Click
Here
for
**Full
Article**

# Validation of community models: 2. Development of a baseline using the Wang-Sheeley-Arge model

Peter MacNeice[1]

[1] This paper is the second in a series providing independent validation of community models of the outer corona and inner heliosphere. Here I present a comprehensive validation of the Wang-Sheeley-Arge (WSA) model. These results will serve as a baseline against which to compare the next generation of comparable forecasting models. The WSA model is used by a number of agencies to predict Solar wind conditions at Earth up to 4 days into the future. Given its importance to both the research and forecasting communities, it is essential that its performance be measured systematically and independently. I offer just such an independent and systematic validation. I report skill scores for the model's predictions of wind speed and interplanetary magnetic field (IMF) polarity for a large set of Carrington rotations. The model was run in all its routinely used configurations. It ingests synoptic line of sight magnetograms. For this study I generated model results for monthly magnetograms from multiple observatories, spanning the Carrington rotation range from 1650 to 2074. I compare the influence of the different magnetogram sources and performance at quiet and active times. I also consider the ability of the WSA model to forecast both sharp transitions in wind speed from slow to fast wind and reversals in the polarity of the radial component of the IMF. These results will serve as a baseline against which to compare future versions of the model as well as the current and future generation of magnetohydrodynamic models under development for forecasting use.

## 1. Introduction

[2] Independent validation is an essential stage in the migration of forecast-capable models from the research community to the operational world. There are a number of models currently in development in the heliophysics community which are expected to transition to operational use by the space weather forecasting community in the next five to ten years. In this paper I lay the groundwork for a systematic validation of a particular class of models, namely models of the corona and inner heliosphere. I present a comprehensive validation of the Wang-Sheeley-Arge (WSA) model [Arge and Pizzo, 2000; Arge et al., 2003], and present the results in a form which is intended to serve as a baseline against which other model's performances can be gauged.

[3] The WSA model is the most advanced of a class of models of the corona and inner heliosphere, that are based upon potential field approximations. It is used by a number of agencies as a prediction tool for Space Weather, and its coronal component is also used in the research community as a driver of kinematic and magnetohydrodynamic (MHD) models of the inner heliosphere [Odstrcil, 2003; Cohen et al., 2007; Fry et al., 2007]. It is essential therefore that the quality of its predictions be thoroughly validated.

[4] The WSA model authors have reported on its performance in a number of publications [Arge and Pizzo, 2000; Arge et al., 2003] which track its development and refinement over the last decade. In addition, Lee et al. [2009] have reported a detailed comparison of the predictions of the WSA model when combined with the ENLIL 3-D MHD model [Toth and Odstrcil, 1996; Odstrcil, 2003] of the heliosphere. These studies have been of a more scientific caste, focussing on the ability of the models to reproduce certain types of structure in the solar wind. Owens et al. [2005] provided an analysis of the WSA model's forecasting ability using a more systematic approach based on the use of a formal definition of skill score, and also measured the models ability to reproduce transitions from slow to fast wind. Owens et al. [2008] extended this study to compare the WSA model's performance with that of two coupled models, the WSA/ENLIL coupled model and the

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

MAS/ENLIL coupled model [*Linker et al.*, 1999; *Mikic et al.*, 1999; *Odstrcil et al.*, 2004].

[5] In this paper I report on a validation study of the WSA model (Version 1.6) installed at the Community Coordinated Modeling Center, which, when appropriate follows the approach of *Owens et al.* [2005], but which complements and extends it in a number of important ways. First, and most importantly, our study has been done independently of the WSA model authors. Second I use a later version of WSA and so to some extent our results can be used to track recent improvements in the model. Third, I use magnetograms from multiple sources, whereas *Owens et al.* [2005] used only Mount Wilson magnetograms, and *Owens et al.* [2008] used only National Solar Observatory magnetograms. Finally, I perform a more detailed study of the influence of all the key factors (including tunable model parameters, magnetogram sources, magnetogram filtering and phase of the solar cycle) which can influence the models performance, as measured in this formal style. I use skill scores which I consider to be more directly relevant to practical reference models, in particular through the use of persistence models as reference models.

[6] Finally I offer a more comprehensive analysis of the ability of the WSA model to forecast specific events. I focus on two timeline features, the transition from slow to high-speed wind, and the occurrence of larger-scale reversals in the polarity of the radial component of the interplanetary magnetic field. A detailed description of our approach in analyzing these events is presented by *MacNeice* [2009]. A number of studies [*Lyatsky et al.*, 2007; *Perreault and Akasofu*, 1978] have shown that the geomagnetic disturbance on the ground in the dayside polar cap region is well correlated with the product of the solar wind speed and the southward component of the interplanetary magnetic field (IMF). Neither the WSA model, or the heliospheric MHD models adequately forecast the southward component of the IMF. Therefore I address only the solar wind speed's influence on the geomagnetic disturbance, which the models can better reproduce.

[7] In section 2 I briefly review the WSA model, and the published tests of its performance. In section 3 I describe a formal measure of the ability of the model when compared to specific measurements of near Earth Solar Wind obtained from the Operating Mission as Nodes on the Internet (OMNI) database [*King and Papitashvili*, 2005]. In section 4 I present the results of this analysis of the WSA model. In section 5 I summarize the implications of these results.

## 2. WSA Model

[8] The WSA model of the corona and heliosphere has been described in detail by *Arge and Pizzo* [2000] and *Arge et al.* [2003]. It extended a model originally conceived by *Wang and Sheeley* [1990]. It has a number of tunable parameters. In the interests of clarity, I provide a short summary of the model here, detailing the specific settings I have adopted for any of the tunable parameters.

[9] The WSA model has three components. Between the solar surface and a source surface radius (here chosen to be $r_{ss} = 2.5r_o$, where $r_o$ is the Solar radius) it uses a standard potential source surface (PFSS) model [e.g., *Altschuler and Newkirk*, 1969]. The input to the PFSS component is a synoptic line of sight photospheric magnetogram obtained from any one of a group of observatories. The model interpolates this magnetogram data onto a uniformly spaced grid on the solar surface. In our study the grid spacing is 2.5° in both latitude and longitude. The magnetic field is assumed to be radial, both at the solar surface and at the source surface.

[10] The second component is a pseudopotential model of the field in the region between $r_{ss}$ and an outer radius, designated $r_{cs}$. This component, based on the approach of *Schatten* [1971, 1972], temporarily modifies the sign of the radial field at $r_{ss}$ to be everywhere positive. It creates a potential solution between $r_{ss}$ and $r_{cs}$, assuming radial field boundary conditions. Finally it restores the true radial polarity in this new solution. The result of this numerical artifice of modifying the radial polarity, is to produce potential-like solutions while preventing any field line reconnection between radially outward and inward field. At the boundary between regions of opposite polarity there will be a thin current sheet which effectively models the base of the heliospheric current sheet.

[11] The third WSA component extends the model from $r_{cs}$ to 1 AU. It does this by assuming that the solar wind flows radially from $r_{cs}$, with a constant flow speed at $r_{cs}$ determined from an empirical formula influenced by two factors, the rate of divergence of the magnetic field at $r_{ss}$, and the proximity of the given field line to a coronal hole boundary. As the sun rotates, the wind speed at any fixed point on the nonrotating sphere of radius $r_{ss}$ can change, and so, along a given radius in an inertial frame, faster wind packets may catch up with slower packets launched along that radial line at an earlier time. To accommodate this, every 1/8 AU distance along the radius, the wind packets are permitted to interact, with the general result that faster packets cause slower ones to speed up while slower ones retard the faster ones. The packets are propagated to 1 AU where their speed, IMF polarity and arrival time are recorded.

[12] When run in stand-alone mode, the WSA model is usually run with $r_{cs} = 5r_o$. When used to drive MHD models of the heliosphere, such as ENLIL, it is run with $r_{cs}$ chosen to be safely outside both the wind's sonic and alfvenic points. This enables the heliospheric codes to assume supersonic inflow boundary conditions. The existing studies of the WSA/ENLIL code, for example, assumed $r_{cs} = 21.5r_o$. Therefore, in this study, I report results for both $r_{cs} = 5r_o$ and $21.5r_o$.

## 3. Metric Definitions

[13] I measure the model's performance in two different ways. The first uses a formal skill score approach. The second tests the ability of the model to accurately forecast

**Table 1.** Empirical Formulae for Wind Speed at $r_{ss}$[a]

| Designation | $a_1$ | $a_2$ | $a_3$ | $a_4$ | $a_5$ | $a_6$ | $a_7$ | $a_8$ |
|---|---|---|---|---|---|---|---|---|
| A | 240 | 675 | 1/4.5 | 1 | 0.8 | 2.8 | 1.25 | 3 |
| B | 200 | 750 | 1/4.5 | 1 | 0.8 | 3.8 | 3.6 | 3 |
| C | 250 | 680 | 1/3.0 | 1 | 0.8 | 4.0 | 4.0 | 1 |

[a]A for all cases except B for the NSO case when $r_{cs} = 21.5r_o$ and C for MWO.

the occurrence of transitions in the solar wind state, which are important for space weather. In particular, I search for sharp transitions from slow to fast wind speeds, and for sector boundary transitions where the radial component of the IMF changes sign.

### 3.1. Skill Score Definition

[14] In defining formal metrics for the model's performance I compare its prediction for two quantities, the solar wind bulk speed and the polarity of the radial component of the IMF, near Earth. The near-Earth measurements of these quantities were obtained in the form of hourly averages from the OMNI 2 database [*King and Papitashvili*, 2005]. Since the angular resolution of the interpolated magnetogram used in our WSA runs was 2.5°, and since one Carrington rotation takes 27.2753 days, the WSA model output was sampled at a 27.2753 × 2.5/360 days (or every 4.546 hours). The OMNI 2 data was averaged at the same times as the WSA samples, by first constructing a continuous time line using a piece-wise linear fit to the OMNI 2 hourly averages, and then integrating over the time bin for each WSA sample to recover the average OMNI 2 data value for that time bin.

[15] I measure the WSA model's performance relative to a set of simple standard reference models. The simplest of these is the "Mean" model. In the Mean model, the predicted value for each variable is simply the mean value of that variable in the observation data set. In this study I construct separate means for each Carrington rotation.

The other reference models I use are persistence models. For example, the "1 day persistence" model assumes that the expected value of a variable is given by its actual measured value 1 day before.

[16] Suppose I wish to evaluate the relative performance of two different models in matching a set of observations for the quantity $F$. The observed values during the specified time interval are given by the set $F_o(i)$ with $i = 1, N$. The corresponding model predictions from each model are given by the sets $F_m^A(i)$, $i = 1, n$, where the superscript $A$ designates the different model.

[17] For each model I compute the mean square difference of the model predictions with the observations

$$D_F^A = \frac{1}{N} \sum_{i=1}^{N} \left( F_m^A(i) - F_o(i) \right)^2. \qquad (1)$$

To evaluate the performances of models $A$ and $B$ relative to each other I compute a "Skill Score" [*Brier*, 1950; *Wilks*, 1995], defined as

$$M_F^{AB} = 1 - \frac{D_F^A}{D_F^B}. \qquad (2)$$

Skill scores range from $-\infty$ to $+1$. Values of $M_F^{AB}$ greater than zero indicate that model $A$ does better than model $B$. Values less than zero indicate model $B$ is better than model $A$. Note, our skill score definition differs by a factor of 1/100 from that used by *Owens et al.* [2008]. To compare WSA with the Mean model, for example, model $A$ would be the WSA model and model $B$ the Mean model.

### 3.2. Timeline Event Detection

[18] The skill score approach reduces the model performance to a single number. This approach has the virtues of simplicity, reproducibility and a history of acceptance in
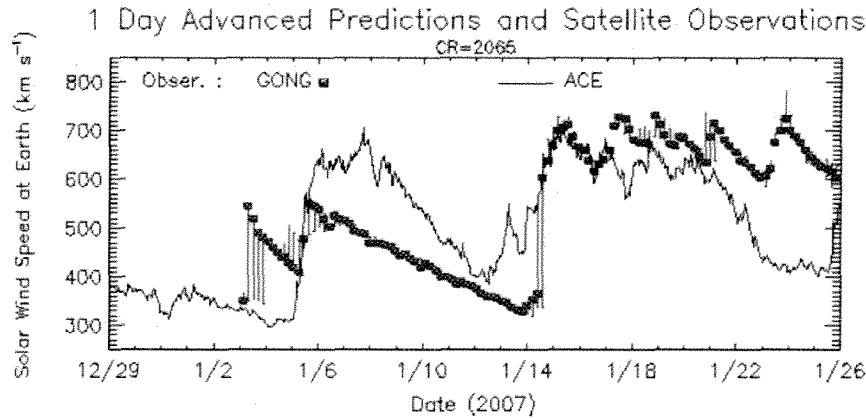


**Figure 1.** A typical 1 day advance prediction plot for solar wind speed produced by the WSA model(blue line) compared with OMNI measurements. For this prediction a magnetogram from the GONG network was used. Empirical velocity formula A from Table 1 was used.
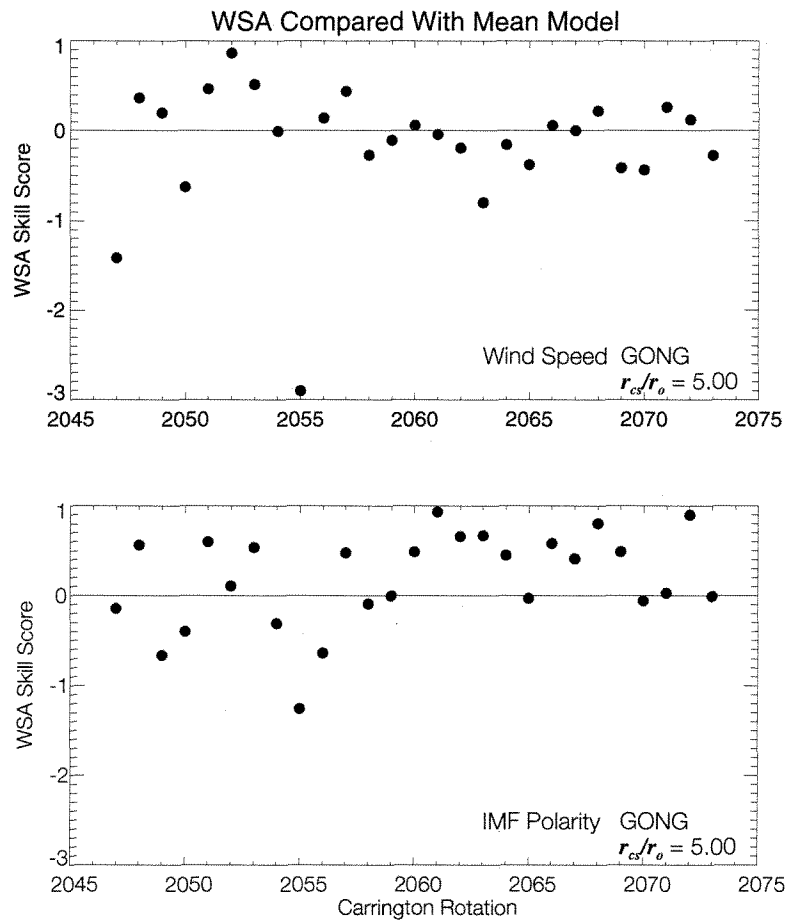
MACNEICE: VALIDATION OF COMMUNITY MODELS, 2

WSA Compared With Mean Model



**Figure 2.** Skill scores for the WSA solar wind speed and IMF polarity predictions for near Earth, relative to the mean model, for Carrington rotations from 2047 to 2070, based on GONG synoptic magnetograms, using $r_{cs} = 5r_o$ and the velocity formula A.

the scientific literature. However its simplicity is also its greatest weakness. It is not hard to construct data sets which would return skill scores for competing models that would contradict a readers expectation of the merits of the competing models, based on simple visual inspection.

[19] The skill score approach also gives no indication of the model's ability to forecast specific types of signal in the data. For example, sharp transitions from slow to fast wind can cause geomagnetic disturbances. It is important to characterize the model's ability to predict these.

[20] *Owens et al.* [2005] devised an approach to identify these HSEs in the model output and associate them with observed enhancements. They then characterized the hit/miss performance of the model. Their approach was to define a high-speed event (HSE) as an event in which a speed gradient threshold, in their case 50 km/d, was sustained for a minimum duration, in their case 2 days. I have essentially followed this approach. I found the *Owens et al.* [2005] event description overly simplistic and so made some modifications in the details of its imple-

mentation, which are described in detail by *MacNeice* [2009]. The most significant of these modifications ensured that HSEs with a rise time of less than 1 day were not excluded from the list of detected enhancements.

[21] One of our principal goals is to cast the results of our analysis in terms of forecast probabilities. The statistical summary of the numbers of hit and misses gives one view of the quality of the model. However it is not obvious how this translates into forecast probabilities. A forecaster wants to know the answer to the question(s), "If the model predicts (does not predict) a HSE in the next 24 hours, what is the probability that there will (not) be a HSE in the next 24 hours?" To answer this question, I determine the answer for each time point in our timelines. Then I total the results for all possible outcomes. This is the most direct way to provide the answer. Each of our time points represents an experimental test of the model. However these tests are not independent for time points separated by less than the "forecast window" (which in our case is

**Table 2.** Average Skill Scores for Solar Wind Speed and $B_r$ Polarity Forecasts at 1 AU for All Available Data[a]

| | NSO | | MWO | | GONG | |
|---|---|---|---|---|---|---|
| | $r_{cs}/r_o = 5.0$ | $r_{cs}/r_o = 21.5$ | $r_{cs}/r_o = 5.0$ | $r_{cs}/r_o = 21.5$ | $r_{cs}/r_o = 5.0$ | $r_{cs}/r_o = 21.5$ |
| | | | *Wind Speed* | | | |
| Reference Model | | | | | | |
|   Mean | −0.59 | −2.71 | −0.81 | −0.89 | −0.16 | −0.31 |
|   Persistence (1 day) | −0.95 | −3.59 | −1.19 | −1.27 | −0.70 | −0.98 |
|   Persistence (2 day) | −0.02 | −1.39 | −0.16 | −0.21 | 0.27 | 0.14 |
|   Persistence (4 day) | 0.29 | −0.66 | 0.18 | 0.14 | 0.54 | 0.46 |
|   Persistence (8 day) | 0.29 | −0.63 | 0.21 | 0.17 | 0.44 | 0.35 |
| | | | *$B_r$ Polarity* | | | |
| Reference Model | | | | | | |
|   Mean | −0.11 | −0.19 | −0.07 | −0.28 | 0.19 | 0.14 |
|   Persistence (1 day) | −0.88 | −1.00 | −0.83 | −0.86 | −0.41 | −0.46 |
|   Persistence (2 day) | −0.01 | −0.10 | 0.01 | −0.01 | 0.25 | 0.21 |
|   Persistence (4 day) | 0.29 | 0.33 | 0.40 | 0.38 | 0.59 | 0.57 |
|   Persistence (8 day) | 0.29 | 0.52 | 0.56 | 0.55 | 0.72 | 0.70 |

[a]Here observatories are the National Solar Observatory (NSO), Mount Wilson (MWO) archives, and Global Oscillation Network Group (GONG).
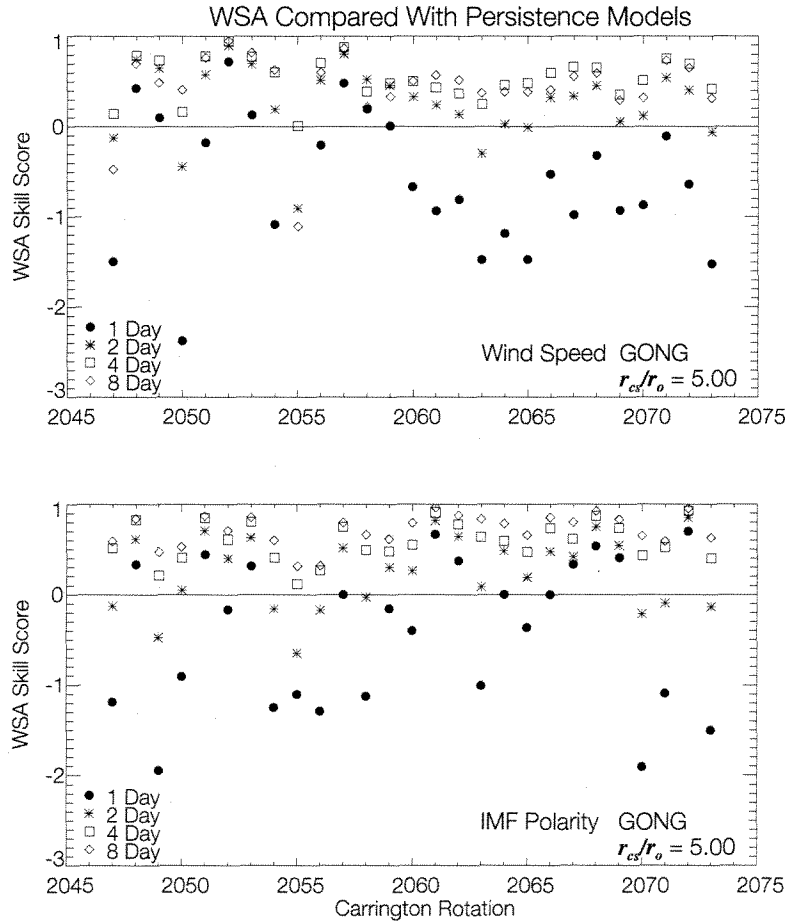


**Figure 3.** Skill scores for the WSA solar wind speed and IMF polarity predictions near Earth, relative to a set of persistence models, for Carrington rotations from 2047 to 2070, based on GONG synoptic magnetograms, using $r_{cs} = 5r_o$ and the velocity formula A. Symbols are labeled according to the period of persistence in the bottom left.
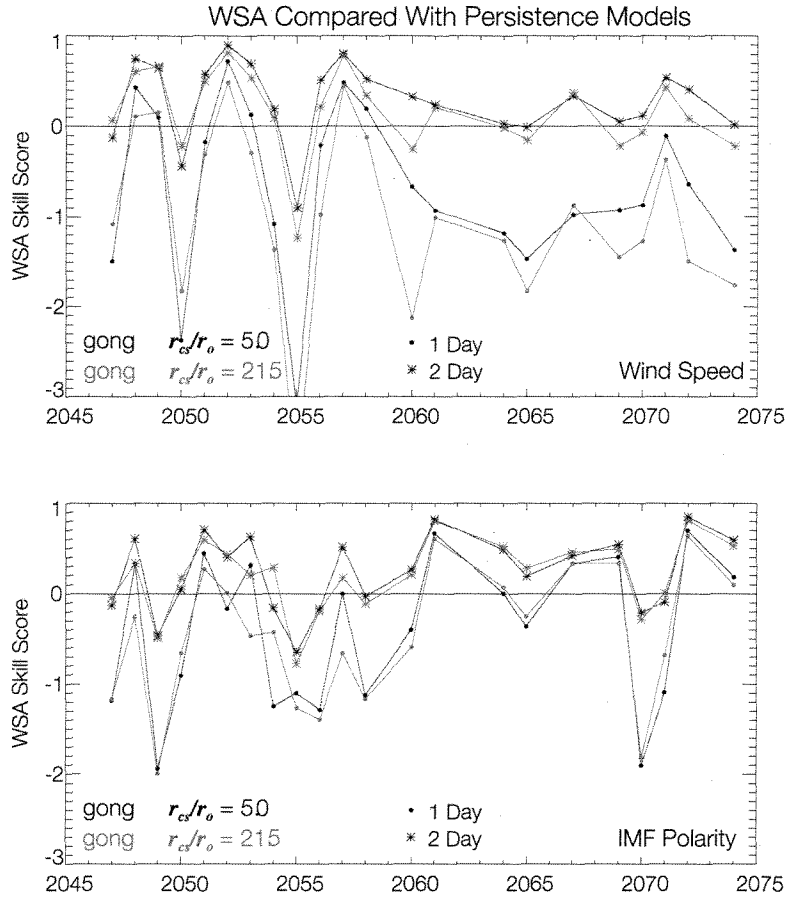
**Figure 4.** Skill scores, relative to the 1 and 2 day persistence models, of the WSA model for $r_{cs} = 5r_o$ and $21.5r_o$, using velocity formula A and Gong synoptic magnetograms around solar minimum.

24 hours), which makes it difficult to assign error bars to the probabilities that I derive.

## 4. Results

[22] I ran the WSA model (version 1.6) for the complete archive of full rotation synoptic maps available from the Global Oscillation Network Group (GONG), National Solar Observatory (NSO) and Mount Wilson (MWO) archives. I excluded any maps with clear flaws, including missing data or excessively noisy field patches (I provide a list of the excluded data in Appendix A). I ran the model for both $r_{cs} = 5r_o$ and $r_{cs} = 21.5r_o$, and with the empirical velocity formulae as tuned both for stand-alone WSA execution or for runs when used as a driver for the ENLIL MHD heliosphere model. For each of these runs I computed skill scores for the model relative to the Mean and Persistence (1 day, 2 days, 4 days, and 8 day) models.

[23] The model is tuned for each magnetogram source and depending on whether the WSA kinematic wind model is being used, or whether the output is intended for use with the ENLIL MHD heliospheric model. The

tuning is accomplished by use of different formulae associating the wind velocity at $r_{ss}$ with the field line divergence and coronal hole proximity. In Table 1 I list the different wind speed tunings that I use in this study. The general formula is [Owens et al., 2008]

$$v(f_s, \theta_b) = a_1 + a_2(1 + f_s)^{-a_3}\left(a_4 - a_5\ e^{-(\theta_b/a_6)^{a_7}}\right)^{a_8}\ \text{km s}^{-1}.$$

$$(3)$$

Here $f_s$ is the rate at which a magnetic flux tube at $r_{ss}$ expands compared to a purely radial expansion and $\theta_b$ is the minimum angular separation at the photosphere between an open field foot point and the nearest coronal hole boundary.

[24] In Figure 1 I show a typical WSA prediction plot, in this case a 1 day advance prediction of the solar wind speed during Carrington rotation number 2065.

[25] In the interests of precision, I note that the predictions in this plot do not begin 1 day after the start date (29 December 2007) of the Carrington Rotation, but typi-
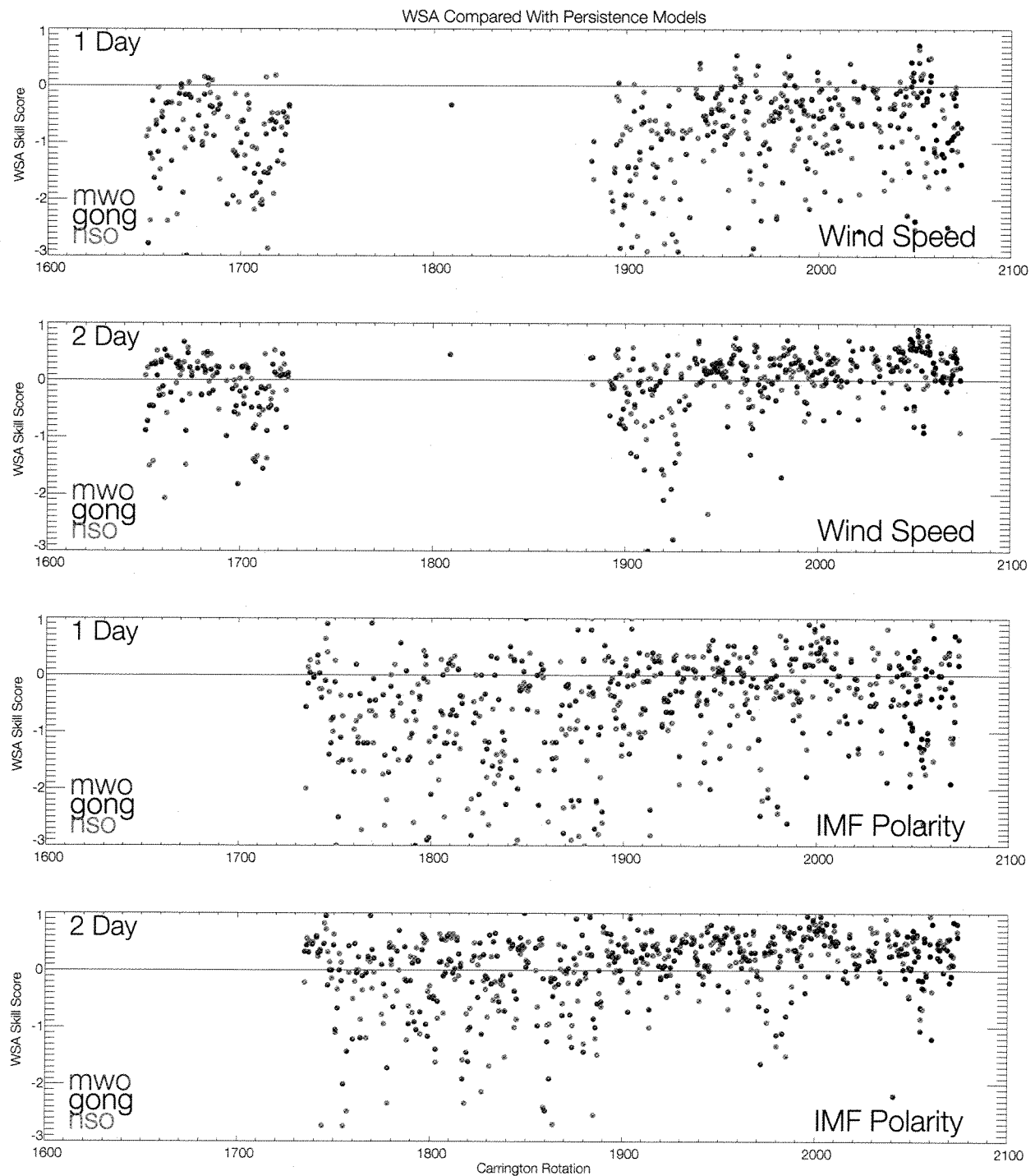
**Figure 5.** Comparison of skill scores, relative to the 1 and 2 day persistence models, when using magnetograms from different observatories, with $r_{cs} = 5r_o$.

**Table 3.** Average Skill Scores for Solar Wind Speed and $B_r$ Polarity Forecast at 1 AU Only for Carrington Rotations With Data From All Three Observatories, With $r_{cs} = 5r_o$

| | Wind Speed | | | $B_r$ Polarity | | |
|---|---|---|---|---|---|---|
| | NSO | MWO | GONG | NSO | MWO | GONG |
| Reference Model | | | | | | |
| Persistence (1 day) | −0.77 | −0.77 | −0.71 | −0.53 | −0.53 | −0.42 |
| Persistence (2 day) | 0.27 | 0.27 | 0.28 | 0.19 | 0.15 | 0.23 |

cally about 4 or 5 days after the start time of the rotation. The delay is due to the finite propagation time of the solar wind from the sun to 1 AU. To make a prediction prior to this would require the model to reference the previous rotation's synoptic magnetogram. While it would be possible for us to modify the model to do this, it would introduce discontinuous behavior in the time profiles as the prediction signal points of origin transitioned from one map to the next. It would also mean I were not validating the WSA model, but our own unique modified version. To avoid these complication I deliberately retain the approach currently adopted by the standard WSA model. Hence, when I report a skill score for a particular Carrington rotation, I am in fact analyzing the model prediction for the time interval from this delayed arrival time to the end time of the chosen rotation.

[26] For our analysis I excluded synoptic magnetograms with missing or bad data. I also limited the list of Carrington rotations included in our metrics by excluding those for which more than one third of the the solar wind data values were bad or missing. The list of exclusions is presented in detail in Appendix A.

## 4.1. Skill Scores

[27] Figure 2 shows the WSA skill scores, when compared with the Mean model, for GONG monthly synoptic maps from rotation 2047 to 2074. For this plot I ran WSA with $r_{cs} = 5r_o$ and using the velocity formula A in Table 1. During this period WSA V1.6 is comparable in quality to the Mean model for both wind speed and IMF polarity predictions. Here, the IMF polarity is defined as $B_r/|B_r|$, where $r$ is the RTN coordinate, with positive $r$ axis pointing away from the Sun. The average skill score for these measures are reported in Table 2.

[28] It should be noted that in April 2008 (CR2069), the GONG network adjusted the way their magnetogram processing algorithm determines polar fields (G. Petrie, private communication, 2009), with the result that polar coronal holes have their field strength more enhanced relative to equatorial coronal holes, than was the case prior to the adjustment. The GONG synoptic magnetograms that I used in this study are affected by this change. However the skill scores reported in Figure 2 show no obvious sensitivity to this change, though the number of points affected (CRs 2069 through 2074) is too small to make a definite conclusion.

[29] Because of its simple definition, the Mean model is useful in demonstrating our formal procedures. However as defined above, it has no use as a practical forecasting model, since it can only be constructed once the particular Carrington rotation is complete. Persistence models are a more practical class of simple reference models.

[30] Figure 3 shows WSA model skill scores relative to a set of persistence models. I consider four persistence reference models. For example, for 1 day persistence, the reference model predicts that the observed OMNI signal on day $d$ predicts the OMNI signal on day $d + 1$.

[31] WSA V1.6 is generally not quite as good as a 1 day persistence model, but is usually a little better than 2 day persistence, and is significantly more reliable than 4 or 8 day persistence. This is true for both wind speed and IMF polarity predictions.

### 4.1.1. Influence of $r_{cs}$ Setting

[32] When the WSA model is run alone, the recommended setting for $r_{cs}$ is $5r_o$. When coupled with an MHD model of the heliosphere, such as ENLIL, $r_{cs}$ must be set to a value which is beyond the point at which the wind speed exceeds both the local sound and Alfven speeds. This is required to ensure that the heliospheric model can assume supercritical inflow boundary conditions at its inner radial boundary. Of course, modification of $r_{cs}$ has implications for the quality of the coronal model. It is therefore of interest to determine if our metric test detects any significant degradation in the quality of this coronal model.

[33] In Figure 4 I show the skill score comparison, relative to the 1 and 2 day persistence models, using GONG data. In this case there no significant change in skill score for prediction of either wind speed or IMF polarity signal as a consequence of the change in $r_{cs}$.

[34] Table 2 confirms that this insensitivity of skill score to the change in $r_{cs}$ is true for both wind speed and IMF polarity, regardless of which observatory I use as a source of magnetogram, and regardless of choice of reference model (i.e., mean or persistence). There are just three exceptions, for wind speed when using NSO magnetograms in combination with mean or 1 or 2 day persistence models, where the choice of $r_{cs} = 5r_o$ is superior on average.

### 4.1.2. Influence of Magnetogram Source

[35] In Figure 5 I compare skill scores relative to 1 and 2 day persistence for model runs using magnetograms from all three observatories. The average skill scores for all "observatory/$r_{cs}$/reference model" combinations are given in Table 2. This shows that the choice of magnetogram source has no significant impact on skill scores. The GONG averages are slightly better than NSO or MWO, but I consider the advantage to be minimal. Since the GONG archive time coverage is much more limited than either NSO or MWO, I also computed average scores for just the rotations common to all three data sources. As Table 3 indicates, for 1 or 2 day persistence reference models, with $r_{cs} = 5r_o$, the model predicts almost identical average skill scores, regardless of magnetogram source.
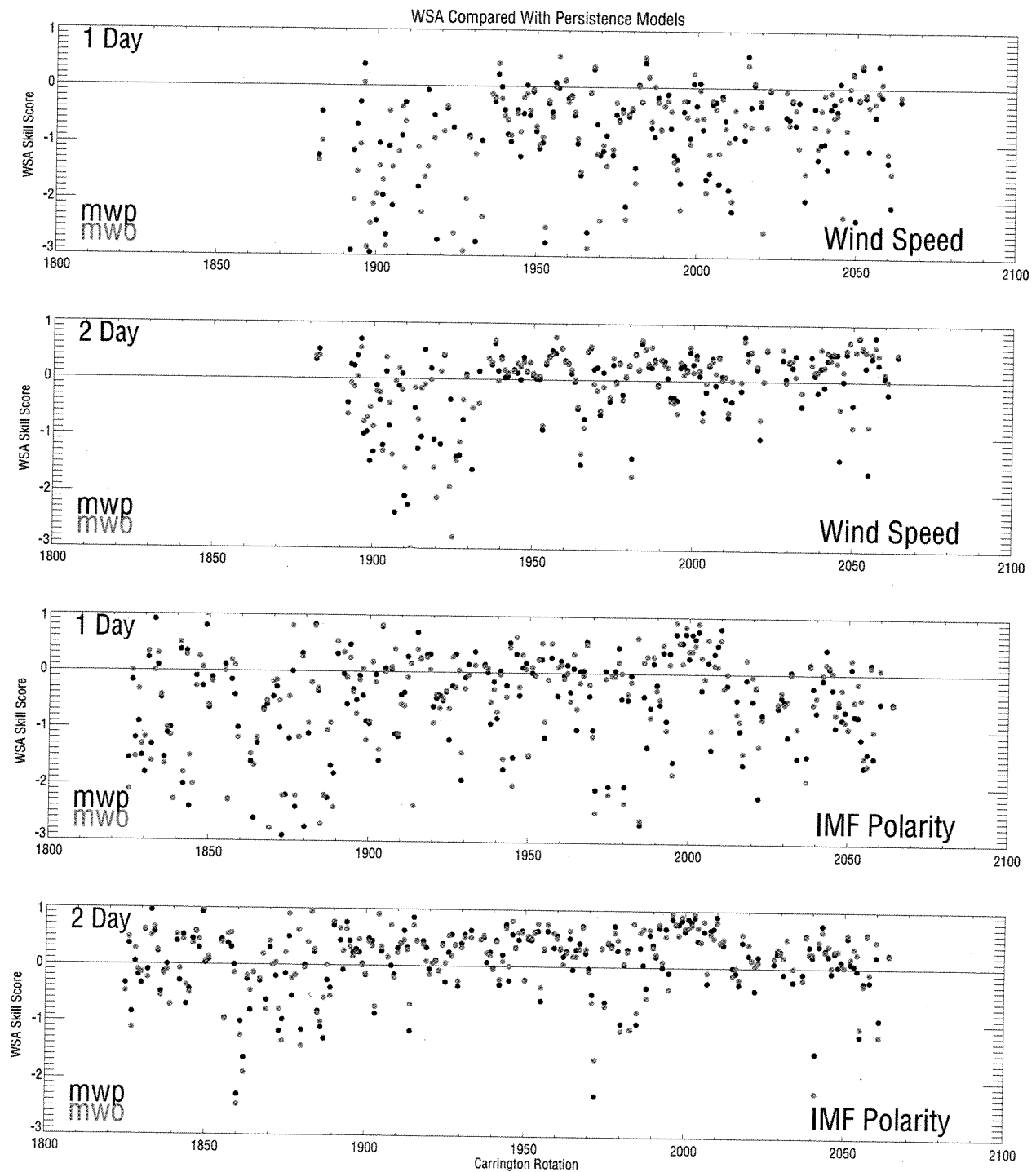
**Figure 6.** Comparison of skill scores relative to the 1 and 2 day persistence models, when using Mount Wilson magnetograms with(MWP) and without(MWO) temporal filtering of the polar fields, with $r_{cs} = 5r_o$.

**Table 4.** Average Skill Scores for Solar Wind Speed and $B_r$ Polarity Forecast at 1 AU for $r_{cs} = 5r_o$ Using Mount Wilson Magnetograms With and Without Temporal Filtering of the Polar Fields[a]

| | Velocity | | $B_r$ Polarity | |
| --- | --- | --- | --- | --- |
| | MWP | MWO | MWP | MWO |
| Reference Model | | | | |
| Persistence (1 day) | −1.34 | −1.19 | −0.68 | −0.61 |
| Persistence (2 day) | −0.23 | −0.15 | 0.10 | 0.13 |

[a]Here MWP is the Mount Wilson magnetogram with temporal filtering of the polar fields and MWO is the Mount Wilson magnetogram without temporal filtering of the polar fields.

### 4.1.3. Effect of Temporal Smoothing of Polar Fields

[36] Measurement of the line of sight (LOS) component of the magnetic field is challenging near the Solar poles for two reasons. For significant periods during each year, each pole is hidden from the view of Earth-centric observers because of the 7.25 degree tilt of the Earth's orbit relative to the Solar equator. In addition if the polar fields are close to radial they will produce a very weak signal in the LOS measurements. As a result, the polar field data measurements are often noisy. *Arge and Pizzo* [2000] have suggested reducing these noise levels by fitting the polar fields using a temporal extrapolation from Carrington rotation maps which are close in time to the current map.

[37] Arge has applied that only to Mount Wilson magnetograms. Using a set of Mount Wilson magnetograms processed using this formula for Carrington rotations 1824 to 2064, which C. N. Arge (private communication, 2009) provided for us. I have tested the effect of this data processing on the skill scores. Results are plotted in Figure 6 and in Table 4 I show the average skill scores for these two approaches. Use of this temporal smoothing of the polar fields has no significant effect on the skill scores.

### 4.1.4. Quiet Versus Active Period Performance

[38] To test if the WSA model is better tuned for quiet or active periods, I divided the Carrington rotations into two groups. The first set of "quiet" rotations (<1665, 1735−1800, 1860−1930 and >2000) covers periods where the sunspot totals shown in Figure 7 are below 2000. The second set is all the remaining rotations in our data set. In Table 5 I present the skill scores for the model runs using NSO magnetograms. This shows that there is no significant difference in the WSA models skill scores for active and quiet periods. This is also true when using Mt Wilson magnetograms.

### 4.2. Feature Specific Validation

[39] As discussed above, I analyzed the model's ability to forecast the transitions from slow to fast wind, and the occurrence of reversals in the sign of $B_r$ associated with large-scale heliospheric current sheet structure. The results are summarized in Table 6, and cast in terms of forecast probabilities in Tables 7 and 8.

### 4.2.1. HSE Forecasts

[40] Table 6 reports the number of HSEs recorded by OMNI and predicted by the model for each combination of observatory and outer radius. It also reports the numbers
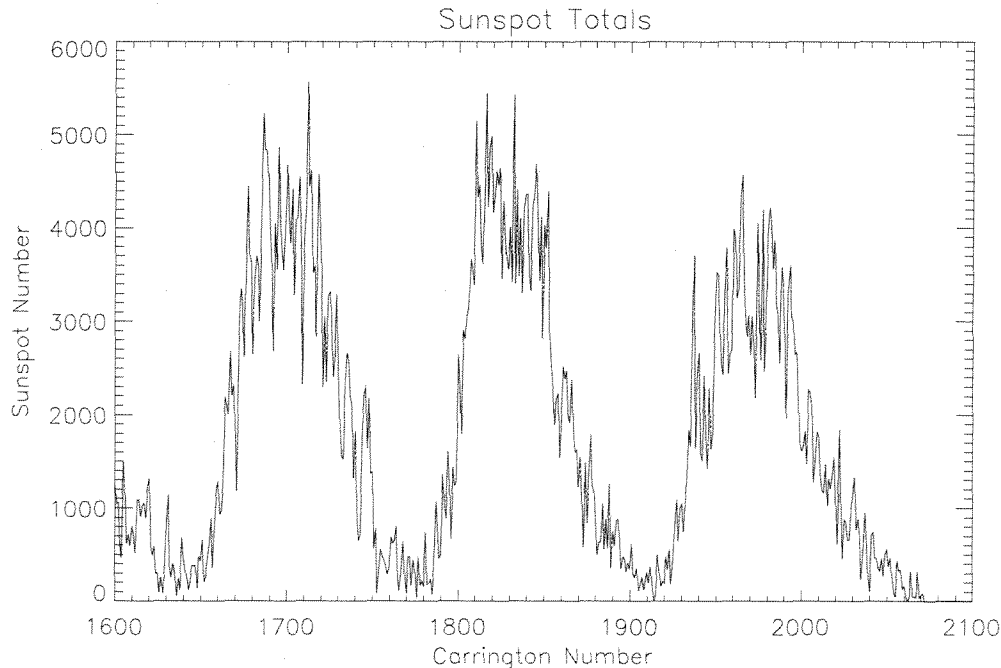


**Figure 7.** Sunspot numbers as a function of the Carrington number.

**Table 5.** Average Skill Scores Solar Wind Speed and $B_r$ Polarity Forecast at 1 AU for Both Quiet and Active Solar Conditions[a]

|  | Velocity | | $B_r$ Polarity | |
|---|---|---|---|---|
|  | Quiet | Active | Quiet | Active |
| Reference Model |  |  |  |  |
| Persistence (1 day) | −1.10 | −0.87 | −0.96 | −0.87 |
| Persistence (2 day) | −0.04 | −0.03 | −0.07 | 0.00 |

[a]Here $r_{cs} = 5r_o$, and I used synoptic magnetograms from the National Solar Observatory.

of hits, misses, and false positives. For $r_{cs} = 5r_o$, for all three magnetogram sources, the model averages a HSE hit rate of 40%, while of the forecast HSEs, 39% were false positives.

[41] With $r_{cs} = 5r_o$, the model performs as well with NSO and Mount Wilson data. The setting $r_{cs} = 21.5r_o$ results in a noticeable increase in HSE misses and false positives. This confirms that the WSA model is better tuned for $r_{cs} = 5r_o$. It implies that MHD codes for the inner heliosphere which use the $r_{cs} = 21.5r_o$ case output at their inner boundary may suffer from the same elevated rates of misses and false positives when forecasting HSEs, unless this tuning is adjusted to allow for the difference in $r_{cs}$. It should be pointed out that WSA(V1.6) does have specific tuning for some observatory/$r_{cs}$ cases specifically for use with ENLIL, which allows for the different propagation time of the wind between the surface $r_{cs}$, and also allows for differences in solar wind propagation in the inner heliosphere between the kinematic treatment of WSA and the more complete physical description in the ENLIL code.

[42] The distribution of timing errors for all HSEs from the $r_{cs} = 5r_o$ cases for all three observatories is shown in Figure 8. The distribution is slightly skewed in favor of hits and misses for which the forecast HSE precedes the associated OMNI HSE. The mean timing error for hits is −0.25 days, while the mean absolute error in $|dt|$ for hits is 0.94 days.

[43] All the significant variation in HSE timing error is within the range of $|dt| < 2$ days, while outside this range the distribution for misses and hits is flat. This distribution of timing errors confirms that the time window of

2 days used in our hit detection algorithm is a reasonable choice.

[44] It should be noted that the time windows and the data binning for each CR used in constructing Table 6 are determined from the WSA model's forecast window, and so are weakly dependent on $r_{cs}$. This is why the number of OMNI HSEs (or OMNI polarity reversals) for a given observatory case seems to vary slightly with $r_{cs}$.

[45] To construct the forecast probabilities presented in Table 7 I consider each time bin in our data set and ask the question, if WSA does or does not predict a HSE within the next 24 hours, does OMNI report a HSE within that time window? Table 7 presents the probabilities for each possible case, as a function of magnetogram source. It also lists the weighted average for our entire data set.

[46] Not surprisingly, since most 24 hour intervals do not have a HSE, the model is much more reliable when asked to predict the absence of an HSE than when asked to match an occurrence. On average, the model is accurate only 17% of the time when it predicts a HSE will occur in the next 24 hours, but is accurate 94% of the time when it predicts there will be no HSE in the next 24 hours.

[47] Notice, I deliberately framed the question from the operational forecaster's perspective, i.e., given the model prediction which the forecaster has in hand, what is the probability that an event will happen.

### 4.2.2. $B_r$ Polarity Forecasts

[48] The model correctly reproduces the observed large-scale radial IMF polarity 82% of the time for GONG based forecasts, 75% of the time for NWO, and 76% for MWO, with an overall average of 76%. The model misses 14% of the polarity phases reported by OMNI. This result is almost completely insensitive to the magnetogram source. Using GONG data it missed 16, or 13%, of the 119 polarity phases reported by OMNI for the same period. For NSO it missed 234 (14%) phases out of 1699, while for MWO it missed 233 (14%) out of 1709.

[49] Table 6 reports the number of polarity reversal recorded by OMNI and predicted by the model for each combination of observatory and outer radius. It also reports the numbers of hits, misses, and false positives. For $r_{cs} = 5r_o$, for all three magnetogram sources, the model

**Table 6.** HSE and $B_r$ Polarity Reversal Matching by the WSA Model[a]

| Observatory | $r_{cs}/r_o$ | HSEs | | | | | $B_r$ Polarity Reversals | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
|  |  | OMNI HSE Total | WSA HSE Total | Hits | Misses | False Positive | OMNI Reversal Total | WSA Reversal Total | Hits | Misses | False Positive |
| GONG | 5.0 | 54 | 36 | 23 | 30 | 11 | 91 | 58 | 54 | 25 | 2 |
| GONG | 21.5 | 55 | 42 | 27 | 27 | 11 | 91 | 53 | 50 | 32 | 2 |
| MWO | 5.0 | 436 | 319 | 170 | 222 | 135 | 1284 | 932 | 777 | 398 | 111 |
| MWO | 21.5 | 458 | 346 | 164 | 243 | 163 | 1313 | 941 | 779 | 425 | 112 |
| NSO | 5.0 | 441 | 305 | 175 | 222 | 111 | 1279 | 936 | 783 | 403 | 104 |
| NSO | 21.5 | 428 | 309 | 121 | 249 | 175 | 1279 | 868 | 741 | 420 | 90 |

[a]Here HSE is, OMNI is, and WSA is the Wang-Sheeley-Arge model.

**Table 7.** Forecast Probabilities for Occurrence of HSEs Within 24 Hours of the Current Time[a]

| Model | Observation | GONG | MWO | NSO | Weighted Average |
|-------|-------------|------|-----|-----|------------------|
| WSA predicts HSE | OMNI HSE | 23 | 15 | 17 | 17 |
| WSA predicts HSE | No OMNI HSE | 77 | 85 | 83 | 83 |
| WSA predicts no HSE | OMNI HSE | 10 | 6 | 6 | 6 |
| WSA predicts no HSE | No OMNI HSE | 90 | 94 | 94 | 94 |

[a]Values given are percentages.

averages a reversal hit rate of 61%, while of the forecast reversals, 11% were false positives.

[50] In contrast to the HSE forecasts, the model forecasts of large-scale polarity reversal seem insensitive to the $r_{cs}$ setting.

[51] The mean timing error between the polarity reversals which were matched as hits was 1.1 days. In the top panel of Figure 9 I have plotted all reversals forecast by the model as a function of their timing error and average latitudinal distance from the heliospheric current sheet (HCS) during the time interval between the OMNI reversal and the matching WSA prediction, i.e., during the period of "error." Our method for computing this latitudinal distance from the HCS is described in detail by *MacNeice* [2009]. For reversal misses, I estimate the timing error by comparing with the closest observed reversal of the same reversal direction (i.e., + to −, or, − to +). 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 $360/27.27°/d$.

[52] Approximately 35% of the reversal hits have a timing error of less than 1 day. About 23% of the remaining matches average less 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°.

[53] The bottom left panel in Figure 9 shows a histogram of the generalized angular error. Generalized angular error is the distance of each point from the origin in the top panel. A generalized angular error of less than 14° is equivalent to a timing error of less than 1 day, which is indicated by the vertical dashed line. 35% of the matched reversals have a generalized error of less than 1 day.

[54] The bottom right panel in Figure 9 shows the distribution of the average offset in the model HCS from the ecliptic for polarity phases in the OMNI data which are entirely missed by the WSA model. 228 out of the 483, or 47%, of missed phases show a time averaged separation between the ecliptic and HCS of less than 5°.

[55] Table 8 presents the probabilities for the occurrence of polarity reversals based on the model forecasts. To prepare forecast probabilities for polarity reversals, I need also to consider whether the model nowcast is in agreement with the observed polarity when the model forecast is made. For example, if the polarity nowcast is accurate and WSA predicts a reversal in the next 24 hours, it is accurate 32% of the time. If the nowcast is accurate and

WSA predicts no reversal during the next 24 hours, its is correct 93% of the time.

## 5. Conclusions

[56] Reproducible quantitative measurements are vital in assessing the performance of models destined for Space weather forecasting use. The WSA model is of particular importance because it is generally considered to be the most accurate forecast model currently available to predict solar wind speed and IMF polarity at Earth. As the community develops more complex first principles based models to replace WSA, they will need to demonstrate their performance relative to the baseline which WSA establishes. In this paper I have presented such an evaluation.

[57] The skill score results indicate that for both wind speed and IMF polarity forecasts, the WSA model is generally inferior to a 1 day persistence model, but comparable to 2 day persistence, and generally superior to 4 and 8 day persistence. With current model tuning, this result is true regardless of the source observatory used. It is also true for both quiet and active periods.

[58] When the outer radius $r_{cs}$ of the current sheet component was pushed outward from 5 to 21.5 $r_o$, there was a slight degradation in average skill score for wind speed, but the average IMF polarity skill scores were virtually unchanged. The model skill scores were largely insensitive to the source of the magnetogram data for $r_{cs} = 5r_o$, and only weakly sensitive when $r_{cs} = 21.5r_o$. Surprisingly, the temporal filtering of the polar fields for the Mount Wilson magnetograms made no significant difference.

[59] The model correctly forecasts the IMF polarity 82% of the time. At current magnetogram resolution (2.5°) it predicts significantly fewer polarity reversals than were actually observed, and completely missed 14% of the observed polarity phases. It achieved a polarity reversal hit rate of 61%, and a false positive rate for reversals of 11%.

[60] The model also predicted fewer HSEs than were observed. It achieved a hit rate of 40%, and a false positive rate of 39%. This result is insensitive to the source of magnetogram, but the $r_{cs} = 21.5r_o$ setting resulted in a significant increase in HSE misses and false positives. The model slightly favored HSE timing errors in which the forecast HSE occurs before the observed HSE.

[61] I computed forecast probabilities for both polarity reversals and HSEs within a 24 hour forecast window.

**Table 8.** Forecast Probabilities for Predictions of Polarity of $B_r$ for All Combinations, When the Model Nowcast is Correct or Incorrect, and the Model Predicts or Does Not Predict a Polarity Reversal[a]

| Model | Observation | GONG Agreement | GONG Disagreement | MWO Agreement | MWO Disagreement | NSO Agreement | NSO Disagreement | Weighted Average Agreement | Weighted Average Disagreement |
|---|---|---|---|---|---|---|---|---|---|
| WSA predicts reversal | OMNI reversal | 36 | 9 | 31 | 6 | 33 | 6 | 32 | 6 |
| WSA predicts reversal | No OMNI reversal | 64 | 91 | 69 | 94 | 67 | 94 | 68 | 94 |
| WSA predicts no reversal | OMNI reversal | 7 | 37 | 7 | 29 | 7 | 29 | 7 | 29 |
| WSA predicts no reversal | No OMNI reversal | 93 | 63 | 93 | 71 | 93 | 71 | 93 | 71 |

[a]Values given are percentages. For each case the comment with the number indicates whether the model prediction is good or bad.

Averaged over all data sets, a HSE forecast within the next 24 hours will be correct 17% of the time. A forecast of no HSE in the net 24 hours will be accurate 94% of the time. For IMF polarity reversal forecasts, a forecast reversal within the next 24 hours will be accurate 32% of the time if the nowcast is accurate. If the nowcast is not accurate then if a reversal is forecast, there is a 94% chance that no reversal will occur, meaning that the model polarity will be in agreement with observation after the 24 hour period. If the now cast is accurate and no reversal is forecast, this will be accurate 93% of the time.

[62] From these results I can conclude that the WSA model is better at reproducing the ambient IMF polarity and forecasting reversals than it is at reproducing the wind speed and occurrence of HSEs.

[63] Direct comparison of our results with those of *Owens et al.* [2005, 2008] is not straightforward, given the differences in magnetogram data sets, model spatial resolution, size of time bins used in each analysis, methodology of analysis, and in the empirical formulae used for wind speed. The charcteristic properties of the observed and model solar wind speeds are similar in our study and that of *Owens et al.* [2008]. For example, the mean and standard deviation of the observed wind speed in our study for the (NSO, $r_{cs} = 5$) case are 436 km/s and 87.5 km/s, compared with 434 km/s and 99.2 km/s, respectively, by *Owens et al.* [2008]. For WSA, for the same case, I find a mean and standard deviation of the wind speed of 427 km/s and 69.9 km/s, compared with 411 km/s and 84.3 km/s, respectively, by *Owens et al.* [2008]. Our average root mean square error is 99.8 km/s compared with 94.9 km/s in their study. Since the mean square error is the basis for skill score computation, these similarities suggest that our skill scores are in general agreement with the results of *Owens et al.* [2008].

[64] Our algorithm for HSE detection gives a significantly lower hit rate (40%) and higher miss (60%) and false positive rates(39%) than *Owens et al.* [2008], who report 59% hit, 41% miss and 16% false positive. This difference is significant in assessing the absolute probability of successful forecasting, but is not important when assessing relative performance of different models. Finally, *Owens et al.* [2008] report no detectable average timing error for HSEs. Given that their HSE analysis uses 8 hour time binning, the average timing error of −0.25 days found in this study is consistent with their result.

[65] With the exception of the HSE hit rate, our results are generally consistent with those of *Owens et al.* [2005, 2008], but will serve as a more comprehensive and specific baseline for comparisons with future versions of the WSA models, and for MHD based forecast models in development.

## Appendix A: Data Selection

[66] I ran the WSA model with synoptic magnetograms for Carrington Rotations 1650 through 2074 from both NSO and Mount Wilson, and for rotations 2047 through
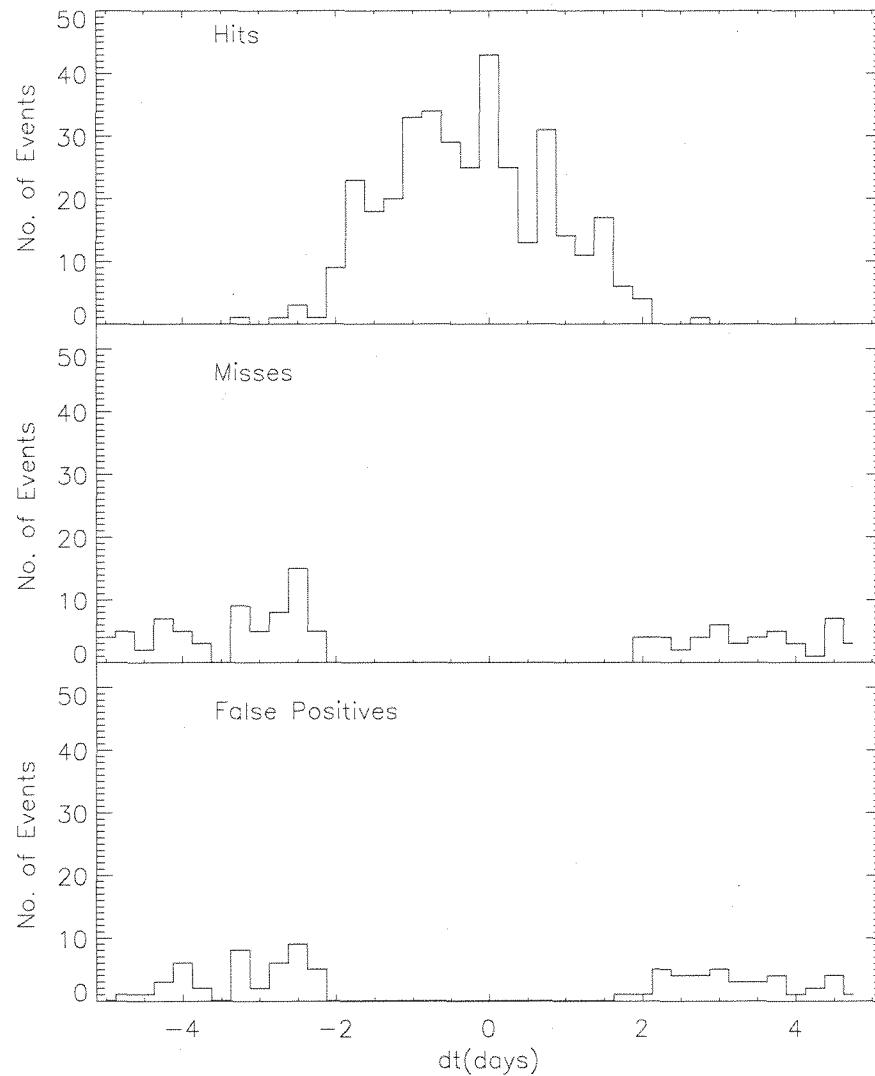
**Figure 8.** Summary of timing errors for HSE hits, misses, and false positives. Positive *dt* means the OMNI event occurred before the associated WSA model event.

2074 from the GONG network. Some of these synoptic magnetograms have bad or missing data. I have excluded these cases. The following is a list of the Carrington Rotations which were excluded from our skill score and timeline event analyses for this reason: 1650, 1663, 1665, 1666, 1676, 1677, 1679, 1690−1692, 1695, 1705, 1720, 1726, 1731−1734, 1738, 1740, 1756, 1759, 1766, 1772−1774, 1801, 1824, 1840, 1852−1854, 1866, 1891, 1917, 1932, 1934, 1935, 1958, 1973, 2013, 2024, 2025, 2026, 2027, 2036, and 2066 (from MWO) and 1661, 1663, 1665, 1837, 1860, 1973, 1981, 2008, 2009, 2014, 2015, 2016, 2019, 2026, 2033, 2035, 2059, 2062, 2063, and 2068 (from NSO).

[67] In almost all of the maps, the polar fields have large noise levels. I chose not to exclude maps because of excessively noisy polar fields because to do so would have reduced the available data set to a minimal subset, and because, in a forecasting environment, the model will have to function with data which has this flaw.

[68] For our analysis of the model's accuracy in forecasting reversals in the sign of $B_r$ chose to eliminate Carrington Rotation 1730 and earlier because there were too many dropouts in the measured signal in the OMNI database.

[69] For the analysis of wind speed forecast, I eliminate any rotations for which the fraction of bad or missing OMNI wind speed measurements was higher than 33%. The rotations that I excluded for this reason are CRs 1687, 1727−1882, 1884−1890.
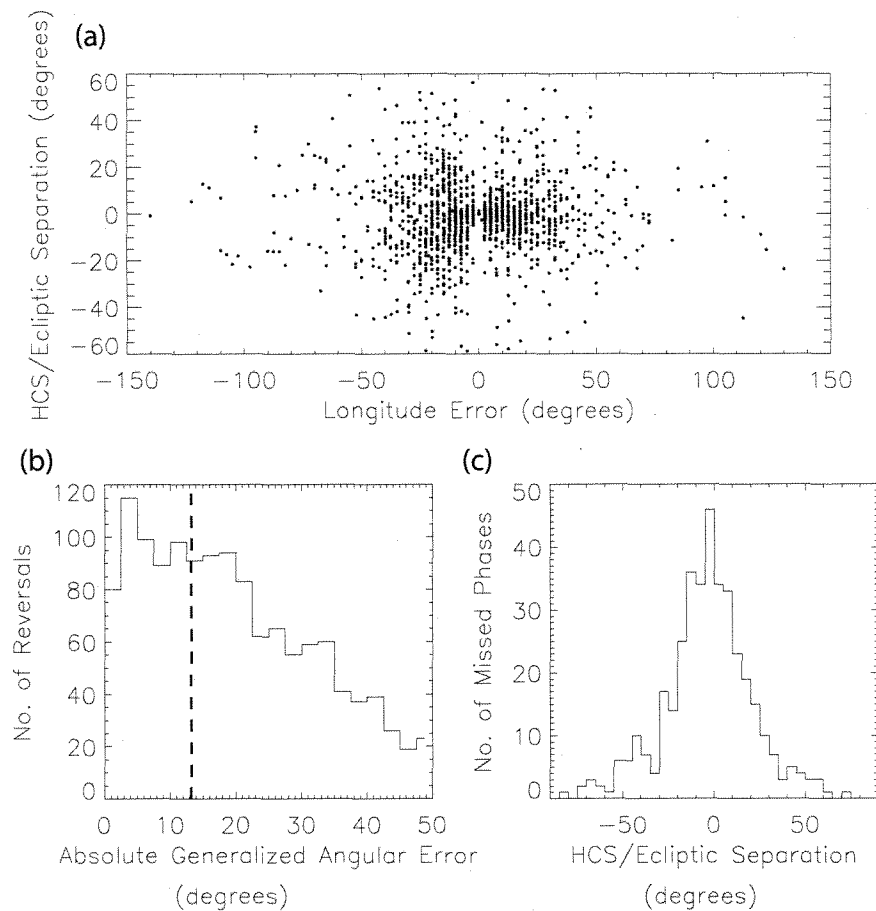
**Figure 9.** Summary of IMF $B_r$ polarity errors. (a) The timing errors and mean offset of the model HCS from the ecliptic during the period of polarity mismatch, for each reversal and the closest available match. (b) A histogram of the absolute generalized angular error for each reversal, and (c) the distribution of the mean offset of the model HCS from the ecliptic during polarity phases which were completely missed by the model.

# References

Altschuler, M. D., and G. Newkirk (1969), Magnetic fields and the structure of the solar corona. Part I: Methods of calculating coronal fields, *Solar 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(A5), 10,465–10,479.

Arge, C. N., D. Odstrcil, V. J. Pizzo, and L. R. Mayer (2003), Improved method for specifying solar wind speed near the Sun, *AIP Conf. Proc.*, 679, 190–193.

Brier, G. W. (1950), Verification of forecasts expressed in terms of probabilities, *Mon. Weather Rev.*, 78, 1–3.

Cohen, O., et al. (2007), A semiempirical magnetohydrodynamical model of the solar wind, *Astrophys. J.*, 654, L163–L166.

Fry, C. D., T. R. Detman, M. Dryer, Z. Smith, W. Sun, C. S. Deehr, S.-I. Akasofu, C.-C. Wu, and S. McKenna-Lawlor (2007), Real-time solar wind forecasting: Capabilities and challenges, *J. Atmos. Sol. Terr. Phys.*, 69, 109–115.

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.*, 110, A02104, doi:10.1029/2004JA010649.

Lee, C. O., J. G. Luhmann, D. Odstrcil, P. MacNeice, I. de Pater, P. Riley, and C. N. Arge (2009), The solar wind at 1 AU during the declining phase of solar cycle 23: Comparison of 3D numerical model results with observations, *Sol. Phys.*, 254, 155–183.

Linker, J. A., Z. Mikic, D. A. Biesecker, R. J. Forsyth, S. E. Gibson, A. J. Lazarus, A. Lecinski, P. Riley, A. Szabo, and B. J. Thompson (1999), Magnetohydrodynamic modeling of the solar corona during Whole Sun Month, *J. Geophys. Res.*, 104(A5), 9809–9830.

Lyatsky, W., S. Lyatskaya, and A. Tan (2007), A coupling function for solar wind effect on geomagnetic activity, *Geophys. Res. Lett.*, 34, L02107, doi:10.1029/2006GL027666.

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

Mikic, Z., J. A. Linker, D. D. Schnack, R. Lionello, and A. Tarditi (1999), Magnetohdrodynamic modeling of the global solar corona, *Phys. Plasmas*, 6, 2217–2224.

Odstrcil, D. (2003), Modeling 3-D solar wind structure, *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. Sol. 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), Metris for solar wind prediction models: Comparison of empirical, hybrid, and physics-based schemes with 8 years of L1 observations, *Space Weather*, 6, S08001, doi:10.1029/2007SW000380.

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

Schatten, K. H. (1971), Current sheet magnetic model for the solar corona, *Cosmic Electrodyn.*, 2, 232–245.

Schatten, K. H. (1972), Current sheet magnetic model for the solar corona, in *Solar Wind*, edited by C. P. Sonett, P. J. Coleman Jr., and J. M. Wilcox, pp. 44–54, NASA, Washington, D. C.

Toth, G., and D. Odstrcil (1996), Comparison of some flux corrected transport and total variation diminishing numerical schemes for hydrodynamic and magnetohydrodynamic problems, *J. Comput. Phys.*, 128, 82–100.

Wang, Y.-M., and N. R. Sheeley (1990), Solar wind speed and coronal flux-tube expansion, *Astrophys. J.*, 355, 726–732.

Wilks, D. S. (1995), *Statistical Methods in the Atmospheric Sciences*, Academic Press, San Diego, Calif.

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