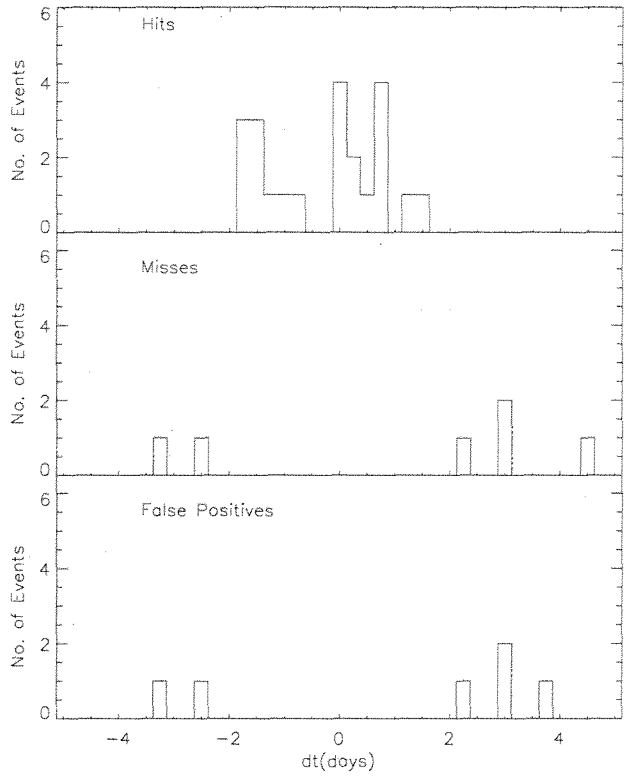

Figure 3. The distribution of HSE timing errors for hits, misses, and false positives.

[30] So which model rating is more credible? This is inevitably a matter of subjective judgement. For a given candidate event there is no "right" answer as to whether it should or should not be labeled an event. Comparison between algorithms based on observer intuition is simply too subjective a criterion. The most important point is to apply the same algorithm to each model.

4. Radial IMF Polarity Transitions

[31] The WSA model makes clear predictions of changes in magnetic sector, due to crossing of the Heliospheric Current Sheet (HCS), that can be associated with reversal in the sign of B_r. Comparison of these predictions with observations can be used to measure the quality of the model's reproduction of the HCS, and overall magnetic configuration of the model.

[32] However, the OMNI data for the polarity of B_r consistently indicates a very large number (typically more than one hundred per Carrington Rotation) of short-term fluctuations of B_r superimposed upon the longer-time trends associated with the large-scale IMF (Figure 4 (top)). These fluctuations have amplitudes much greater than the measurement accuracy of the spacecraft magnetometers. The causes of these fluctuations are not yet clear [see, e.g., *Roberts et al.*, 2005; *Smith*, 2001]. They may be due to waves in the HCS or due to spatial structure inherent in the current sheet.

[33] Given the temporal and spatial resolution of the models, it is not appropriate to expect the model to match each individual reversal in the polarity of B_r. Our goal is to approximately identify the times associated with HCS crossings which result in prolonged periods of reversed polarity, and which reflect the large-scale structure of the current sheet and sector locations. I will refer to these prolonged periods of fixed polarity as "polarity phases."

[34] Developing an automated procedure to search for these crossings is not trivial. They rarely, if ever, appear as a single polarity reversal. In searching for HCS crossings *Roberts et al.* [2005] have categorized the observed reversal in the sign of B_r/B as either "simple" or "complex." A crossing is deemed simple if the transition happens in less than 5 h and involves three or fewer crossings. The transition is complex if it takes longer and involves more than three crossings. They claim that almost all their observed cases fit one or other of these categories.

[35] *Crooker et al.* [1996] suggested using the plasma entropy as an indicator of change in the original solar source region of the plasma. Sharp entropy minima which are temporally coincident with a reversal could in principle confirm passage of the spacecraft through the streamer belt. *Crooker et al.* [1996] also proposed a heat flux polarity test to distinguish between crossings of the true sector boundaries, and crossings of local current sheets which are created by folds in the HCS.

[36] The heat flux analysis can help clarify the nature of individual field reversals, but is less useful for our more limited goal of identifying the large-scale HCS shape. There are two reasons for this. The first is that it is undependable over long periods, since, as pointed out by *Kahler et al.* [1998], the heat flux data is sometimes either missing or ambiguous. The second is that the WSA model's temporal resolution is currently too coarse to benefit from the heat flux analysis. The chief benefit of the use of heat flux is to identify apparent polarity reversals due to local folds in the HCS. These "false" reversals are usually short lived, with the true polarity being recovered on timescales shorter than the temporal resolution of our model. Indeed, when comparing the polarities determined from field direction alone with that from the difference in direction of the field and heat flux in Figure 2 of *Kahler et al.* [1998], the reduction in scatter which they achieve appears to have little or no impact on determination of the large-scale sector boundary structure.

[37] I experimented with use of both the entropy and heat flux tests, but found that both were plagued with problems of data dropout when applied for extended periods. Therefore I conclude that our purpose in detecting the large-scale reversals can be most easily achieved with a simple temporal filtering of the field direction data.

[38] In Figure 4 I have plotted the timelines of normalized radial field, the wind entropy, and the field angle ϕ in the RT plane of RTN coordinates (with $\phi = 0$ indicating

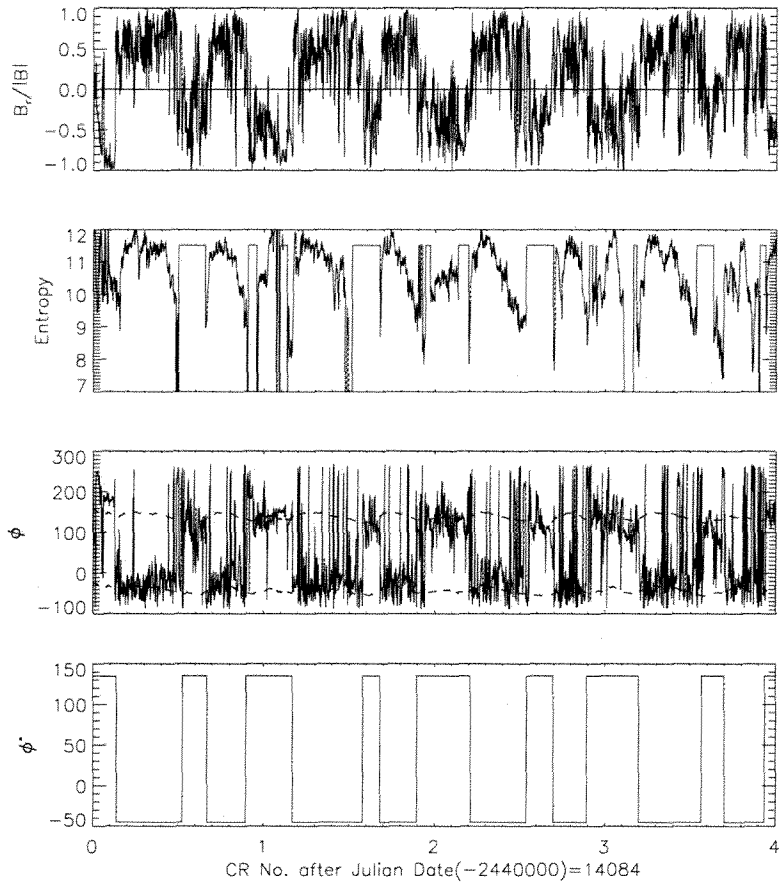

Figure 4. OMNI (ACE) measurements of the IMF for Carrington rotations 2051–2054. (top) Normalized B_r, (top middle) the solar wind entropy, (bottom middle) the angle ϕ between B and the radial direction, and (bottom) the angle ϕ^* determined by our algorithm. The dashed lines in Figure 4 (bottom middle) show the direction of the Parker spiral, with the lower line showing the outward direction and the upper line showing the inward direction.

the radial direction), for a set of four consecutive Carrington rotations. What I would like to extract from these plots is an estimate of the times of the true sector transitions for direct comparison with the WSA model predictions. In Figure 4 (bottom middle), the bottom dashed line is approximately 45° behind radial indicating the outward direction along the archimedean spiral, while the upper dashed line is rotated 180° from that direction indicating the inward spiral direction. The variation in direction of the spiral is caused by the varying wind speed.

[39] The algorithm I have developed has three basic steps. First I designate each data point as inward or outward depending on which direction is closest. Then I apply a broad running average to filter out almost all high-frequency signal. Finally I tidy up the result by surgically eliminating any obvious remaining isolated reversals.

[40] Specifically, to determine the HCS crossings representative of the overall sector structure, I apply the following processing steps to the angle ϕ: (1) If ϕ is greater than 45° set it to 135°(inward). Otherwise set to −45°(outward). (2) Construct a running 60-h average, $<\phi>$ of ϕ. Again if $<\phi>$ is greater than 45° set it to 135°. Otherwise set to −45°. (3) Remove any single point reversals which have at least a 3 h period of opposite polarity on each side. (4) Eliminate any periods of reversal which are no longer than τ = 0.1 days and are separated from the next reversal by at least 3τ days on either side. (5) Repeat the previous step in turn for τ = 0.2, 0.3, 0.4, 0.5 days.

[41] The result, ϕ^*, of this filtering of the original ϕ data is shown in Figure 4 (bottom). It clearly reproduces all the obvious large-scale transitions. In cases where the polarity is highly variable as a large-scale sector crossing is in progress, it gives a reasonable average location.

[42] In Figure 5 I illustrate the result of our polarity reversal analysis for Carrington rotation 2051. The solid

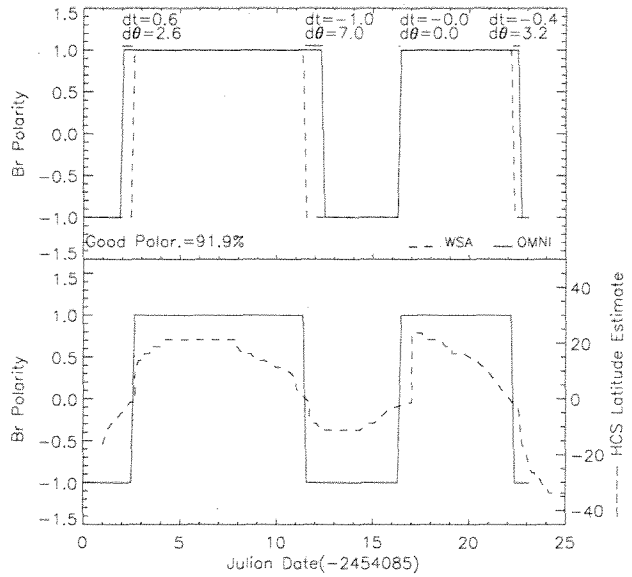

Figure 5. The WSA prediction for the polarity of B_r (solid line in Figure 5 (bottom) and dashed line in Figure 5 (top)) compared with the OMNI (ACE) equivalent (solid line in Figure 5 (top)), as determined using our analysis algorithm. The dashed line in Figure 5 (bottom) indicates approximate distance in latitude from the ecliptic to the HCS at 1 AU as predicted by WSA. In Figure 5 (top) *dt* signifies the time offset, in days, between each observed reversal and the associated predicted reversal and $d\theta$ denotes the average absolute value of latitude difference, in degrees, between the HCS and ecliptic during the periods of polarity mismatch.

line in Figure 5 (top) shows the OMNI(ACE) observation, after processing by our algorithm. The solid line in Figure 5 (bottom) shows the WSA forecast, which is also shown (as a dashed line) in Figure 5 (top) for ease of comparison with the observations.

[43] Each matched reversal is labeled with a timing error *dt*. The dashed line in Figure 5 (bottom) shows the latitude difference between the HCS and ecliptic at 1AU as predicted by the WSA model. Each matched reversal is also labeled in Figure 5 (top) with $d\theta$, the average value of this latitude offset during the period of polarity mismatch.

[44] I can measure the model's performance in three ways, by reporting the fraction of time in which the polarity matches the "measured" polarity, by reporting the timing errors between actual reversals and their closest available predicted reversal (labeled *dt* in Figure 5 (top)), and finally by reporting on the number of missed and false polarity phases. I define a polarity phase as an extended period of time with a constant value of $B_r/|B_r|$.

[45] Where the model appears to perform poorly, I can also consider the model's prediction for the latitudinal distance of the HCS from the L1 location in the plane of the ecliptic to see if the HCS may be close to the ecliptic. If this were true then the model would actually be much closer to an accurate prediction than would be apparent from our initial analysis. To examine this possibility I have plotted the latitude of the HCS at 1 AU along the Earth's longitude in the dashed line in Figure 5 (top). For each matched reversal in Figure 5 I report the average latitude difference, $d\theta$, between the HCS and the plane of the ecliptic at 1 AU, during the period of polarity mismatch.

[46] I deliberately chose to work with the WSA model predictions as they would be generated by the current version (V1.6) of the model when used in a realistic forecasting environment. Therefore, for reasons discussed in Appendix A, when compiling statistics on the likelihood of a successful polarity reversal prediction, I eliminate from consideration, OMNI reversals which are the first (last) during the rotation and which do not match the sign change of the first (last) reversal in the WSA timeline. Similarly, in assessing whether WSA misses a polarity phase completely, I do not count missed phases which end less than two days into the current time window, or which begin less than two days before the end of the time window, since it is possible that with updated magnetogram data from the next CR, the model might not have completely missed those phases.

[47] I illustrate application of the polarity reversal detection algorithm with a limited validation using the GONG synoptic magnetogram archive. For the Carrington rotations 2047 through 2074, the WSA model correctly matched the polarity of B_r, 82% of the time. During that time it missed 16 polarity phases out of a total of 119 phases in the OMNI signal. The WSA model with its current resolution, predicts fewer distinct phases than I determine through our processing of the OMNI timeline. There were 86 distinct WSA polarity phases of which only one could not be matched in our processed OMNI polarity signal.

[48] There were 54 reversal matches. The model failed to predict 37 reversals, of which 25 were unambiguously missed while the remaining 12 were close to the start or end of the associated Carrington rotation and might be matched by a multirotation analysis.

[49] The average timing error between the polarity reversals which were matched was 0.65 days. In Figure 6 I have plotted each matched reversal as a function of timing error and average latitudinal distance from the HCS during the time interval between the OMNI reversal and the matching WSA prediction, i.e., during the period of "error." To place the timing error and HCS latitude offset on equal footing, I express the timing error as an effective longitude error by multiplying the time error, measured in days, by a factor of 360/27.27 degrees/day. The dashed circle indicates a generalized angular error equivalent to a time error of 1 day. Approximately 46% of the matched reversals have a timing error of less than 1 day. About 30% of the remaining matches average less

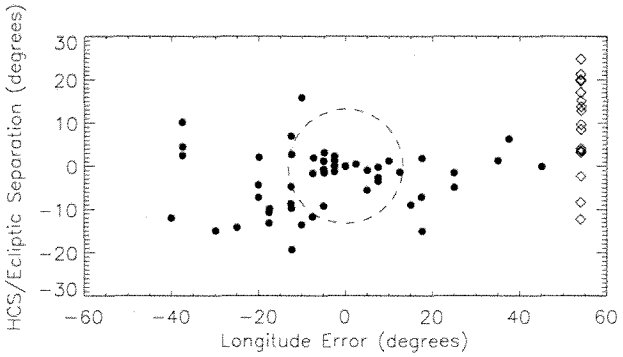

Figure 6. The time error and average distance between the ecliptic and the HCS during the "error" interval, for each reversal of B_r polarity in the OMNI data for which I find a matching WSA reversal prediction. The time difference is expressed as an effective longitudinal angle. The dashed circle denotes the locus of points with a combined angular error equivalent to a timing error of 1 day. The column of diamonds by the right axis shows the average angular separation of the ecliptic from the HCS during polarity phases which are completely missed by WSA.

than 5° in latitude from the HCS through the time of the polarity mismatch, indicating a close miss. This is comparable to the model resolution since the model grid spacing is 2.5°. For comparative purposes I superimpose in diamonds, the average HCS offsets for polarity phases in the OMNI data which are entirely missed by the WSA model. These points are arbitrarily plotted in a column to the right of the plot to avoid confusion. Again 4 out of the 16 show a separation between the ecliptic and HCS of less than 5°.

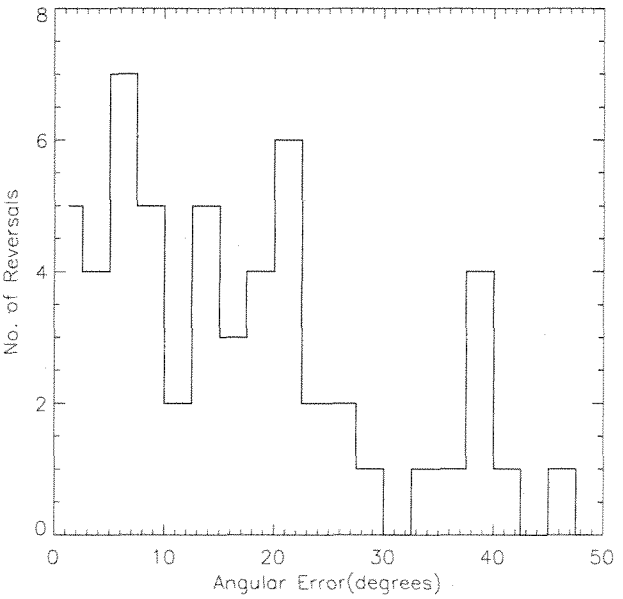

Figure 7. The frequency of absolute generalized angular error for the forecasted polarity reversals in our GONG based sample of WSA forecasts.

[50] Finally, in Figure 7 I plot in histogram form the distribution of the absolute generalized angular error. The absolute generalized angular error is the distance from the origin in Figure 6. The mean of this distribution is 16 degrees.

5. Forecast Probabilities

[51] Forecasters need these results expressed in terms of the probabilities that the model can accurately predict these events. For HSEs the issue is simplest. When the solar wind is slow (<500 km/s), if the WSA model predicts or does not predict an HSE in the next 24 h, what is the probability that an HSE will or will not occur. These forecast probabilities are presented in Table 1.

[52] Approximately 30% of the time when WSA predicts a HSE in the next 24 h, its forecast is accurate. Not surprisingly, the null forecast is much more accurate. When WSA predicts no HSE will occur, it is correct approximately 90% of the time.

[53] For the polarity of B_r the forecaster's question is conditioned by their knowledge of the current polarity. The forecaster wants to predict periods of a given polarity and the probabilities of polarity reversals. Forecasters operate with a knowledge of the present polarity and the model's nowcast. So a typical question they might ask is for example, if both model and data currently agree, and the model predicts a polarity reversal in the next 24 h, what is the probability that there will be a reversal in that time period? To compute these probabilities, I consider each time point in our data set and record the answer to the question. I have done this for the eight questions I can formulate and the results are presented in Table 2. For example, when both model nowcast and OMNI are in agreement and WSA predicts a reversal some time in the next 24 h, the model is correct 36% of the time, and incorrect 64% of the time. Care should be taken in assessing the probabilities when the model nowcast disagrees with the current OMNI measurement. For example, if the nowcast is wrong and WSA predicts a reversal, I find no reversal occurring 92% of the time, which means that after 24 h the model polarity will actually be correct. This can occur when WSA errs by making a late forecast of a

Table 1. Forecast Probabilities for Occurrence of HSEs Within 24 h of the Current Time

Model	Observation	Percent
WSA predicts HSE	OMNI HSE	29
WSA predicts HSE	OMNI no HSE	71
WSA predicts no HSE	OMNI HSE	11
WSA predicts no HSE	OMNI no HSE	89

Table 2. Forecast Probabilities for Predictions of Polarity of B_r[a]

Model	Observation	Current Polarities	
		Agree	Disagree
WSA predicts reversal	OMNI reversal	36 (good)	8 (bad)
WSA predicts reversal	OMNI no reversal	64 (bad)	92 (good)
WSA predicts no reversal	OMNI reversal	7 (bad)	63 (good)
WSA predicts no reversal	OMNI No reversal	93 (good)	37 (bad)

[a]For all combinations, when the model nowcast is correct or incorrect, and the model predicts or does not predict a polarity reversal. Values given are percentages. For each case the comment in parentheses with the number indicates whether the model prediction is good or bad.

reversal which has already occurred, or when the model has previously predicted a reversal which did not happen. In either case the model is actually doing better by predicting the "nonexistent" reversal, because in doing so it is catching up with the current solar wind state.

[54] The probabilities in Tables 1 and 2 are based on data for a small set of Carrington rotations, and are presented here as a sample of the analysis that I will apply to a much large data set from a number of different observatories. The probabilities are derived by considering each time point as a distinct test. It should be remembered that consecutive "tests" are not independent. Therefore I do not estimate error bars for these probabilities since our tests will not be governed by gaussian statistics.

6. Conclusions

[55] I have presented in detail the event detection algorithms to be used by the CCMC in evaluating the WSA model and the heliospheric MHD models in use by the CCMC. These algorithms, which I have refined through testing on OMNI data covering the period from January 1977 (CR1650) to August 2008 (CR2074), will supplement skill score evaluations of the same models.

[56] Our algorithms represent one possible, and effective, approach to validate this class of forecast model. I do not claim this to be the only effective approach and are not promoting it for adoption as a community standard. It is presented so I can openly disclose the method to be used in our subsequent papers, in which I will report comprehensive validations of these models. As the models and data supply evolve in the future I expect these algorithms will need to be adapted to accommodate that evolution.

[57] The limited validation of the WSA model that I present here is intended to illustrate the application of our methodology, and not to serve as a comprehensive validation. While the results are accurate, they are limited because they apply only to Carrington rotations 2047 through 2074, and only when the model uses magnetograms from the GONG network. P. MacNeice (manuscript in preparation, 2009) will present a comprehensive validation of WSA, including both event detection and skill scores. It will examine the influence of all the factors upon which the model results depend, including the influence of magnetogram source, outer radial boundary of the current sheet component, difference between active and quiet solar conditions, and the effects of different temporal smoothing of polar fields.

[58] The WSA model constructs solutions for ambient solar wind. One might reasonably ask, what effect would coronal mass ejections have on our validation of the model's HSE and polarity reversal forecasts. Some HSEs, particularly during solar maxima, will be associated with arrival of ICMEs, but the model will not be able to predict them. It is also likely that in the wake of an Earth directed ICME, the IMF and solar wind are perturbed for some time, which would also be expected to reduce the quality of the model forecast. Therefore I would expect that the quality of the model forecasts would decrease around solar maxima because of the effects of increased ICME production. Because I have demonstrated the application of our validation algorithm using the GONG magnetogram archive, which is limited to the most recent solar minimum, our validation demonstration cannot test this influence. This issue will be addressed in detail by P. MacNeice (manuscript in preparation, 2009).

Appendix A

[59] When compiling the validation results for HSEs and polarity reversals, I do not include unmatched events which are within 2 days of the start or end of the Carrington rotation. For polarity reversals I make the additional exclusion of forecast reversals which are the first (or last) during a rotation and whose sign change (i.e., positive to negative or negative to positive) does not match the first (or last) reversal sign change in the observed polarity timeline.

[60] The reason for this is that I want our results to reflect the accuracy of the model when run in a realistic forecasting environment, but I are using archived full rotation magnetograms, not the daily updated magnetograms that a forecaster would be using. Synoptic magnetograms are constructed by combining into a weighted time average, all the available full disk magnetograms taken during the time interval of the given Carrington rotation. A detailed explanation of the typical method for constructing synoptic magnetograms is given by *Harvey and Worden* [1998].

[61] Each synoptic magnetogram provides an approximate global representation of the radial photospheric magnetic field as if it were unchanging during the 27.27 days preceding the last full disk magnetogram used in its construction. An archived full disk synoptic map for Carrington rotation C would cover Carrington times (C, 0°) to (C, 360°). When WSA forecasts are made in a realistic forecasting environment, the forecaster uses the most recent daily updated synoptic map. These maps span

Carrington time ($C - 1$, ϕ) to (C, ϕ), where ϕ is the latest Carrington longitude contributing data to the synoptic map (typically about 60° east of the sub-Earth longitude at the time of the last observation).

[62] The surface fields most relevant for predicting the current wind state at 1AU were recorded during the previous five days and so are the most up to date section of the daily updated synoptic magnetogram. Daily forecasts of events for the next 24 h can be produced using a sequence of daily maps whose differences reflect just one days evolution of the solar field.

[63] When I try to mimic this process with archived full disk synoptic magnetograms I use a single fixed magnetogram for all times during the Carrington rotation. This is an acceptable analog to the realistic forecasting approach, for most times during the rotation because of the way in which the spatial and temporal weighting is done. It is not a good analog for times close to the start and end of the rotation.

[64] Consider what happens when I plot our wind speed forecasts for two successive Carrington rotations, designated C_1 and $C_2 = C_1 + 1$. The wind at 1 AU at time t, originated at the sun at time $t - \delta t$ where δt ($v(t)$) depends on the wind speed $v(t)$ at the source region, and is typically about 3 to 4 days. The earliest wind forecast associated with the rotation C_2 synoptic magnetogram is due to wind emanating from the earliest Carrington longitude 360°, i.e., Carrington time (C_2, 360°) of rotation c_2. If t_0 is the time at which the Earth is at this Carrington longitude, I can denote the time delay associated with this wind parcel as $\delta t(v_2(t_0))$, where the subscript on v indicates that v has been determined by WSA using the second magnetogram. Now consider a forecast made using synoptic magnetogram C_1 for wind originating at the same Carrington time (C_1, 0°) = (C_2, 360°). This will arrive at 1 AU after a time delay $\delta t(v_1(t_0))$. Since $v_1(t_0)$ and $v_2(t_0)$ are not necessarily the same, it should be clear that there will almost always be a discontinuity in the wind speed forecast as I transition from one synoptic magnetogram to the next. This discontinuity may result in two forecasts for the same time, or may result in a time interval with no forecast. To avoid the problems which this introduces, the simplest remedy is to ignore any unmatched events (HSEs or polarity reversals) which are within 2 days of the start or end of a rotation.

[65] **Acknowledgments.** I would like to acknowledge that without C. N. Arge's extensive support this study would not have been possible. I also wish to acknowledge the support of numerous members of the support staff at the Community Coordinated Modeling Center. The OMNI data were obtained from the GSFC/SPDF Omniweb interface at http://omniweb.gsfc.nasa.gov. I would also like to acknowledge the anonymous referees whose comments helped to improve the papers clarity.

References

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*(A5), 10,465–10,479.

Crooker, N. U., M. E. Burton, G. L. Siscoe, S. W. Kahler, J. T. Gosling, and E. J. Smith (1996), Solar wind streamer belt structure, *J. Geophys. Res.*, *101*(A11), 24,331–24,341.

Harvey, J. W., and J. R. Worden (1998), New types and uses of synoptic maps, in *Synoptic Solar Physics, Astron. Soc. Pac. Conf. Ser.*, vol. *140*, edited by K. S. Balasubramaniam, pp. 155–160, Astron. Soc. of the Pac., San Francisco, Calif.

Kahler, S., N. U. Crooker, and J. T. Gosling (1998), Properties of interplanetary magnetic sector boundaries based on electron heat-flux flow directions, *J. Geophys. Res.*, *103*(A9), 20,603–20,612.

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

McPherron, R. L., G. Siscoe, and C. N. Arge (2004), Probabilistic forecasting of the 3-h ap index, *IEEE Trans. Plasma Sci.*, *32*, 1425–1438, doi:10.1109/TPS.2004.833387.

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.

Roberts, D. A., P. A. Keiter, and M. L. Goldstein (2005), Origin and dynamics of the heliospheric streamer belt and current sheet, *J. Geophys. Res.*, *110*, A06102, doi:10.1029/2004JA010541.

Smith, E. J. (2001), The heliospheric current sheet, *J. Geophys. Res.*, *106*(A8), 15,819–15,831.

Tsurutani, B. T., et al. (2006), Corotating solar wind streams and recurrent geomagnetic activity: A review, *J. Geophys. Res.*, *111*, A07S01, doi:10.1029/2005JA011273.

Zhang, J., M. W. Liemohn, J. U. Kozyra, M. F. Thomsen, H. A. Elliott, and J. M. Weygand (2006), A statistical comparison of solar wind sources of moderate and intense geomagnetic storms at solar minimum and maximum, *J. Geophys. Res.*, *111*, A01104, doi:10.1029/2005JA011065.

P. MacNeice, NASA Goddard Space Flight Center, Mail Code 674, Greenbelt, MD 20771, USA. (peter.j.macNeice@nasa.gov)