

Constraining Global Coronal Models with Multiple Independent Observables

Samuel T. Badman, David H. Brooks, Nicolas Poirier, Harry P. Warren, Gordon Petrie, Alexis P. Rouillard, C. Nick Arge, Stuart D. Bale, Diego de Pablos Agüero, Louise Harra, et al.

▶ **To cite this version:**

Samuel T. Badman, David H. Brooks, Nicolas Poirier, Harry P. Warren, Gordon Petrie, et al.. Constraining Global Coronal Models with Multiple Independent Observables. The Astrophysical Journal, 2022, 932, 10.3847/1538-4357/ac6610 . insu-03867486

HAL Id: insu-03867486

https://insu.hal.science/insu-03867486

Submitted on 23 Nov 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

The Astrophysical Journal, 932:135 (29pp), 2022 June 20 https://doi.org/10.3847/1538-4357/ac6610
© 2022. The Author(s). Published by the American Astronomical Society.
OPEN ACCESS

Constraining Global Coronal Models with Multiple Independent Observables

Samuel T. Badman[1,2], David H. Brooks[3,4], Nicolas Poirier[5], Harry P. Warren[6], Gordon Petrie[7], Alexis P. Rouillard[5], C. Nick Arge[8], Stuart D. Bale[1,2], Diego de Pablos Agüero[9], Louise Harra[10,11], Shaela I. Jones[8,12], Athanasios Kouloumvakos[5], Pete Riley[13], Olga Panasenco[14], Marco Velli[15], and Samantha Wallace[8]
[1] Physics Department, University of California, Berkeley, CA 94720-7300, USA; samuel_badman@berkeley.edu
[2] Space Sciences Laboratory, University of California, Berkeley, CA 94720-7450, USA
[3] College of Science, George Mason University, 4400 University Drive, Fairfax, VA 22030 USA
[4] Hinode Team, ISAS/JAXA, 3-1-1 Yoshinodai, Chuo-ku, Sagamihara, Kanagawa 252-5210, Japan
[5] IRAP, Université Toulouse III—Paul Sabatier, CNRS, CNES, Toulouse, France
[6] Space Science Division, Naval Research Laboratory, Washington, DC 20375, USA
[7] National Solar Observatory, 3665 Discovery Drive, 3rd Floor, Boulder, CO 80303, USA
[8] GSFC, Greenbelt, MD 20771, USA
[9] UCL-MSSL, Holmbury St. Mary, Dorking, Surrey RH5 6NT, UK
[10] PMOD/WRC, Dorfstrasse 33, CH-7260 Davos Dorf, Switzerland
[11] ETH-Zurich, Hönggerberg Campus, HIT Building, Zürich, Switzerland
[12] Catholic University of America, 620 Michigan Avenue NE, Washington, DC 20064, USA
[13] Predictive Science Inc., San Diego, CA 92121, USA
[14] Advanced Heliophysics, Pasadena, CA 91106, USA
[15] EPSS, UCLA, Los Angeles, CA 90095, USA
Received 2022 January 27; revised 2022 March 16; accepted 2022 April 9; published 2022 June 27

Abstract

Global coronal models seek to produce an accurate physical representation of the Sun's atmosphere that can be used, for example, to drive space-weather models. Assessing their accuracy is a complex task, and there are multiple observational pathways to provide constraints and tune model parameters. Here, we combine several such independent constraints, defining a model-agnostic framework for standardized comparison. We require models to predict the distribution of coronal holes at the photosphere, and neutral line topology at the model's outer boundary. We compare these predictions to extreme-ultraviolet (EUV) observations of coronal hole locations, white-light Carrington maps of the streamer belt, and the magnetic sector structure measured in situ by Parker Solar Probe and 1 au spacecraft. We study these metrics for potential field source surface (PFSS) models as a function of source surface height and magnetogram choice, as well as comparing to the more physical Wang–Sheeley–Arge (WSA) and the Magnetohydrodynamic Algorithm outside a Sphere (MAS) models. We find that simultaneous optimization of PFSS models to all three metrics is not currently possible, implying a trade-off between the quality of representation of coronal holes and streamer belt topology. WSA and MAS results show the additional physics that they include address this by flattening the streamer belt while maintaining coronal hole sizes, with MAS also improving coronal hole representation relative to WSA. We conclude that this framework is highly useful for inter- and intra-model comparisons. Integral to the framework is the standardization of observables required of each model, evaluating different model aspects.

Unified Astronomy Thesaurus concepts: Solar physics (1476); Solar corona (1483); Solar coronal holes (1484); Astronomical models (86)

Supporting material: animations

1. Introduction

The solar corona is the tenuous outer layer of the Sun's atmosphere extending out to a few solar radii, which comprises hot magnetized plasma. This system is of critical importance for understanding how the Sun interacts with its local environment by creating the solar wind, and how it can affect the Earth. While much can be learned about the corona indirectly from remote observations including plasma density and temperature structure (e.g., Warren & Brooks 2009) or magnetic topology (e.g., Boe et al. 2020), such investigations are usually limited to the plane of the sky as viewed from Earth or one of the STEREO spacecraft. With direct in situ measurements within 1–5 $R_\odot$ currently technologically out of reach, a comprehensive synchronic representation of the full 3D structure can only currently be achieved using physics-based modeling.

Accurately reproducing this structure with models is of vital importance for innumerable endeavours in solar and heliophysics. For example, coronal models form the central component of space weather simulation and prediction frameworks (see, e.g., review by Feng 2020); they enable contextualization and prediction of in situ spacecraft data (e.g., Riley et al. 2019; Badman et al. 2020) and ultimately hold the key to fundamental plasma physics problems that are still unsolved, including coronal heating.

The most commonly used and one of the earliest implemented coronal models was the potential field source surface (PFSS) model (Altschuler & Newkirk 1969; Schatten et al. 1969) wherein the corona is assumed to be magnetostatic (i.e., current-free) in an annular domain with an inner boundary of the radial magnetic field derived from direct magnetograph observations of the

Original content from this work may be used under the terms of the Creative Commons Attribution 4.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

photosphere, and the outer boundary assumed to be an equipotential of radius R_{SS}, called the source surface. (Typically this is set to a value around $2.5\,R_\odot$; Hoeksema 1984.) The equipotential surface simulates the outflowing solar wind by forcing field lines intersecting it to be radial and produces *open* field lines along which plasma can escape, without needing detailed plasma modeling.

PFSS models have some obvious drawbacks. While they produce open field typically rooted in observed coronal holes, the highest altitude closed loops are rounded while eclipse observations show that in fact the last closed loops are kinked and spike outwards in a streamer configuration (Altschuler & Newkirk 1969). Further, the height where field lines are observed to become uniformly radial differs not just from the standard (Hoeksema 1984) $2.5\,R_\odot$ value, but also varies from location to location on the Sun (Riley et al. 2006; Boe et al. 2020) and in time with the evolution of the solar cycle (Réville & Brun 2017). PFSS models also only directly predict coronal magnetic structure, and have an indirect connection to solar wind flow speed (Wang & Sheeley 1992) but otherwise pose no constraints on other plasma properties such as density, temperature, or composition. They are also manifestly time independent. As such, numerous more realistic models have been developed, usually trading improved physical content at the price of computational tractability.

PFSS extrapolations can be extended and improved beyond $2.5\ R_\odot$ with a Schatten Current Sheet (SCS; Schatten 1972), as implemented in the Wang–Sheeley–Arge (WSA) model (Arge & Pizzo 2000; Arge et al. 2003, 2004), which produces more kink-shaped streamer topology and a semiempirical solar wind velocity field, which is a function of the field line expansion factor and the distance between traced open-field-line footpoints and coronal hole boundaries.

Magnetohydrodynamic (MHD) simulations of the solar corona represent the state of the art of global coronal modeling. To the extent that MHD represents the true physical processes in the Sun's atmosphere, these models constitute a complete, time-dependent representation of the associated plasma flows including plasma moments, wave normal modes, and instabilities. Variations between MHD models of the corona often stem from how energy is supplied to the corona and hence drives supersonic outflow, and generally come down to the energy equation in the underlying numerical scheme. Numerous model implementations for the Sun's corona exist; examples include the Alfvén Wave Solar Model (AWSoM; Sokolov et al. 2013; van der Holst et al. 2014), the Multi-Scale Fluid-Kinetic Simulation Suite (MS-FLUKSS; Pogorelov et al. 2013), and the Magnetohydrodynamic Algorithm outside a Sphere (MAS; e.g., Mikić & Linker 1996; Mikić et al. 1999), the latter of which is used in the present work.

Although the three categories of models mentioned above comprise those examined in this article, there are numerous other coronal models that are being developed and seek to improve on PFSS models without going as far as full MHD. For example Kruse et al. (2020) has produced a PFSS extension with an elliptical source surface. Rice & Yeates (2021) proposed a magnetofrictional model that is similar to PFSS models except it produces its coronal magnetic field self-consistently with an accelerating radially outwards flow, which leads to much more compelling streamer belt shapes. Nonlinear force-free (NLFF) models (see review by Wiegelmann & Sakurai 2021) are also an improvement on potential fields in that they reduce the severity of the no-current constraint, allowing currents to flow along field lines.

It is clear that there are a plethora of options for coronal modeling, with trade-offs between the physical content of the model, computational time, and accessibility of model software. However, there is an implicit assumption here that *more physics* should produce more accurate coronal models. To support this assumption, it is clearly desirable to compare models to each other and assess for improvements in accuracy.

Almost any publication introducing a new or tuned coronal model includes a qualitative or quantitative comparison to empirical measurements of the Sun (either remote or in situ). For example, Lee et al. (2011), Arden et al. (2014), Badman et al. (2020), and Panasenco et al. (2020) all investigated PFSS source surface height, with the first comparing the extent of coronal holes, the second trying to match the open magnetic flux in the heliosphere, and the latter two trying to match magnetic polarity changes with Parker Solar Probe (PSP) measurements. Similarly, Kruse et al. (2021) used ballistically mapped magnetic polarity to study an elliptical outer boundary condition to a potential model. Even the first PFSS papers investigated different observables where Altschuler & Newkirk (1969) focused on comparison to white-light (WL) eclipse observations while Schatten et al. (1969) investigated interplanetary magnetic field strength. While this single metric approach is usually useful to support an author's assertion of improvement, especially when comparing their model to a prior iteration of the same model, there are a couple of issues. First, such comparisons made with one type of metric (for example prediction of current sheet crossing timing, or visual correspondence of model field lines to a coronagraph snapshot) will test and optimize a single aspect of the models' representation of the corona. As we show in this work (Section 5.1), it is possible to improve one aspect of a model to the detriment of another. Second, this leads to highly nonstandardized comparisons across the literature. For illustration, we may know model X produces a more accurate forward model of an eclipse observation than model Y, and that model Y produces a much more accurate time-series prediction of velocities compared to the in situ measurements than model Z. However this does not allow us to compare model X to model Z.

The goal of this work is to define a model-agnostic framework to evaluate quantitatively the representation of the corona for different coronal models and to use multiple independent comparisons to do so. We study PFSS, WSA, and MAS model results, and for consistency, we study the minimum set of information that is provided directly by all three models, i.e., the coronal magnetic field. For each model, we require a prediction in a standardized format of (1) the location of coronal holes at the coronal base (i.e., the location of open magnetic field in the model), and (2) the location of the coronal neutral line (NL) at the outer boundary of each model that corresponds to the coronal streamer belt and shapes the heliospheric current sheet (HCS). To test these predictions, we identify three independent observations. The first is extreme-ultraviolet observations of the locations of coronal holes, the second is the location of the streamer belt at $5\ R_S$ observed from WL coronagraphs, and last is the timing of crossings of the HCS as measured by heliospheric spacecraft (PSP, "near-Earth" spacecraft via the OMNI data set, and STEREO A). Each of these three types of observations is compared to the relevant model prediction, and a quantitative metric is

developed to compare them. We study time intervals of ±30 days from the date of the first three perihelia of PSP to focus our investigation and take advantage of novel in situ measurements of PSP closer to the Sun than ever before. Our metric scores are probed as a function of the choice of model, input parameters, boundary conditions, and time over our model data set. We present results and inferences for these specific studies (Section 5) as well as more general conclusions about our model framework (Section 6).

Similar issues have recently been discussed and studied in complementary recent work by Wagner et al. (2022) with the implementation goal of optimizing the EUropean Heliospheric FORecasting Information Asset (EUHFORIA; Pomoell & Poedts 2018) model, another example of an MHD model. As in the present work, these authors specify individual independent evaluation metrics between models and observations. In this work, where relevant, we explain the relationship or difference of implementation in our metrics as compared to those of Wagner et al. (2022).

2. Coronal Field Models

As mentioned above, several coronal models are used in this work to define and test our evaluation framework. In this section, we introduce the model types and specific implementations used to then discuss the model-independent format of "observables" that each model is required to produce to facilitate quantitative comparisons.

Models were obtained for three specific time intervals, distributed ± 30 days either side of the first three perihelia of PSP. As mentioned in the introduction, this choice was made to focus our investigation and make use of novel new in situ measurements from PSP. For each model, at least one run per interval was generated, but where possible provided full, daily updates to the magnetogram input boundary conditions spanning these time intervals.

2.1. Potential Field Source Surface (PFSS) Models

PFSS models for this work were generated from two sources:

1. *pfsspy*,[16] an open-source Python implementation (Stansby et al. 2020) with Air force Data Assimilative Flux Transport (ADAPT; Arge et al. 2010, 2011, 2013; Hickmann et al. 2015) magnetograms as their input boundary conditions. ADAPT maps enhance typical magnetogram observations by assimilating direct measurements of magnetic fields on the solar near side into a flux transport model of the whole Sun. ADAPT map data products include 12 individual realizations obtained by running an ensemble of models. For the PFSS model run here, we average these realizations together, subtracting any residual monopole moment induced by the averaging before feeding it into *pfsspy*. ADAPT maps are regularly produced using magnetographs from both Global Oscillation Network Group (GONG; Harvey et al. 1996) and from the Helioseismic and Magnetic Imager (HMI; Scherrer et al. 2012) on board the Solar Dynamics Observatory (SDO; Pesnell et al. 2012). Both are used and their separate models generated with *pfsspy* in this model data set. In this work, these two magnetograms will be referred to as "ADAPT-GONG" and "ADAPT-HMI" respectively.
2. The standard PFSS implementation by the Global Oscillation Network Group (GONG; Harvey et al. 1996) utilizes the zero-corrected GONG magnetogram data product (Hill 2018). This magnetogram will be referred to as "GONGz" in this work.

As well as being tabulated for each day in the studied time intervals, individual model runs were produced for differing values of source surface height (R_{SS}) ranging from 1.5 to 3.0 $R_\odot$, spanning most of the range typically investigated in studies of R_{SS} (e.g., Lee et al. 2011; Arden et al. 2014).

2.2. Wang–Sheeley–Arge (WSA) Model

The Wang–Sheeley–Arge (WSA) model (Arge & Pizzo 2000; Arge et al. 2003, 2004) extends the capabilities of PFSS models with a SCS (Schatten 1972) model extending above the potential layer, providing magnetic coronal structure out to 5 $R_\odot$. WSA additionally generates an empirical velocity field at the model's outer boundary (generated as a function of expansion factor and the distance of magnetic footpoints to coronal holes), although only the magnetic structure is accessed in the present work. As with *pfsspy*, for this project, these models are driven by ADAPT magnetograms. Selection of magnetogram realization varied across different time intervals.

WSA model parameters include a source surface height (R_{SS}) marking the boundary between the PFSS and SCS components of the model, and an outer boundary for the SCS (R_{SCS}). Unless otherwise noted, these values were set to $R_{SS} = 2.5\,R_\odot$ and $R_{SCS} = 5\,R_\odot$.

Full, daily evolved models were provided for our data set for the perihelia 1 and 3 intervals. For perihelion 2, a single model run was investigated using an ADAPT-GONG map from 2021 April 10. As will be discussed in Section 5, encounter 2 was challenging for modelers due to far side active region emergence. This single model run was chosen for encounter 2 as it provided the best representation of PSP in situ magnetic polarity.

2.3. Magnetohydrodynamic Algorithm outside a Sphere (MAS) MHD Model

Lastly, we use results from the Magnetohydrodynamic Algorithm outside a Sphere (MAS) model (Mikic & Linker 1994; Mikić et al. 1999; Riley et al. 2001) for the studied time intervals. As touched on in the introduction, MHD models are significantly more computationally intensive, and as such the daily evolution of the magnetogram input was not available for this project. Instead a single model run is provided for each perihelion date based on the closest Carrington rotation number.

The MAS models extend from the corona out to 30 $R_\odot$. The supersonic solar wind outflow in this case is generated via the PSI thermodynamic energy equation (Lionello et al. 2001, 2009). The magnetograms used to drive them from the inner boundary used HMI data (Mikić et al. 2018; Riley et al. 2019). Again, although the MAS model provides a litany of other physical properties (e.g., density, flow velocity, temperature), only the magnetic structure is extracted for this work.

[16] https://github.com/dstansby/pfsspy

2.4. Derived Observables from Coronal Field Models

A very important step in the framework outlined in this paper is defining a standardized set of "observables" to be produced by each model. This allows for intermodel comparison and portability of this method to future models by ensuring that the comparison between observations and models is made in a controlled way that does not depend on any internal properties of the model used to produce them.

Specifically, we require models to produce the following:

1. A 2D open field map at the photosphere uniformly binned in sine heliographic latitude and heliographic Carrington longitude. Each pixel takes either the value 1 to indicate that a field line traced from that point on the photosphere reaches the model's outer boundary and is magnetically *open* or the value 0 to indicate a field line traced from the point is *closed* and reintersects the photosphere. The left-hand edge of the 2D map is required to be at 0° Carrington longitude.
2. A 2D map of the position of the HCS at the outer boundary of the model binned in the same way as the coronal hole map. Pixels have the value of $+1$ for locations where the field is directed outwards (positive polarity) and -1 where the field is directed inwards (negative polarity). Presented in this way, the HCS can be extracted from the map using contouring software in IDL or Python at the zero contour level, which allows for the possibility that the HCS consists of multiple closed loops. Although such topologies were not encountered in this study due to the deep solar minimum, this allows this framework to be useful at future (or previous) solar maxima, where such a topology may occur.

Thus, we require models to predict coronal structure at both their inner and outer boundaries. Examples of these two observables may be seen in the top panels of Figures 1 and 2 respectively. Note we are only examining magnetic field structure. While coronal plasma of course comprises many other physical quantities (such as density, flow, temperature, and composition), here we opt to restrict to magnetism because this is the minimal physical content common to all realistic coronal models, and so it allows direct comparison between all commonly used models, including those to come in the future.

3. Observations and Measurements

In this section, we describe the corresponding observational data used and processed to facilitate comparison to the models. Data is obtained concurrently with the models, i.e., for the time intervals around each of PSP's first three perihelia.

3.1. EUV Coronal Hole Synoptic Maps

We use SDO Atmospheric Imaging Assembly (AIA; Lemen et al. 2012) observations to prepare our EUV Carrington maps from which coronal holes can be detected. We downloaded 180 s worth of data for the 304 Å, 171 Å, 193 Å, and 211 Å filters each day of the Carrington rotations relevant to PSP encounters 1–3. The data were processed using standard AIA analysis software. We performed a deconvolution of the point-spread function on each image using `deconvolve` from aiapy[17] and produced a median-filtered composite image in each wavelength for each day. The daily images were converted from helioprojective coordinates to Carrington longitude and sine latitude using `wcs_convert_to_coord`,[18] a routine in the Solar Software (SSW) IDL package (Freeland & Handy 1998). Finally, these are combined to produce an EUV Carrington map for the Carrington rotation. The blending overlap is about 12^o in longitude. PSP perihelia 1 and 3 overlapped mostly with Carrington rotations 2210 and 2221 respectively, but encounter 2 occurred near the boundary of Carrington rotations 2215 and 2216. In this case, we prepared maps for both rotations, switching between them at the appropriate time.

Several methods of coronal hole detection have been used by the community dating back to early efforts using hand-drawn maps (Harvey & Recely 2002). Subsequently, a number of independent detection methods have been developed. For example, using local intensity thresholding (Krista & Gallagher 2009), forming high-contrast images from linear combinations of spectral line properties (Malanushenko & Jones 2005), segmenting images by region-growing algorithms (Caplan et al. 2016), and applying intensity thresholding across multitemperature image passbands (Garton et al. 2018). Each of these techniques has its advantages and weaknesses. Automatic detection algorithms, for example, are very powerful for processing large volumes of data.

In this work, we experimented with several methods: simple EUV intensity thresholding, the coronal hole identification via multithermal emission recognition algorithm (CHIMERA; Garton et al. 2018), and the Predictive Science Inc. (PSI) EZSEG image segmentation code (Caplan et al. 2016). Ultimately we chose to use the EZSEG algorithm. For this work, the chosen algorithm needed to be able to accurately produce contours of coronal holes from 2D Carrington maps to allow direct comparison to our chosen observables (see Section 4.1). By visual comparison, the EZSEG algorithm was found to be the best for our circumstances, particularly proving useful at contouring the polar coronal holes that form a large fraction of the global coronal hole area at solar minimum.

The software is available as part of the EUV2CHM Matlab package from the PSI website.[19] We converted the EZSEG algorithm from FORTRAN to IDL to plug-in to our software pipeline. This algorithm uses an initial intensity threshold to acquire coronal hole locations in an EUV image, and then uses an area growing technique to define connected regions. This continues until a second intensity threshold is reached, or the condition for connectivity is not met. The dual-thresholds and connectivity conditions (essentially the number of consecutive pixels) are defined on input. We experimented with the optimal input parameters for each of our 193 Å Carrington maps. As an example, for encounters 1 and 3, thresholds of ∼20% and ∼60%–70% of the peak EUV intensity worked well, with connectivity defined as 5 consecutive pixels. For encounter 2, thresholds of ∼15% and ∼45% worked better. This reflects the presence of several brighter active regions around the solar surface during encounter 2. An example, of our 193 Å Carrington maps and the EZSEG output, is shown in the middle panel of Figure 1, showing how the algorithm performs well at contouring coronal hole regions identified by eye as dark regions in the EUV carrington map. This figure as a whole

[17] https://aiapy.readthedocs.io/en/latest/

[18] https://hesperia.gsfc.nasa.gov/ssw/gen/idl/wcs/conversion/wcs_convert_to_coord.pro

[19] http://www.predsci.com/chd/

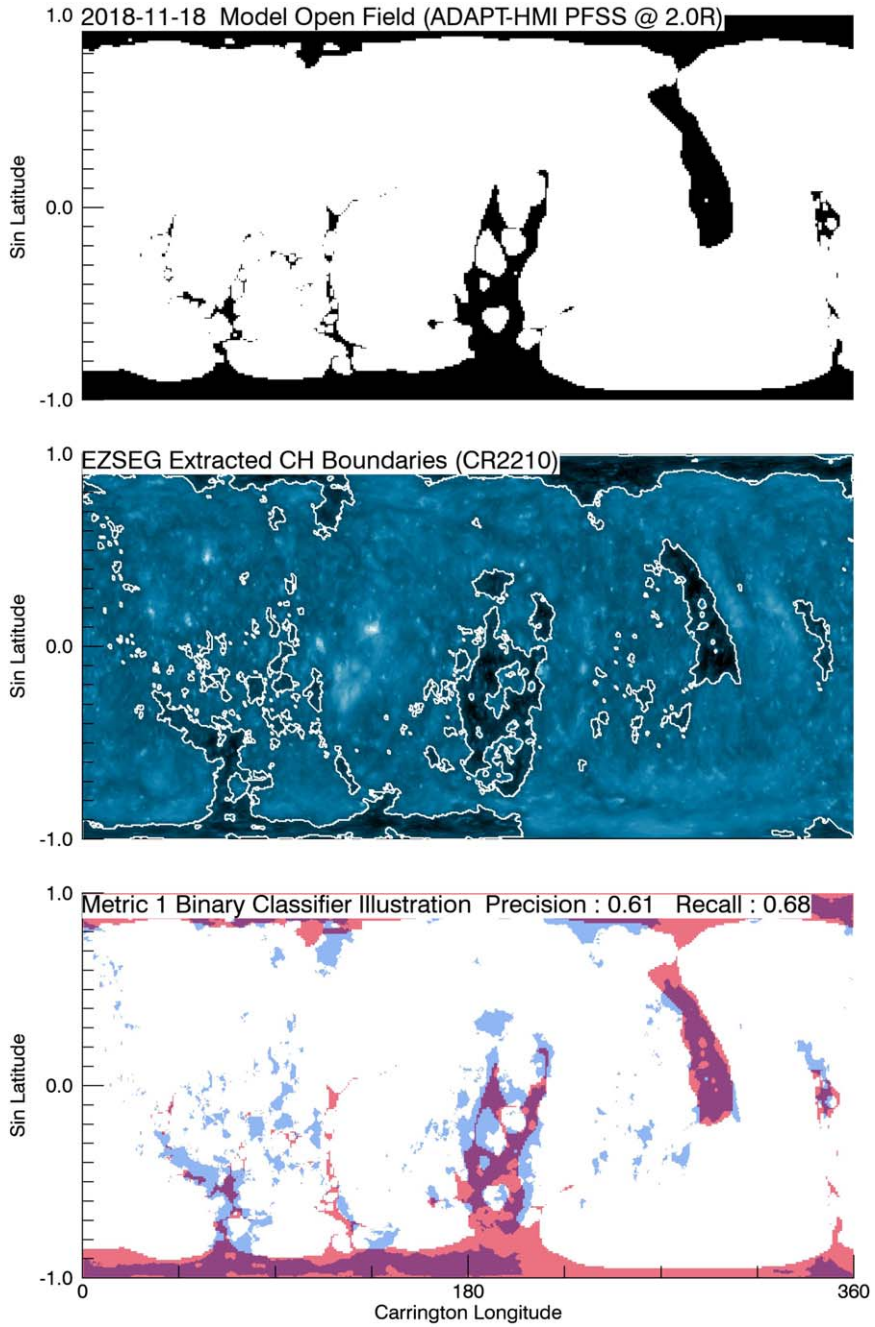


Figure 1. Illustration of construction of metric 1 using the example of an ADAPT-HMI PFSS model from 2021 November 18 during PSP Encounter 1 with a source surface height parameter of 2.0 $R_\odot$. Top panel: mask of coronal hole locations (black shading) predicted by the PFSS mode model, i.e., pixels in the input magnetogram that map to open field lines. Middle panel: coronal hole locations as determined by applying the EZSEG algorithm to AIA/EUVI maps of 193 Å coronal emission, as described in the main text. The background image depicts the EUV Carrington map and the white contours indicate the areas identified by EZSEG. Bottom panel: shading depicting the comparison of the observed and modeled coronal holes and metric construction. Blue shading indicates pixels that are identified as coronal holes by EZSEG, and red shading indicates locations predicted open by the model. Where they overlap, purple shading indicates the pixels are predicted and measured as coronal holes. The main text (Section 4.1) describes how these regions are used to generate the precision, recall, and final "F-score" metrics to quantify the agreement.

and Section 4.1 explain how we make the comparison of our maps with modeled coronal hole areas.

3.2. White-light Synoptic Maps

We exploit continuous WL observations from the Solar and Heliospheric Observatory (SOHO; Domingo et al. 1995) Large Angle and Spectromeric Coronagraph Experiment (LASCO) C2 coronagraph (Brueckner et al. 1995) to produce WL synoptic maps.

The light received by such a detector located at 1 au is the result of multiple contributions along the lines of sight. The most significant contribution is the photospheric light scattered by dust particles, the F-corona. The latter is removed by subtracting background images from raw images in order to reveal the faint amount of photospheric light that is scattered by

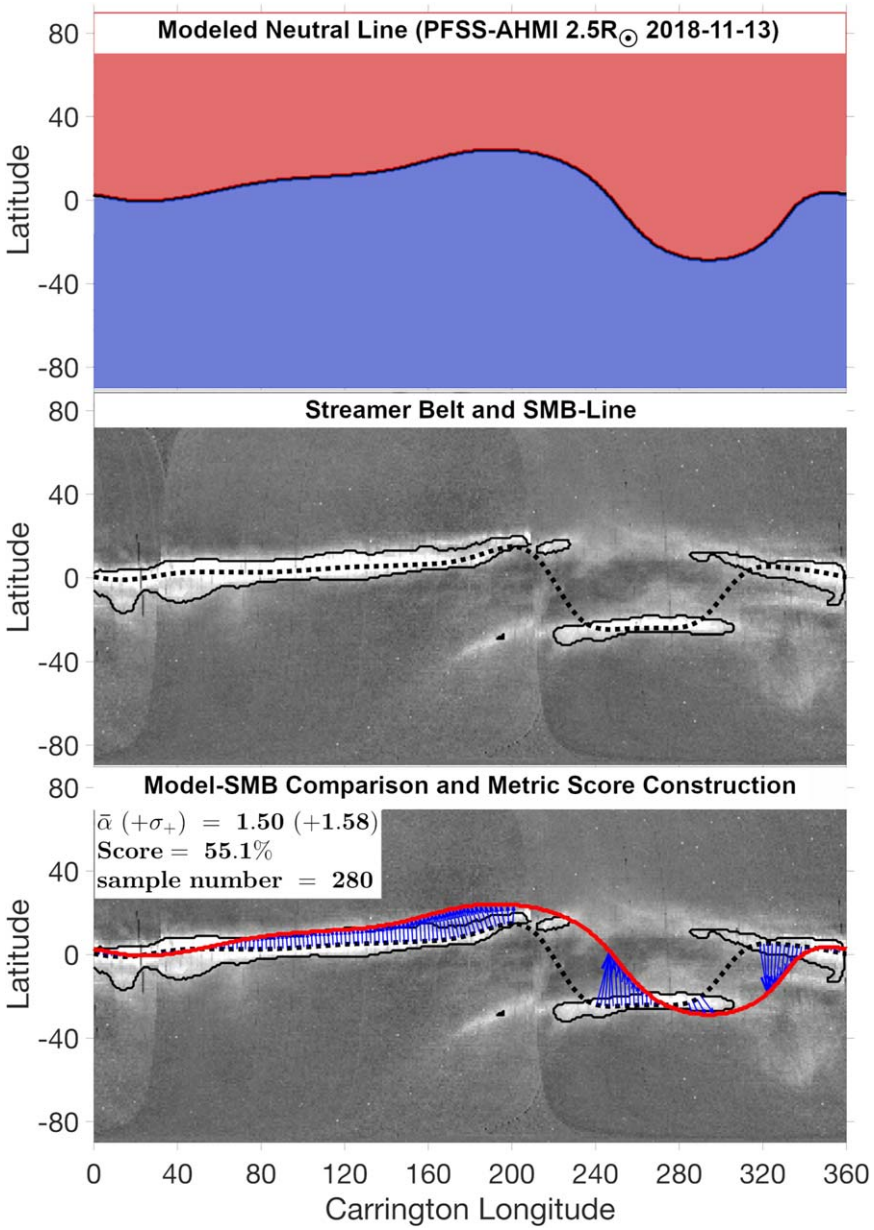


Figure 2. Illustration of the construction of metric 2. Top panel: model input—contour of polarity inversion line as predicted by coronal models. Middle panel: observation input—a WL synoptic map of the streamer belt (Carrington map) constructed at a height of 5 $R_\odot$ from SOHO-LASCO-C2 images taken between 2018 October 30 and 2018 November 13. The streamer belt contour is plotted in dark. The core of the streamer belt identified by the SMB line is traced as a dark dashed line. Bottom panel: comparison between the SMB line and the modeled polarity inversion line (solid red line). The angular deviations between the two lines that are used to compute the WL metric are represented by the blue arrows. Results from the WL metric are given in the legend, of which the WL global score (in percent) is calculated from the mean $\bar{\alpha}$ and mean deviation σ_+.

electrons, the K-corona. Hence this process provides not only access to the overall structure of the corona but also details often observed in eclipse images for instance (Boe et al. 2020; Druckmuller 2008). Monthly minimum averaged images are used as background images to remove the F-corona. They are accessible via the widely used SolarSoft IDL library. The resulting image can now be used to visualize features of interest such as bright coronal rays that are signatures of enhanced density structures. Streamer rays are good proxies to visualize the dense plasma emanating from the tip of helmet streamers, which in turn supplies the heliospheric plasma sheet (HPS). Hence streamer rays are also useful to get an approximate location of the HCS, which is likely located inside the thick HPS.

The second step is to concatenate and convert continuous observations into a synoptic format, which is heliographic latitude versus longitude at a given height in the corona. Such maps, if defined in the Carrington frame rotating with the Sun, are called Carrington maps. They have been significantly used in the past to exploit 1 au observations (see, e.g., Wang & Sheeley 1992; Wang et al. 1998; de Patoul et al. 2015; Pinto & Rouillard 2017; Sasso et al. 2019; Rouillard et al. 2020).

To generate such a 2D map from LASCO-C2 images, the following steps are followed. For 1 au observations and small heliocentric distances, a widely used and reasonable approximation is to project images onto the helioprojective sphere (that is centered at the observer, also known as the plane-of-the-sky). The projected image is then converted from

helioprojective to Carrington coordinates. As the Sun rotates on itself, the LASCO-C2 coronagraph images the plane of the sky above both the east and west limbs, performing a full scan in half of the solar synodic period (13 days). A 3D annular volume of WL intensity is therefore progressively filled up, providing a global overview of the structure of the solar corona at multiple altitudes, similarly to tomographic imagery techniques (Morgan & Cook 2020). Finally, we extract a single slice from this 3D volume at a specific heliocentric distance, getting at last a 2D-synoptic, or Carrington, WL map. As long as new images are available, the Carrington synoptic map is continuously added to and updated, allowing for the production of a near-real-time map. An example of a WL synoptic map is shown in the middle panel of Figure 2, which has been extracted at 5.0 Rs and combines SOHO-LASCO-C2 images from 2018 October 30 to November 13.

Once the WL synoptic map is complete, we perform 2D median filtering in order to reduce *salt and pepper* noise while preserving edges of features of interest. The streamer belt is a thick band of enhanced brightness in each WL synoptic map, which shows the overall shape of the HPS around the Sun. It is not always easily distinctive and sometimes can be confused with pseudostreamers. One needs to adjust the height at which the WL synoptic map is extracted in order to dispel the ambiguity. For instance, the usual larger radial extent of streamers compared to pseudostreamers can help to discriminate between both (Wang et al. 2007). To start, we selected a height of 5.0 Rs, large enough to remove most of the pseudostreamer signatures, while also keeping good image quality (further details can be found in Poirier et al. 2021). The outermost boundary of the streamer belt is identified with a simple brightness threshold. The threshold has been adjusted manually until a coherent shape was obtained (see the dark contour in the middle panel of Figure 2). That provides a rough idea of the shape and thickness of the HPS. For this work, the key feature of interest is the HCS that is likely standing inside the streamer belt and can reasonably be expected to be approximately located where the brightness is maximum within the streamer belt. As such, we trace the locus of maximal brightness at each longitude (see the dark dashed line in the middle panel of Figure 2), hereafter called the Streamer-Maximum-Brightness (SMB) line, via the method described later in Algorithm Box 2. A good correlation has been found between the SMB line and the magnetic sector structure observed in situ at 1 au (see Poirier et al. 2021). Similarities between the brightest regions of the K-corona and the magnetic sector boundary have also been highlighted in past studies (Hansen et al. 1974; Howard & Koomen 1974).

3.3. PSP and 1 au in situ Magnetic Polarity Measurements

The final observation we use in our model comparisons are direct in situ measurements of the magnetic field polarity in the inner heliosphere over the time intervals of interest combined with knowledge of the corresponding spacecraft's position. We use data from the FIELDS instrument (Bale et al. 2016) on the *Parker Solar Probe* spacecraft (Fox et al. 2016) down to 35.7 $R_\odot$, as well as the OMNI "near-Earth" data set[20] and measurements from the IMPACT (Luhmann et al. 2008) instrument on board STEREO AHEAD (Kaiser et al. 2008) at 1 au. The magnetic field polarity is determined by the sign of the radial component of the field (B_R). We consistently choose to determine this by computing the most probable value over 1 hr intervals of higher-cadence data following Badman et al. (2020). This removes small-scale fluctuations and gives a resulting observable that indicates which side of the HCS that specific spacecraft was located at a particular time. One further step is required for comparison with the location of the NL predicted by our coronal models. Since the coronal models predict the NL at a distance of a few solar radii, the spatial information of the spacecraft farther out must be corrected to account for the Parker spiral. We utilize ballistic propagation (Nolte & Roelof 1973; Macneil et al. 2021) whereby a constant solar wind velocity is assumed for a Parker spiral field line intersecting the spacecraft to set the spiral curvature. This velocity is chosen either from in situ measurements of the velocity at the spacecraft or using a compromise value of slow wind of 360 km s^{-1} (typical of low heliographic latitudes). By combining the polarity time series with the spatial location of the spacecraft, we can build up a picture of where warps in the HCS are located. An example of data from PSP/FIELDS during its first solar encounter (2018 October/November) is shown in Figure 3 where, in the middle panel, we see PSP's trajectory (ballistically mapped to the edge of the corona) colorized by the polarity measured as it traveled through each position.

The final derived observation is then a list of time stamps, polarity values, and ballistically corrected heliographic latitudes and longitudes of the spacecraft at the outer boundary of the coronal model (typically 2.0–2.5 $R_\odot$ for PFSS models). Our method of comparison to this observation is described below in Section 4.3.

4. Prediction-observable Comparisons

We next introduce our three metrics for comparing the data and modeled observables introduced in the previous sections.

4.1. Coronal Holes

Our aim here is to provide some measure of how well the coronal magnetic field models are able to predict the true open magnetic field on the Sun during each of the PSP encounters. We compare the model predictions of the locations of open field (black regions, top panel, Figure 1) with the locations and extent of coronal holes detected in our 193 Å Carrington maps (white contours, middle panel, Figure 1). Each of the models outputs a Carrington map of open and closed magnetic field. Similarly, EZSEG (Section 3.1) produces a pixel map of coronal hole detections. We take the EZSEG map of coronal hole detections to be the ground-truth open field data set and apply a binary classifier algorithm to make the model comparison. This overall procedure is summarized in Algorithm Box 1.

The algorithm computes three metrics: the precision (p), recall (r), and f-measure (f). These are familiar concepts in machine learning and statistics and are defined as

$$p = t_p/(t_p + f_p) \quad (1)$$

$$r = t_p/(t_p + f_n) \quad (2)$$

$$f = 2pr/(p + r) \quad (3)$$

where t_p is the number of true positives, f_p is the number of false positives, and f_n is the number of false negatives. Each metric falls in the range of 0–1. Here, a true positive is recorded

[20] omniweb.gsfc.nasa.gov/

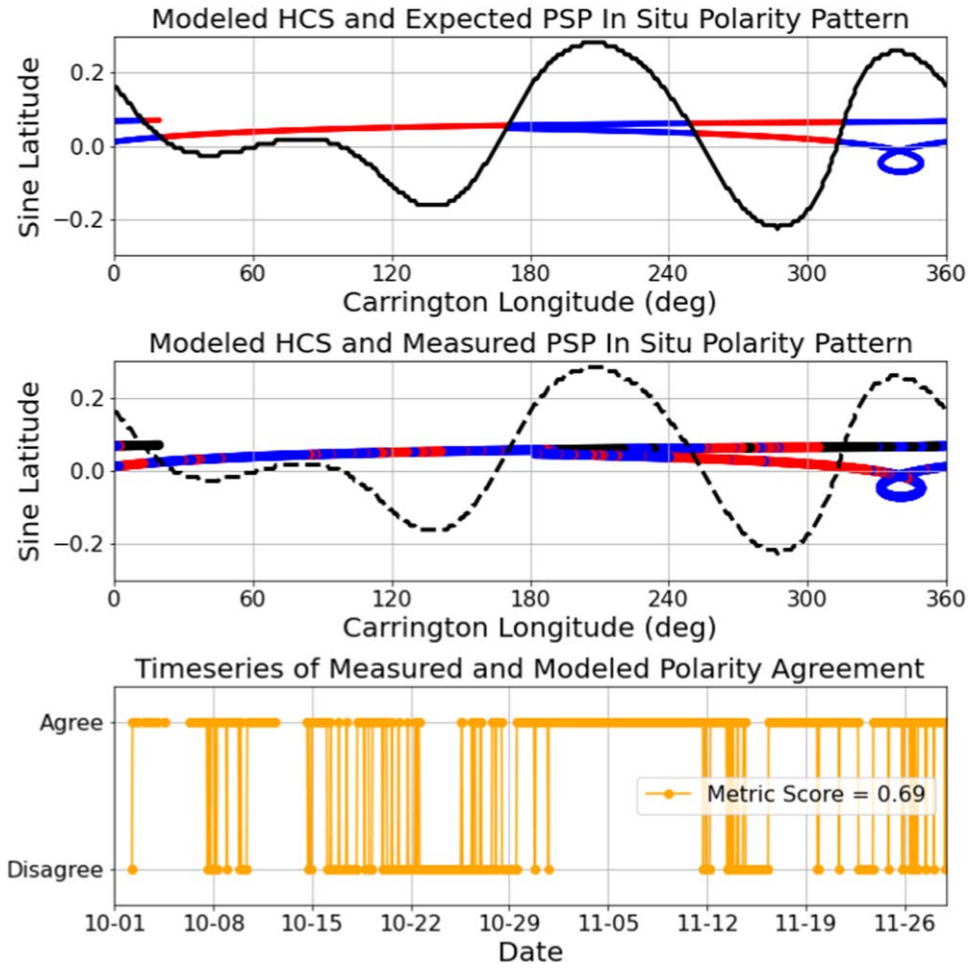


Figure 3. Illustration of construction of metric 3. Top panel: model input—contour of polarity inversion line as predicted by coronal models and resulting predicted polarity that PSP's trajectory is expected to trace out according to the model. Middle panel: observation input—same trajectory of PSP but colored by measured polarity. Dashed curve indicates the same polarity inversion line from the top panel showing the model-data agreement is imperfect. Bottom panel: product of measured and predicted polarity time series. Values are +1 for agreement and −1 for disagreement. Metric is the ratio of agreements to the number of measurements, as described in the main text, Section 4.3.

Algorithm 1. Process to Generate Coronal Hole F-score

1) Prepare (or download from Zenodo, doi:10.5281/zenodo.6342187) a 193 Å EUV Carrington map for time interval of interest.
2) Prepare coronal hole map from model output (0 = Closed, 1 = Open), Map A.
3) Apply EZSEG (Caplan et al. 2016) to 193 Å EUV Carrington map to obtain a binary map of "measured" open and closed pixels, again 0 = Closed, 1 = Open: Map B.
4) Interpolate model coronal hole map to the same resolution as the binary map.
5) Compute true positive rate t_p: number of pixels open in Map A *and* Map B.
6) Compute false-positive rate f_p: number pixels that are open in Map A and closed in Map B.
7) Compute false negative rate f_n: number of pixels that are closed in Map A and open in Map B.
8) Follow Equations (1)–(3) to compute precision, recall, and F-score respectively.

when a pixel marked as a location of open field in the model is found to be open in the ground-truth map (purple or "overlap" pixels in the bottom panel of Figure 1). A false positive is recorded when a pixel marked as a location of open field in the model is found to be closed in the ground-truth map (red pixels in Figure 1). Similarly, a false negative is recorded when a pixel location marked as closed field in the model is found to be open in the ground-truth map (blue pixels in Figure 1). The precision is therefore the fraction of the open field predicted by the model that is actually open according to the detection of coronal holes in the 193 Å Carrington synoptic map, which is the number of purple pixels as a fraction of the number of pixels that are red or purple. Conversely, recall is the fraction of actual open field (according to the coronal hole detection map) that is predicted to be open by the model, or number of purple pixels as a fraction of the number of blue or purple pixels. These two metrics are useful for understanding how the models are performing, but, as can be seen in panel-by-panel comparison in Appendix A.1, they are anticorrelated, so we calculate the f-measure (the harmonic mean of the precision and recall) to classify which models are best. The harmonic mean gives high values if both constituents have high values, but gives low values if *either* constituent is low-value independent of the other.

By now it should be clear that there are some limitations to our comparisons. First, we assume that all the magnetic field within all the detected coronal holes is open. There is, of course, a transition to closed field close to the boundary. Second, we assume that there is no open field outside of coronal holes. Comparisons of PFSS extrapolations with actual solar features indicate that this is not the case—see, e.g., the full Sun intensity and open-field maps of Brooks et al. (2015) that suggest a large contribution to the slow solar wind from active region outflows. Although the time intervals studied in this work occur during deep solar minimum, we do have one interval when there was significant solar activity.

As will be shown later, metric scores typically did not exceed 60%. Figure 1 illustrates some of the reasons for this. For the example shown, the model appears to overexpand the polar coronal holes and underexpand the mid-latitude ones. Since the polar coronal holes are the largest open regions by solid angle, they became an important point of comparison for the models and are what led us to prefer EZSEG as the algorithm for detection (recall the discussion in Section 3.1). As we can see in the middle panel of Figure 1 (white contours), EZSEG does a good job detecting the polar coronal holes while other algorithms were less effective. Conversely, there are quite a few small observed coronal holes that appear dubious. These, however, constitute a small fraction of the total area of detections so are a minor contributor to our results. An example showing the f-measure comparisons for all the models that we investigated for encounter 1 is included in the Appendix (Figure 8) and linked to in the full encounter animations in the online version of the article.

Wagner et al. (2022) also make comparisons between EUV images and open field models (specifically from the EUHFORIA model). In their work, they use a simple threshold condition after Krista & Gallagher (2009) to obtain observed coronal holes, and use a similar set of binary classification metrics (see Wagner et al. 2022 for details). This work addresses the difficulty with polar coronal holes by truncating their binary classification to between $\pm 60^{\circ}$ latitude, and therefore score the model based on mid- and low-latitude coronal holes only. In the present work, our choice is to use a more complex coronal hole extraction software (EZSEG) and therefore a slightly more convoluted analysis pipeline with the advantage of having a higher degree of confidence in our extracted polar coronal holes. Indeed, we should note that the polar coronal holes, in general, are poorly observed both in EUV and the magnetograms driving the models. However, their equatorward extent is usually well observed, at least in one hemisphere. Accepting this uncertainty, we make use of sine latitude (or cylindrical-equal-area, CEA) projections in our maps, which limit the area occupied by the polar coronal holes, effectively assigning a lower weighting to those pixels by preserving pixel area at higher latitudes. Future work may seek to improve Equations (1)–(3) with an explicit weighting for different types of coronal holes, and even further afield, the direct solar polar observations may negate such considerations altogether. However, in the present work, we proceed with this straightforward classification scheme and argue that due to the deep solar-minimum conditions that our studied time intervals occur during, it is important to assess the polar coronal holes in our metrics.

4.2. Streamer Belt

As discussed in Section 3.2, we built and exploited WL synoptic maps to get an approximate shape of the HCS, or magnetic sector structure, via what we called the SMB line. As a consequence, the SMB line can then be easily compared to any global coronal model. The steps we follow are summarized in Algorithm Box 2. However with the SMB being extracted at a specific height, some precautions must be taken prior to any comparison being made.

In this work, we performed PFSS modeling with a source surface height up to 3.0 $R_{\odot}$. Above this height, the magnetic field is assumed to be purely radial, and hence the line of polarity inversion or NL remains the same in a latitude versus longitude format. Assuming that the streamer topology does not change significantly between the PFSS source surface height and 5 $R_{\odot}$, the SMB can be compared with the PFSS NL. The overall shape of the streamer belt is not expected to vary significantly, but a slight narrowing of the streamer belt could be observed still. The outer boundary of the WSA runs is located at 5 $R_{\odot}$, at the exact same height at which the SMB line is extracted, allowing a direct comparison. The MAS MHD simulations from PSI extend much further, and the NL is extracted at 30 $R_{\odot}$. In principle, this means the HCS could be shifted up to 8° due to Parker spiral curvature,

Algorithm 2. Process to Generate White Light Streamer Belt Score (Poirier et al. 2021)

1) Prepare (or download from Zenodo, doi:10.5281/zenodo.6342187) WL Carrington map for time interval of interest.
1a)(recommended) Clean up the WL map with a 2D Gaussian filtering (e.g., using the `medfilt2()` function in MATLAB).
2) Prepare neutral line from model output (2D map of polarity at model outer boundary, pixel values $= \pm 1$).
3) Extract (or download) the SMB line. If downloaded then jump directly to step 4.
3a) Find longitudes and latitudes of max brightness points in the WL Carrington map.
3b)(optional) Smooth out the SMB line with a 1D Gaussian filtering (e.g., using the `smoothdata()` function in MATLAB).
4) Measure (or download) the streamer belt (half) width (T_i) (provided in the SMB .ascii files in the Zenodo repository). One could also simply assume a constant streamer belt width and jump to step 5. If you choose the download option or want to perform your own measurements, please carefully read the footnote.[a] Follow steps 4a and 4b if you decide to perform your own measurements or else jump to step 5.
4a) Draw streamer envelope with a simple contour threshold.
4b) Measure the angular distances (T_i) between each point of the SMB line and the extracted contour of the streamer envelope.
5) Measure for each grid point along the SMB line, the angular distance (α_i) to the model neutral line extracted at step 2. Make sure to make these measurements along the same directions as in step 4b.
6) Normalize each angular distance (α_i) computed in step 5 by its associated streamer (half) width (T_i) measured in step 4.
7) Apply Equation (4) to obtain the streamer belt metric score $C(\bar{\alpha}, \sigma_+)$.

Notes.
[a] In this work and in the SMB files provided online (Zenodo doi:10.5281/zenodo.6342187), all angular distances are measured in the direction perpendicular to the SMB line, which requires to compute normal vectors beforehand. For simplicity one could choose to measure these angular distances over latitude (i.e., along the vertical axis) or just retain the closest distance to the streamer belt contour. Two streamer widths are given in the SMB files, for the northern and southern portion of the streamer belt, of which only one is used depending on the relative position of the modeled neutral line with the SMB line. One could also simplify the procedure by computing an average of the northern and southern (half) widths.

but in practice this is a small correction compared to the dispersion of NL topology among our data set of model runs.

In order to compare in a statistical manner the WL observations and global coronal models, we need to define a metric that is accurate and robust. To compare the SMB with the modeled NL, various techniques can be tested. We decided to measure the angular distances between the two lines. These angular distances α_i are then normalized by the local streamer (half) thickness T_i, which is obtained from the streamer belt contour identified beforehand. The reason for choosing this normalization is twofold: the streamer thickness is a physical measure that helps to interpret the final metric; and also the final metric is therefore not affected by the streamer thickness variability over space and time. In particular, this normalization leads to the useful intuition that scores above 50% corresponding to the model NL lying on average within the bright WL emission of the streamer belt. The final step consists of calculating an averaged score over the full Carrington longitude range. The global score is built so that a model having a good overall match with WL coronagraphic observations would still be penalized in the case of local and large deviations from the SMB line. Further details on the derivation of this global score can be found in Poirier et al. (2021) where the following formula has been retained:

$$C(\bar{\alpha}, \sigma_+) = g(\bar{\alpha}) \times [1 - p(\sigma_+)]$$
$$g(\bar{\alpha}) = \frac{1}{1 + (99)^{\bar{\alpha}/2-1}}$$
$$p(\sigma_+) = \frac{1}{1 + \left(\frac{1}{99}\right)^{\sigma_+/2-1}} \quad (4)$$

where $\bar{\alpha} = \frac{1}{n}\sum_{i=1}^{n}|\alpha_i|/T_i$ is the mean over longitude of the normalized angular distances to the SMB line, and $\sigma_+ = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(|\alpha_i|/T_i - \bar{\alpha})^2}$ represents the mean deviation. Note that a looser version of the formula has been adopted in this work to better cover the dispersion among all model runs.

An example is illustrated in Figure 2 where the WL metric has been applied to a PFSS model based on the ADAPT/HMI map of 2018 November 13. For this particular case, the polarity inversion line remains mostly at the edge of the streamer belt that is reflected with a midrange WL score of 55%. In regions where the HPS is significantly warped, the WL signature is too faint to allow a precise detection of the shape of the streamer. This happens when the line of sight of an imager, or coronagraph, gets significantly inclined with respect to the HPS such that the light scattered by electrons is integrated over a much shorter path. Hence, the WL metric is configured to not evaluate the model in these specific regions, as shown by the absence of blue arrows in Figure 2. This constrains the effectiveness of this metric to simple solar configurations where the HCS shape is not too complex. As a consequence, the reliability of this metric is uncertain for highly structured coronal configurations typically observed during periods of high solar activity. Future improvements that may increase the reliability of this metric for more complex configurations are discussed in further detail in Poirier et al. (2021). The conversion in an all-in-one metric that evaluates both the pseudostreamers and regular streamer locations should alleviate the uncertainty in using the SMB line alone for complex coronal configurations. WL synoptic maps produced by merging multiviewpoint coronagraphic observations (see, e.g., Sasso et al. 2019) should also be more appropriate to catch the dynamic reconfigurations of the solar corona arising with higher solar activity.

Algorithm 3. Process to Generate In Situ Polarity Score

1) Prepare (or download from Zenodo, doi:10.5281/zenodo.6342187) measured in situ polarity time series. If preparing yourself, recommended to follow processing routine here where for each hour of data, generate histogram of B_R, and take polarity of most probable value as the polarity "measurement" for that time.
2) Produce model-predicted polarity time series by following these steps:
2a) Use spice kernels to generate spacecraft position in the heliographic frame at an hourly cadence over the time interval of interest.
2b) Ballistically propagate this position inwards to 2.5 $R_\odot$ using the equation for the Parker spiral (Equation (1) of Badman et al. 2020).
2c) Sample the modeled 2D map of the neutral line using this projected orbit (via interpolation or nearest neighbor) to generate a time series of predicted polarity.
3) Combine modeled and measured time series via Equation (5) to obtain the in situ polarity metric score.

Wagner et al. (2022) also define metrics to compare WL observations to model results, focusing on snapshots of coronagraph observations and projecting their model results into the relevant plane of the sky. Their method requires a user to "point and click" to associate features in model and data and measure, for example, the angular difference between a modeled and observed streamer structure. This method enables very detailed comparisons at different coronal altitudes, albeit at the cost of requiring human intervention, and also only focuses on one plane cut through the model at a given time. Our chosen method after Poirier et al. (2021) is automated and evaluates agreement at all longitudes, but at the cost of not tracking the variation with coronal altitude.

4.3. Magnetic Sector Structure

From our in situ observations described in Section 3.3, we have a measured time stamp, polarity, and a ballistically mapped spacecraft location at the outer boundary of a given coronal model. To compare this observation to the model's prediction of the location of the HCS, we produce a predicted time series of magnetic polarity. For each time stamp, we take the model output closest in time to that time stamp and sample the polarity at the outer boundary of the model at the location of the associated ballistically mapped footpoint location. We then have, for each time stamp, i, a measured (m), and predicted (p) value of the polarity. While each individual measurement-prediction comparison gives very little information, we can multiply the whole time series together, and divide the total number of agreements (value of +1) by the total number of measurements to give a single value representing the fraction of measurements and predictions that agree, giving a single number between 0 and 1 evaluating the model and in situ measurement agreement for a given encounter. This operation may be mathematically expressed as with the Heaviside step function $\Theta(x)$ as follows:

$$\sigma(\mathrm{sign}(B_R)) = \frac{1}{N}\sum_{i=1}^{N}\Theta(B_R^{m,i}B_R^{p,i}). \quad (5)$$

We note that a value of this metric of 0.5 would indicate as many agreements as disagreements in the time series and thus is approximately equivalent to random chance. Thus, values above 0.5 are required for this metric to convey *good* agreement.

The construction of this metric is illustrated in Figure 3 in a similar three-panel format as the previous two metrics. In the top panel, a modeled HCS is depicted as a solid black curve, and the trajectory of PSP projected onto the model's outerboundary is superimposed. The color of the trajectory (red or blue) shows the model's prediction of the in situ polarity measurement according to which side, of the HCS, the PSP is located. The equivalent data is shown in the middle panel where the same trajectory is colorized by the measurements for the relevant time interval. For reference, the model HCS is again included as a dashed curve to show that the measurement and model do not agree perfectly. Finally, the construction of the metric is illustrated in the bottom panel where a time series of predicted and observed in situ measurements are combined to produce a time series of agreement or disagreement. The score shown in the metric is produced from this data according to Equation (5). This overall procedure is also summarized in Algorithm Box 3.

5. Results

Having introduced our models, derived observables, sources of observation, and the construction of each metric in the prior sections, we proceed to compute metric scores and explore the results. We study the time intervals for ±30 days for each of PSP's first three encounters, and for each encounter, we study metric scores for PFSS models with different source surface heights ranging from $1.5\,R_\odot$ to $3.0\,R_\odot$, and with different magnetogram inputs (ADAPT-GONG, ADAPT-HMI, and GONGz), and for WSA and MAS models for the same time intervals. Given the large number of parameters to explore (time interval, magnetogram, source surface height, model physical content), we divide the presentation of the results into a series of specific studies. A more exhaustive exposition of the scores as plots and movies is shown in the Appendix.

5.1. PFSS Models: Source Surface Height Study

We begin by studying the metric scores for the PFSS models. In Figure 4, the metric scores are depicted for each encounter (column by column) as a function of source surface height. Results using different input magnetograms are shown in different colors (blue for ADAPT-GONG, red for ADAPT-HMI, and black for GONGz). For metrics 1 and 2, the score shown is the average over that encounter, and the error bar is the standard deviation. For metric 3, there is only one value as the metric is integrated over the encounter, and there is therefore no equivalent error bar.

The coronal hole metric is shown in the top row of the figure. All models depict a lower score for the higher source surface heights and have the same qualitative variation for each encounter. The two ADAPT magnetograms (blue and red) follow each other very closely, with the GONGz diverging in qualitative behavior: The ADAPT magnetograms consistently show a maximum coronal hole F-score for a source surface height of $2.0\,R_\odot$ with a small reduction at either lower or higher values. GONGz magnetograms result in a maximum F-score for the lowest source surface height ($1.5\,R_\odot$) and a monotonic decrease for higher values and gives significantly lower values for 2.5 and $3.0\,R_\odot$ source surface heights. The actual average scores vary with a maximum of 0.6 and a minimum for GONGz of ∼0.3 and for the ADAPT maps around 0.55. Error bars indicate the variation of the metric over time within each encounter is small (<10%), which will be verified later in Section 5.3. The bifurcation between the coronal hole representation of the ADAPT maps compared to GONGz, independent of which time interval is examined, suggests the surface flux transport treatment of the ADAPT maps has a stronger impact than variation between GONG or HMI data. Specifically, the ADAPT scores optimize around $2.0\,R_\odot$ while, for GONGz, the lower the source surface height the better. As discussed further in Section 6, this appears to relate to the extent of the polar coronal holes in the model and therefore may relate to the treatment of polar flux. The worsening scores with increasing source surface height are driven by the smaller extent of coronal holes predicted by PFSS as the outer boundary of the model is raised (e.g., Lee et al. 2011). The maximum F-score derived is around 60%, which is far from perfect. As a reminder, the F-score is the harmonic mean of the precision and recall, which have the physical meanings respectively of *fraction of measured open field captured by the model* and *fraction of modeled open field that is actually open*.

Results for the streamer belt metric are depicted in the middle row of Figure 4. Conversely to the coronal hole metric, the PFSS representation of the streamer belt according to our metric *improves* monotonically with source surface height. This is true for each encounter, and independent of the magnetogram driving the model. The scores vary from near zero for the lowest source surface heights ($1.5\,R_\odot$) to greater than 80% for the highest values. The improvement with source surface height starts to flatten above $2.5\,R_\odot$. A dashed curve at a metric score of 50% indicates the score at which the majority of the modeled HCS remains within streamer belt latitudes, as defined by the intensity threshold discussed in Section 4.2. All PFSS models exceed this score for $R_{SS} \geqslant 2.5\,R_\odot$, and many exceed it by $2.0\,R_\odot$. There is variation in the ordering of scores according to the magnetogram from encounter to encounter: For encounter 1, all model scores are nearly on top of each other with ADAPT-GONG being slightly better at $R_{SS} = 2.0\,R_\odot$. For encounter 2, there is significant divergence, with GONGz doing the best followed by ADAPT-GONG and ADAPT-HMI. For encounter 3, GONGz has the lowest score, with ADAPT-HMI slightly better and ADAPT-GONG the best. The monotonic increase in score with source surface height is driven by the extent to which the modeled HCS is warped: At lower values of R_{SS}, PFSS-derived HCS curves are very warped and span large ranges in latitude, while at higher values the HCS flattens out (tending to a flat line at infinite R_{SS}). Clearly, the observed latitudinal extent of the streamer belt is smaller than the HCS variation produced by PFSS models for lower source surface heights.

Finally, the bottom row shows the in situ polarity scores. In addition to the difference between different models, the results are shown for in situ data taken by PSP from ∼37.5–50 $R_\odot$ (circles), and at 1 au by OMNI (diamonds) and STEREO A (squares). A dashed curve at a score of 0.5 indicates the threshold above which the metric score is better than random chance; the majority of scores do exceed this value. In general, the scores are relatively flat with respect to source surface height. Interestingly, although they are nominally testing the same model output as the streamer belt metric (model HCS), the profiles are more similar to the coronal hole metric. The highest metric scores are typically largest for 2.0 or $2.5\,R_\odot$ with a small drop-off for the low and high source surface height

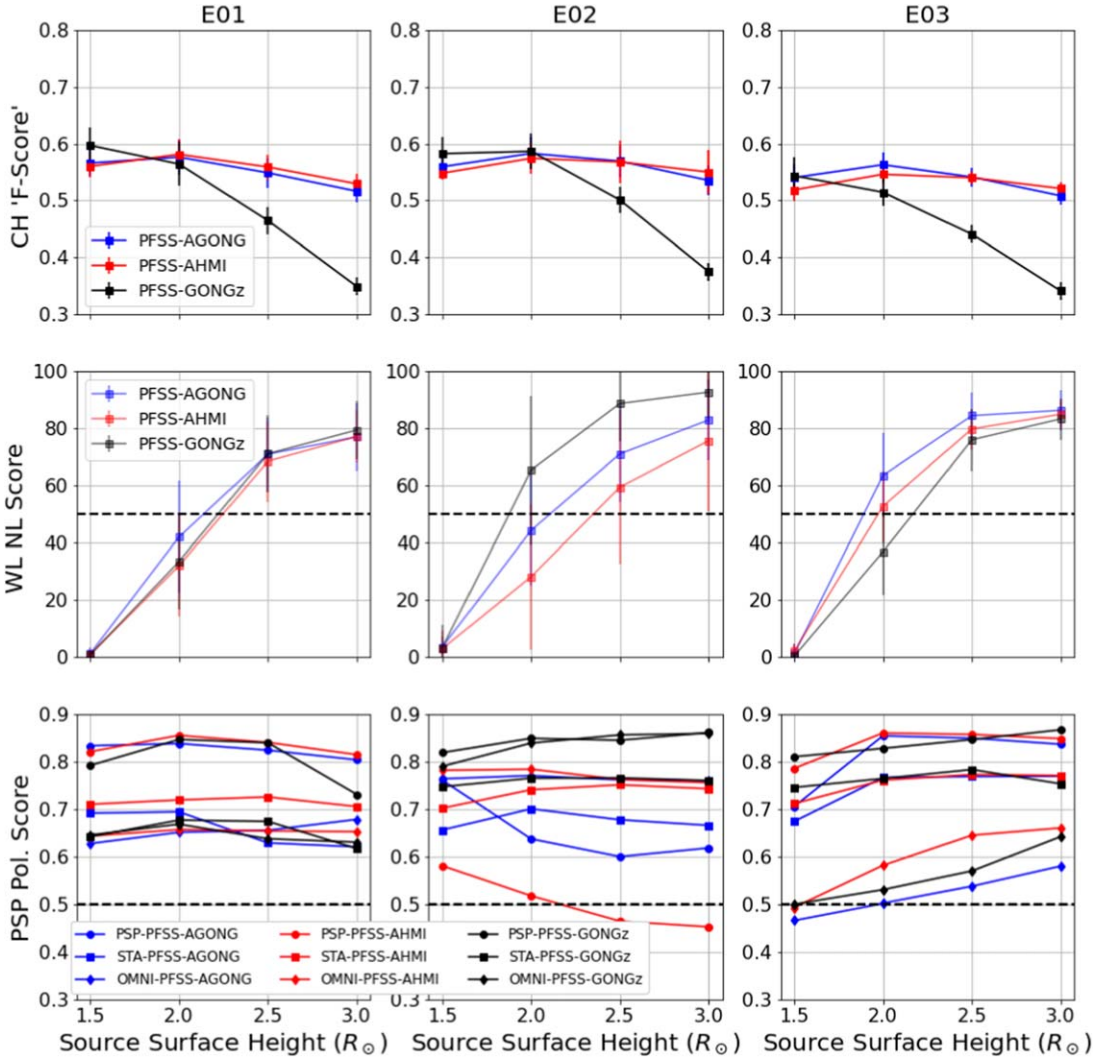


Figure 4. Summary of metric scores for PFSS models as a function of source surface height. Left, middle, and right columns show PSP Encounters 1, 2, and 3 respectively. Top, middle, and bottom rows show metric 1 (coronal holes—Section 4.1), metric 2 (streamer belt topology—Section 4.2), and metric 3 (in situ magnetic polarity—Section 4.3). Each panel shows the respective metric score (averaged over the corresponding PSP encounter for metrics 1 and 2) as a function of source surface height of the PFSS model. Different colors indicate results utilizing different magnetograms (blue, ADAPT-GONG; red, ADAPT-HMI; black, GONGz). The bottom row showing metric 3 has in situ polarity scores for PSP/FIELDS, STEREO A/MAG, and OMNI data shown by different symbols (circles, squares, and diamonds respectively). All metrics are oriented such that a higher metric value indicates a *better* score. The middle row depicting metric 2 has a dashed horizontal line marking the 50% score threshold, which indicates that the modeled HCS lies primarily within the observed extent of the streamer belt. The bottom row has a similar horizontal dashed line indicating the 0.5 score threshold corresponding to the result that randomly generated measured polarities would achieve. Metric 1 and 2 scores have error bars indicating the standard deviation of the score throughout the encounter.

extremes. In terms of comparison between spacecraft, for encounters 1 and 3, PSP gave the highest scores for all source surface heights independent of a magnetogram. For encounter 2, PSP gave both the lowest scores (ADAPT-HMI near 0.5) and the best (GONGz near 0.75). Other than for encounter 1, it is not clear that PSP's in situ measurements close to the Sun are better modeled than in situ measurements at 1 au. For encounter 3, PSP and STA performed identically (with high scores) while OMNI data gave significantly worse scores near the random chance threshold. The variation between the spacecraft metric scores and different magnetogram scores for the same spacecraft is not consistent. For encounters 1 and 3, the scores for a given spacecraft are tightly clustered together for different magnetograms while for encounter 2 there can be wide variation in a given spacecraft score for different magnetograms.

Taken together, there is a clear tension between our different metrics. The higher the source surface height, the better agreement is produced between the modeled HCS and the observed streamer belt location, while at the same time the coronal hole extent produced by the model decreases and gives worsening coronal hole scoring. The higher source surface height also gives slightly worsening in situ scores, implying that while the amplitude of warping of the HCS is more consistent with streamer belt observations, the locations where those warps intersect the spacecraft orbits are less well represented. These results demonstrate that there is no global optimum source surface height that gives a best representation of all aspects of the magnetic structure of the corona.

To investigate if this tension can be eased by the physical insight introduced in more complex coronal models, we compare PFSS results to those of other models in the next section.

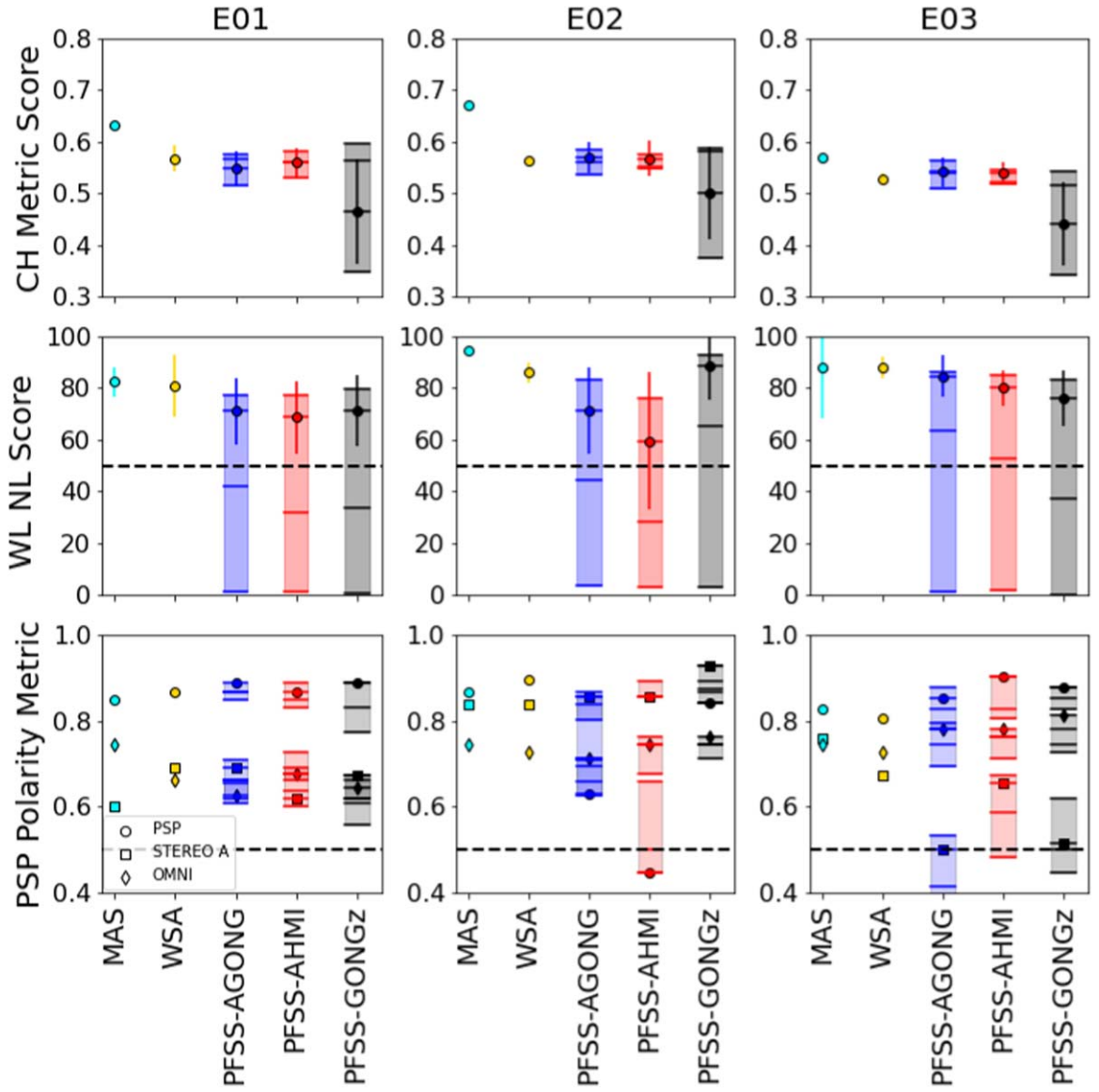


Figure 5. Intermodel comparison of metric scores. Columns and rows arranged as in Figure 4 separating out results for each metric and PSP encounter. In each panel, results are shown for the MAS model (cyan), WSA model (yellow), and PFSS models driven by ADAPT-GONG (blue), ADAPT-HMI (red), and GONGz (black) magnetograms. Mean and standard deviations in metric scores across each encounter are indicated by the plot markers and error bars. For the PFSS models, the plot marker and error bar indicate results for a source surface height of 2.5 $R_\odot$; shading and horizontal bars indicate the range of metric scores with variation in source surface height. For metric 3 (bottom row), circle, square, and diamond markers show different results for PSP, STEREO A, and OMNI in situ measurements respectively. Dashed horizontal lines in rows 2 and 3 are as in Figure 4.

5.2. Intermodel Comparison

Metric scores for the MAS MHD model, WSA model, and the PFSS results discussed in the previous section are compared in Figure 5. The panels are laid out as in Figure 4 (columns and rows differentiating encounter number and metric respectively). In each panel, a metric score for each model is shown, with the main new information coming from the scores for the MAS model in cyan and WSA models in yellow. For the PFSS results, the main marker and error bar show the results for a source surface height for the standard value of 2.5 $R_\odot$, but the variation within source surface height is captured by shaded regions indicating the range of variation, and horizontal bars indicating the distribution of scores that generate that range. The blue, red, and black colors follow the scheme from Figure 4 in differentiating ADAPT-GONG, ADAPT-HMI, and GONGz magnetograms respectively.

The top row again shows the coronal hole metric scores. For each encounter, the MAS model produces higher scores than the best achievable score for PFSS (although only slightly better for encounter 3). There is no error bar shown for the MAS score because there is only 1 coronal hole map generated due to computational tractability. As a reminder, the MAS model is driven by magnetograms from the Carrington rotation of the perihelion in each encounter (CR2210, CR2215, and CR2221). WSA on the other hand does not show significant improvement over the PFSS models in terms of coronal hole representation. For encounter 1, it is slightly better than the 2.5 $R_\odot$ ADAPT PFSS results but less than the best score generated by lowering the source surface height. For the other encounters, it is slightly lower scoring than 2.5 $R_\odot$ ADAPT PFSS results. Note that the WSA model is driven by ADAPT-GONG magnetograms.

Metric 2 as depicted in the middle row of Figure 5 shows that typically the MAS and WSA models produce better agreement between the model HCS and the observed extent of the streamer belt than the best representation achieved by PFSS results (recall these best scores were obtained by the highest source surface height explored, $3.0\,R_\odot$), with the exception of the encounter 2 PFSS-GONGz results that produced comparable scores to the other models for 2.5–3.0 $R_\odot$ source surface height. The PFSS shaded bars in this row demonstrate the stark decline in PFSS scores with lowering source surface height, and we again see heights of $2.0\,R_\odot$ or below tend to produce scores less than 0.5 (dashed horizontal line), indicating the model HCS is reaching higher latitudes than expected from streamer belt observations. In all cases, the MAS score is the same or higher than the WSA score.

Metric 3 finally is compared between models in the bottom row of Figure 5. As previously mentioned, the circles, squares and diamonds indicate scores for PSP, STEREO A, and OMNI in situ measurements respectively. For encounters 1 and 3, for each model, the best score is for PSP, and the scores worsen for OMNI and STEREO A. The ordering of the spacecraft scores is consistent between metrics (PSP $>$ STEREO A $>$ OMNI for E01 and PSP $>$ OMNI $\gg$ STEREO A for E03). Comparing scores for a single spacecraft between models, for E01 and E03, MAS and WSA models give comparable or slightly worse scores compared to PFSS. The exception is the STEREO A scores for E03 that are significantly better in the WSA and MAS models as compared to the PFSS scores that are close to the 0.5 random chance threshold.

Staying with the in situ polarity metric, the result for encounter 2 is significantly different: the ordering of the spacecraft scores is reversed between the MAS and WSA models compared to PFSS models; PSP continues to give the best score for MAS and WSA, but is the worst for the ADAPT PFSS models (middle scoring for GONGz). We also note the variation in the PFSS scores as a function of source surface is often comparable to the difference between different spacecraft at the same source surface height, indicating that the location in the heliosphere at which the in situ data is taken can be as important as the specifics of the model. We note that, since the boundary condition for the WSA model in encounter 2 is a single magnetogram chosen specifically to match PSP polarity measurements, it is not surprising that this specific value of metric 3 is one of the highest observed.

Lastly, bringing the metrics together, we make some observations about the differences between the models overall. We see the following:

1. Despite only utilizing a single magnetogram compared to the time series of magnetograms for the PFSS Models, the MAS model's use of MHD physics leads to improvements of both the coronal hole metric and the streamer belt representation. We note that the quiescent solar-minimum conditions may be responsible for the applicability of the MHD result from a single magnetogram over a long time interval. It does not significantly improve (or may even worsen) predictions of HCS crossing times (which drives the score of metric 3), except for the second PSP encounter specifically for PSP in situ measurements.
2. The WSA model improves the score of metric 2 (streamer belt shape) compared to raw PFSS models, but it does not lead to a noticeable improvement in the coronal hole metric score. It also does not significantly improve scores for the in situ polarity metric, except for encounter 2 where this was specifically tuned for. As discussed in Section 6, this is an unsurprising result based on the WSA physics, which only adds information above the source surface, and therefore forms a good sanity check of our metrics.

Next, we close our exposition of the metric-score results by looking at the details of the metric-score time series for each encounter, from which the averages and error bars in the previous plots were generated.

5.3. Metric-score Time Series

As a reminder, for each encounter for metric 1 and 2, the model output (coronal hole map and NL/HCS map) for PFSS and WSA models was evolved daily by using a new magnetogram (excepting E02 for WSA). The observed WL Carrington map was also updated daily as the longitudes of the streamer belt visible from Earth updated. This means that, for metrics 1 and 2, we have a time series of metric scores for the 60 days comprising each encounter for WSA and PFSS models. For metric 3, the score is the output of integrating the full 60 day "measured" and "modeled" polarity time series together and therefore is not itself a time series.

5.3.1. Coronal Hole and Streamer Belt Metric Time Series

In Figure 6, we depict the time series for metrics 1 and 2. The color scheme follows the preceding figures. The top row shows how the agreement between the modeled coronal holes and the EZSEG output of the EUV Carrington map for the relevant rotation evolved each day through each encounter. The bottom row shows the same for the agreement between the modeled NL and the observed streamer belt Carrington map. For the PFSS models, the semitransparent shading in the relevant color shows the range of metric values explored by varying the source surface height for each time stamp, while the solid lines show the scores for the canonical source surface height of $2.5\,R_\odot$. The upper bound of these shaded regions therefore shows the time series of the optimized source surface height for each mode. Following these time series, we can see how the coronal hole representation and streamer belt representation of each model evolved with each encounter.

Examining the coronal hole representation (top row), we see for encounter 1 (top-left panel) of the ADAPT-driven PFSS models and the WSA model (also driven by ADAPT magnetograms) follow each other closely, showing a periodic variation peaking twice, with an amplitude of approximately 0.1 corresponding to about 10% variation in the score. We note that the "observation" EUV Carrington map is static in each time interval, utilizing data combined over the course of a Carrington rotation; therefore all evolution shown derives from the evolution of the PFSS model driven by updating its magnetogram boundary condition each day. The PFSS-GONGz score (the only daily updated model driven by a non-ADAPT magnetogram) is quite different for the same source surface height ($2.5\,R_\odot$), being much lower scoring the entire time, and not following the same periodic pattern. Its variation as a function of source surface height (gray shaded region) does show that at low source surface height, the GONGz maps can still produce good coronal hole scores. Some intervals in fact show the GONGz optimized score to be

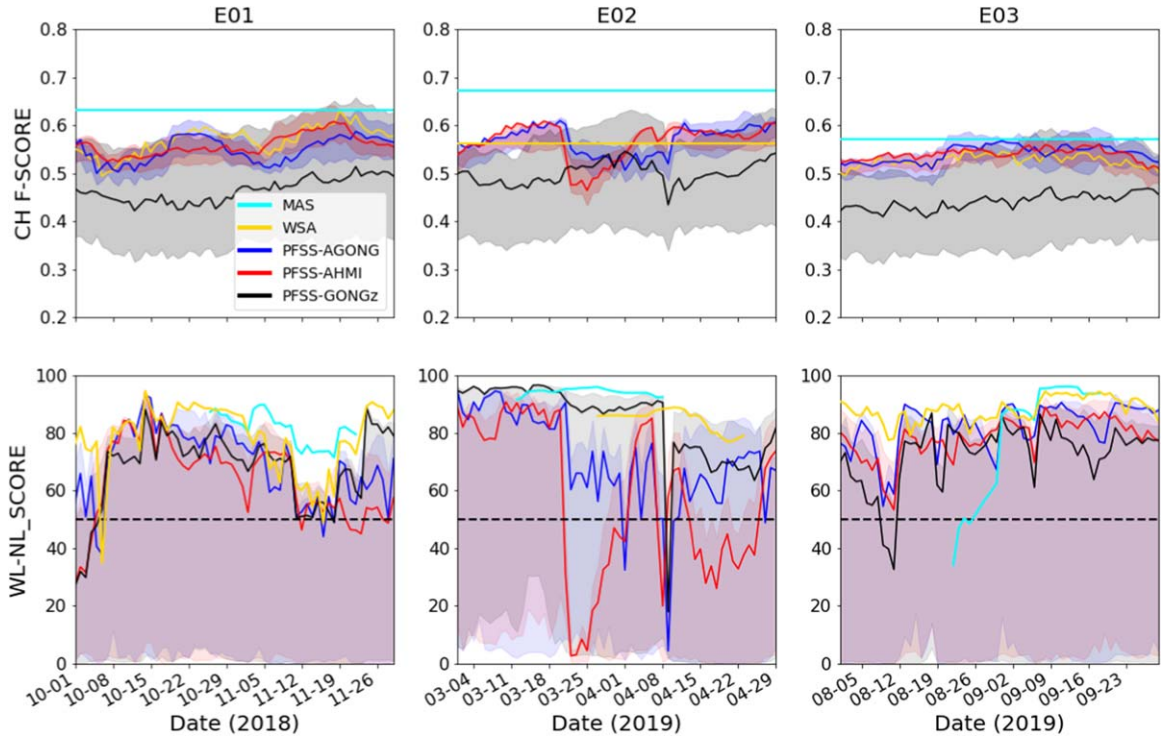


Figure 6. Variation of metric 1 and 2 scores as a function of time within each encounter at a daily cadence. In each panel, colored solid lines indicate different models with the same color scheme as in Figure 5 where a source surface height of 2.5 $R_\odot$ is chosen for the PFSS models (blue, red, black). Variation according to source surface height for the PFSS models is indicated by the shaded regions bounded by the max and min metric score at each time stamp. The top panel shows the coronal hole score (metric 1), the bottom row shows the streamer belt score (metric 2). Note the time series for the MAS model (cyan) is a single value as there is only 1 magnetogram used for each encounter owing to the intractability of running 3D MHD models with time-dependent boundary conditions. For encounter 2, WSA also only uses a single magnetogram and so only has a single value for metric 1 (see main text, Section 4.3).

the best performing of the PFSS models, most notably during the "active region emergence" phase of encounter 2 (see Section 6 conclusion C2).

This bifurcation of scores can be seen in the other encounters' coronal hole metric scores. For encounter 2, the time series is approximately separated into thirds, with the first and last third corresponding to similar scores as encounter 1 for the ADAPT driven models, while in the middle third the scores suddenly drop to similar scores as compared with the GONGz models. Encounter 3 maintains the bifurcation the whole time, and all profiles are very flat. GONGz scores are lower the whole time while the other maps' scores are indistinguishable. Additionally, the MHD score (which is not time varying due to the single model run and single EUV observation) may be compared to the time-varying scores via the horizontal cyan line. For all encounters, the score is never exceeded by the best PFSS (2.5 $R_\odot$) or WSA score. For encounter 2, it is better than all other models and input parameters. In encounters 1 and 3, there are time intervals where different (lower) source surface heights yield scores comparable to or exceeding the MHD value, as indicated by the upper extent of the shading in the figure.

Next, examining the WL metric (bottom row), there are some qualitative similarities and differences to the behavior of the coronal hole scores. For encounters 1 and 3, all models (now including GONGz) follow similar variation in time. The encounter 1 time variation does not show the same periodicity as for the coronal hole metric, instead showing a flat, higher score for the first two-thirds of encounter 1, followed by a sharp drop in score (less pronounced for the MAS model). For encounter 3, a similar drop is observed near the start of the time interval followed by a higher and flat score for the remainder. However, these scores should be interpreted carefully during the first third of encounter 3 since significant portions of the streamer belt were missing due to an issue with the background subtraction (see the animated version of Figure 11 for E3 in the online material). Regions with such artifacts are excluded from the metric computation but deteriorate the reliability by decreasing the number of measurements. Encounter 2 is again distinct from 1 and 3. As with the coronal hole metric, there are distinct thirds of the encounter with the first third having all models having high scores indicating agreement with the observations, followed by a precipitous drop in some models, and a further precipitous, but brief (1–2 days only) drop in all models. It is clear that there are sharp changes detected by both metrics in encounter 2 on approximately 2019 March 23 and 2019 April 9. As will be explored in Section 6, encounter 2 did contrast to encounters 1 and 3 in terms of the presence of dynamic solar activity through that time interval.

5.3.2. *In situ Polarity Time Series*

We conclude our exposition of the metric-score results by examining the in situ polarity metric in detail. As mentioned above, this metric is formed by integrating the measured and modeled polarity time series together, and therefore it only yields one score per encounter. However, we can examine qualitatively how these scores are formed by examining the time series of agreement and disagreement for each model and encounter combination.

This information is encapsulated in Figure 7. Each panel shows model agreement and disagreement for a given

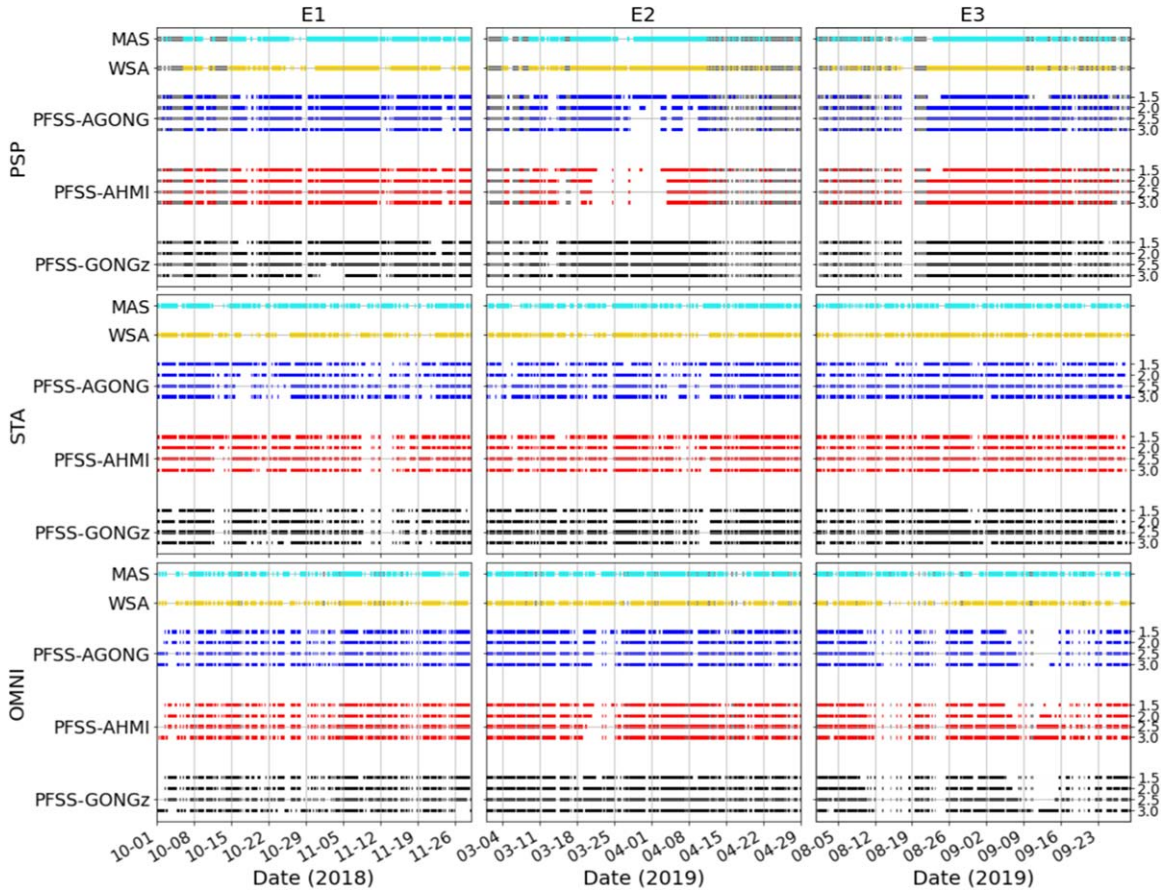


Figure 7. In situ magnetic polarity and prediction comparison as a function of time for each encounter. Metric 3 produces a single score for each encounter as described in the main text, Section 4.3, so it cannot be represented as a time series. This figure depicts how the predicted and measured polarity agreed or disagreed as a function of time and therefore produced a high or low metric score. Columns from left to right show encounters 1, 2, and 3 respectively, while each row shows results from different spacecraft sources: PSP, OMNI, and STEREO A respectively. In each panel, results for the different models are shown, following the color scheme from Figure 5. For the PFSS models, there is a different line of data shown for each source surface height value, ordered from low to high. At an hourly cadence, if the model and measured polarity agree, then that point is filled in with the respective color; if the model and measured polarity disagree, that point is left blank; and if in situ measurements are missing for a given time, they are colored gray. Therefore, gaps in the lines indicate intervals of disagreement between model and measurement. The larger the fraction of the whole time series that is blank, the lower the resulting metric score reported in Figures 4 and 5.

encounter, and given heliospheric spacecraft. For times when the model and data agree, the time series has a colored marker (with color indicating differing models) and is blank if they disagree. For times when there is not in situ data available, the markers are colored gray. Therefore, the times when the measured and modeled polarities disagree (or equivalently, when the positions of predicted HCS crossings are misaligned with the data) may be identified as gaps in the colored lines on the plot. For the PFSS models, the lines for different source surface heights are laid out consecutively (from low to high source surface height). The growth or decay of the white gaps on these consecutive lines indicates if a given HCS crossing improves or worsens with changing source surface height.

Overall, it is clear that most spacecraft data and model HCS combinations agree very well for the vast majority of time stamps, which accounts for the high metric scores. Notable exceptions are with PSP in encounter 2 and OMNI data for encounter 3 for which there are much larger gaps in agreement indicating a severe misalignment between data and modelHCS.

For the PFSS models, we see that typically the disagreements (blank spaces) occur at the same time *for a given spacecraft in all models* and also usually follow the same qualitative trend as a function of source surface height. This indicates that small misalignments between HCS warps and the empirical data correspond to the same heliographic locations independent of the input magnetogram. The MAS and WSA models sometimes have gaps aligned in time with the PFSS models (for example PSP Encounter 1 10/29, top-left panel), but often do not, suggesting the HCS output of these models can be qualitatively different to the PFSS models and sometimes in better agreement with spacecraft data. A very clear example of this is in the top middle panel (E2, PSP) where several large outages in agreement for PFSS models at all source surface heights occur while the MAS and WSA models both correctly predict the polarity of PSP throughout.

The time series are usually very different between different spacecraft. The gaps in agreement can occur at different times for a given model. This is not surprising since the spacecraft are at different solar longitudes and crossing the HCS at different times. However, there are also examples where one spacecraft gives much better agreement than another. Encounter 3 is the clearest case of this where results for the OMNI data set show large intervals of disagreement, while PSP and STEREO A have much more continuous agreement (this may be verified by referring to the ordering of the spacecraft scores in the bottom right panel of Figure 5). Another difference between different

spacecraft is that for PSP the agreement overall is much more continuous, while for the 1 au data sets it is more intermittent with agreement switching on and off from one hour to another. This is caused by the relatively larger fluctuation amplitudes and Parker spiral angle in the B_R data at 1 au (OMNI and STEREO A), meaning that the measured time series include fluctuations in polarity that do not correspond to current sheet crossings.

This concludes our exposition of the metric scoring results in this work. We next discuss the implications of these results, both on the specific time intervals, models, and magnetograms studied here, as well as what they show about this framework for evaluating coronal models.

6. Discussion and Conclusions

In this work, we have defined a model-agnostic framework to evaluate and rank coronal models based on their magnetic field topology using three independent sources of observational data. A given coronal model needs only to provide 2D maps of (a) open magnetic field at the photosphere and (b) magnetic polarity at the model's outer boundary. Models are given a metric score based on (1) how well they represent coronal holes as observed in EUV data for the relevant Carrington rotation, (2) how well they predict streamer belt topology observed by WL coronagraph observations, (3) how well they predict HCS crossing timing measured in situ by various heliospheric spacecraft.

We applied our framework to a data set of coronal models for the first three encounters of PSP using potential field source surface (PFSS) models for a range of source surface heights and input magnetograms, as well as concurrent Wang–Sheeley–Arge (WSA) models and the Magnetohydrodynamic Algorithm outside a Sphere (MAS) models. We carried out three investigations. The first focused on PFSS models only and explored the metric scores as a function of source surface height and choice of magnetogram inner boundary condition. The second compared the PFSS model scores to WSA and MAS scores, evaluating the impact of more physics in the model. The third examined the time evolution of the metric scores in a 60 day interval for each PSP encounter period to look at the stability of the results.

For the PFSS-only investigations, we found the following:

(A1) *Quality of representation of coronal holes at the photosphere is a function of source surface height.*

Values above 2.5 $R_\odot$ lead to observed coronal holes in EUV not being captured by the model (low precision) regardless of input boundary condition. Values too low can also lead to opening fields in the model that are not observed in EUV (low recall), although for models driven by the GONGz data product, the best scores occurred for the lowest source surface height. This indicates that for the same source surface height and observation date, ADAPT maps produce more open flux, as was previously observed in Badman et al. (2020). Examining Figure 8, the optimal score for a given magnetogram occurs when the model and observed boundaries of the polar coronal holes are close together. For ADAPT maps, this occurs near 2.0 $R_\odot$, and below this the polar coronal holes are overexpanded, while for GONGz, the best match is found at 1.5 $R_\odot$. This suggests the coronal hole metric defined here is sensitive to the polar flux treatment in the input magnetogram. The best F-scores were typically around 60%, requiring both reasonable polar and low-latitude coronal hole representation.

(A2) *Representation of streamer belt topology is also strongly dependent on the source surface height of PFSS models, but the dependence is in tension with the coronal hole representation.*

For all magnetogram choices and time intervals, the representation improves monotonically with increasing source surface heights, although with diminishing returns above 2.5 $R_\odot$. This variation is simply explained by how the warps in the HCS produced by PFSS models are smoothed out with increasing source surface heights. WL Carrington maps of the streamer belt depict a typical low-solar-activity HCS with warps at most 40 degrees above or below the solar equator.

(A3) *Prediction of HCS crossing times is shown to be only weakly dependent on source surface height.*

This is interesting since we are testing the same model observable as in the streamer belt metric (HCS location), but against a different data source (in situ magnetic polarity). While metric 2 is investigating amplitudes of warps in the HCS, this metric is investigating the longitudinal position of these warps along the heliographic equator. Therefore we show that warps in the HCS predicted by PFSS are typically in the correct heliographic position, regardless of their amplitude, with the correct spacecraft polarity indicated more than 60% of the time and as much as 90% of the time for certain orbits and spacecraft. It is worth noting that this metric (specifically as applied to 1 au spacecraft) is very similar to the basis for the canonical source surface height of 2.5 $R_\odot$ established by Hoeksema et al. (1983); Hoeksema (1984). We see several examples where this canonical height does give the best polarity score, but it is not universally true, and the score difference between 2.0 and 2.5 $R_\odot$ is typically very small.

(A4) *A "global optimum source surface height" is not in general possible with PFSS models of the solar corona.*

Combining our inferences, we see that, while it is certainly possible to optimize coronal hole representation, or representation of warps in the HCS, these cannot be achieved simultaneously. This impasse is indicative of the bare-bones physical content of PFSS and provides a simple explanation for why different authors who have focused on different physical coronal signatures (e.g., Hoeksema et al. 1983; Lee et al. 2011; Arden et al. 2014) have arrived at different optimum values for the source surface. Nevertheless, we can also see why PFSS modeling is useful and widely used: with generally accepted source surface height parameters (2.0–2.5 Rs), such models can produce good (if not optimum) metric scores in all three metrics. Simultaneous good representation of coronal holes and HCS crossing timing allows for accurate solar wind source mapping and comparison to in situ data.

By extending our investigation from PFSS to include other models with additional physics, we found the following:

(B1) *Inclusion of full MHD physics via the MAS model globally improves model scores for metrics (1) and (2).* This shows that MHD predictions of coronal hole

locations and the general shape of the HCS are more in agreement with observational evidence than PFSS models with any configuration of input parameters.

This shows, unsurprisingly, that the physics present in MHD does produce a more accurate representation of the magnetic structure of the corona at both the inner and outer boundary. At least for encounter 1, there is a marked asymmetry in the dominant warp in the HCS in the MHD model coming from differential flow velocity (Riley et al. 2019), providing an example of the physics present in such models that is neglected in potential models.

(B2) *Comparison of WSA results to PFSS shows global improvement in metric 2 (streamer belt representation) but similar results in the other metrics.*

This is not surprising since what WSA adds to PFSS models is a SCS layer, which slightly intersects the PFSS model domain near the source surface. It is thus not expected to drastically change which coronal loops open up to the solar wind, but it does have a strong effect on the HCS beyond the source surface, flattening it. As mentioned earlier, coronagraph observations of the streamer belt show the HCS is flatter or less warped than PFSS models for $R_{SS} \leqslant 2.0\,R_S$. Therefore WSA offers a solution to the lack of a global optimum source surface height in PFSS models, while maintaining easy computation: the source surface height could be tuned to optimize coronal hole representation, and the WSA SCS layer would then take care of sufficiently flattening the HCS, allowing both inner and outer model boundary conditions to better match observations.

(B3) *For encounter 1 and encounter 3, metric 3 showed little dependence on the model chosen, and some evidence for MHD improvement for encounter 2 where there was solar activity.*

This indicates that, under solar-minimum conditions, the location of warps in the HCS in heliographic longitude are fairly accurately reproduced by PFSS models and do not really change with an addition of more physics. The exception to this for our study was encounter 2 where there was significant solar activity and far side evolution in the magnetograms, and therefore less solar-minimum-like conditions. Here, the score specifically for PSP was significantly improved in the MHD model, but was also dramatically better in GONGz versus the ADAPT PFSS models. This suggests that this difference may be due to inaccurate modeling of far side evolution in the ADAPT magnetograms as well as a lack of accurate model physics.

Overall, then we conclude that our framework for coronal model evaluation successfully allows comparison between models with very different constituents, and moreover shows expected differences between models with different physical content.

Lastly, the time evolution of the metric scores (where available) showed the following:

(C1) *Metric scores do evolve in time with a different profile seen for each time interval studied, but profiles are typically very similar between models using the same magnetogram.*

The score ordering between models is therefore preserved as a function of time. This suggests the time evolution is due to new information being introduced into the magnetogram data or remote observations as the Sun rotates, which affects the model coronal field in qualitatively the same way between models. This has the implication that the far side evolution of the photospheric field, and therefore divergence of the model inner boundary condition from reality, is a fundamental limitation to coronal modeling no matter the physics that is introduced.

(C2) *Far side flux emergence can lead to sudden changes in metric scores.*

The importance of far side evolution is particularly emphasized by PSP encounter 2 (middle column, Figure 6), where step changes in metric scores are observed (around 2019 March 20 and 2019 April 8). For this specific encounter, there was significant solar activity (Pulupa et al. 2020) that evolved and changed the global coronal field while out of view from Earth-based magnetograms, mostly concentrated in two key active regions (AR 12737 and AR12378; Cattell et al. 2021; Harra et al. 2021). This meant a strong photospheric field concentration rotated on disk twice during this time interval and at this time produced sudden and significant changes in magnetograms. This produced simultaneous sharp changes both in the accuracy of representation of coronal holes, and in representation of the global streamer belt topology (both reductions in accuracy as implied by the metrics defined in this work). This change was also observed as a loss of accuracy in metric 3 (middle column Figure 7), most dramatically in PSP data but also to some extent in 1 au data, indicating severe distortion to the model HCS topology by this flux emergence.

(C3) *The agreement between PSP in situ data and the model predictions was typically more continuous than that of STEREO A and OMNI.*

This was illustrated in Figure 7 and demonstrates a simple advantage of sampling the solar wind closer to the Sun for the application of evaluating coronal models: the amplitude of magnetic vector fluctuations to the base value is reduced, and this means a more reliable and continuous measurement of which side of the HCS a given spacecraft is located.

In summary, the focus of this work has been defining a framework for evaluating coronal models using several independent physical observables. A key ingredient was the requirement for models to produce a consistent output (2D maps of the open field of the HCS). This means the authors of future coronal models can easily compare to older models and verify if they produce improved results, and if they only improve certain aspects of the coronal field representation or if the improvement is global across different metrics. We demonstrated this with application to a small subset of models centered around the first three PSP perihelia and were able to compare PFSS, WSA, and MHD models, investigate the source surface height parameter to PFSS models, and observe the impact of far side evolution and flux emergence on coronal model accuracy.

Having specified this evaluation method, there are many examples of future work that can be tackled. We close with listing a few examples.

This work has entirely focused on three time intervals around solar minimum (2018–2019) driven by our use of novel PSP data. As the solar cycle waxes, the behavior of all three metrics on different models will be very interesting to study. For example, it is expected that the accuracy of PFSS models should decay due to the prevalence of nonpotential structures in the active Sun, while MHD models should be more robust.

There is also a wide parameter space of models that should be investigated. The MAS MHD model studied in this work is at the upper end of computational effort and complexity, and there are intermediate codes such as PLUTO (Mignone et al. 2007), MS-FLUKSS (Pogorelov et al. 2013), or AWSoM (van der Holst et al. 2014) that could be cross-compared. Even within the MAS framework, several energetics models exist (Riley et al. 2021) that could be evaluated for impacts on the coronal magnetic structure, thereby allowing these metrics to test different theories of coronal heating and solar wind acceleration. Such investigations could be quite powerful in directly addressing fundamental coronal physics questions.

Further studies of magnetogram–magnetogram comparisons are also clearly motivated by this work. Although we have gone some way toward studying this here through our comparisons of ADAPT-GONG, ADAPT-HMI, and GONGz maps applied to the same models, there remain a large number of other magnetogram sources with different polar coronal hole assumptions or different surface flux transport models and different empirical corrections to correct for the line of sight or magnetograph saturation effects (e.g., Wang et al. 2022). Additionally, we did not touch here on the question of magnetogram resolution. All maps used here had 1×1 degree binning, and just by the visual inspection of, e.g., Figure 8, we see this is more than sufficient to pick out detailed structure of both polar and equatorial coronal holes. Nevertheless, both sharper (e.g., HMI) and coarser (e.g., Wilcox Solar Observatory) resolution maps are available. Caplan et al. (2021) showed how even with potential models, high-resolution magnetograms can be used to produce fine-scale coronal structure that is lost at lower resolution. This framework could be used to evaluate the benefit of having increased resolution in terms of matching observational data.

Finally, although not a specific study, a clear further direction for this work is in the creation of an open-source tool or web application to perform metric evaluations. This tool could be used by any coronal modeler by submitting the coronal hole and NL maps in the format prescribed here and are then presented with the resulting metric scores and supporting plots. While such an integrated software tool is beyond the scope of the current project, we presented in Algorithm Boxes 1, 2, and 3 descriptions of the algorithm to produce the respective metric score. We also supply the base observational data products to perform the comparisons for the time intervals studied in this work as a Zenodo repository (doi:10.5281/zenodo.6342187) in an effort to allow other modelers to try evaluating this framework on their own models. As well as the specific observational data products, Python scripts are included for each metric as examples for loading and plotting the different file types.

We acknowledge support from ISSI for the team 463, entitled "Exploring The Solar Wind In Regions Closer Than Ever Observed Before." Parker Solar Probe was designed, built, and is now operated by the Johns Hopkins Applied Physics Laboratory as part of NASA's Living with a Star (LWS) program (contract NNN06AA01C). Support from the LWS management and technical team has played a critical role in the success of the Parker Solar Probe mission. The FIELDS and SWEAP experiments on the Parker Solar Probe spacecraft were designed and developed under NASA contract NNN06AA01C. The authors acknowledge the extraordinary contributions of the Parker Solar Probe mission operations and spacecraft engineering teams at the Johns Hopkins University Applied Physics Laboratory. This work utilizes data obtained by the Global Oscillation Network Group (GONG) Program, managed by the National Solar Observatory, which is operated by AURA, Inc. under a cooperative agreement with the National Science Foundation. The data were acquired by instruments operated by the Big Bear Solar Observatory, High Altitude Observatory, Learmonth Solar Observatory, Udaipur Solar Observatory, Instituto de Astrofísica de Canarias, and Cerro Tololo Interamerican Observatory. This work utilizes data produced collaboratively between Air Force Research Laboratory (AFRL) & the National Solar Observatory (NSO). The ADAPT model development is supported by AFRL. The input data utilized by ADAPT is obtained by NSO/NISP (NSO Integrated Synoptic Program). This research made use of HelioPy, a community-developed Python package for space physics (Stansby et al. 2019). This research has made use of SunPy v2.0.1 (Mumford et al. 2019), an open-source and free community-developed solar data analysis Python package (SunPy Community et al. 2015). The work of D.H.B. and H.P.W. was performed under contract to the Naval Research Laboratory and was funded by the NASA Hinode program. S.D.B. acknowledges the support of the Leverhulme Trust Visiting Professorship program. S.T.B. was supported by NASA Mary W. Jackson Headquarters under the NASA Earth and Space Science Fellowship Program grant 80NSSC18K1201. The work of A.R. and N.P. was funded by the ERC SLOW_SOURCE project (SLOW_SOURCE—DLV-819189). We acknowledge use of NASA/GSFC's Space Physics Data Facility's OMNIWeb service and OMNI data. P.R. gratefully acknowledges support from NASA (80NSSC18K0100, NNX16AG86G, 80NSSC18K1129, 80NSSC18K0101, 80NSSC20K1285, and NNN06AA01C).

Appendix
Full Model-observation Comparisons

This appendix contains summary plots of the observation-model comparisons for all metrics considered in the work, and the plots are animated as a function of time for each encounter in the online version.

A.1. Metric 1: Coronal Hole—Open Field Comparison

Here, in Figures 8–10, we show schematics of the binary classification from EZSEG and the coronal model open field maps detailed in Section 4.1. The panels are organized to differentiate the models and (for PFSS models) the source surface height. Each panel has the precision and recall shown as a text inset. As a reminder, the quantity quoted in the main text (Section 5), the "F-score", is the harmonic mean of these two quantities. These figures are animated over each 60 day interval in the online version of this article. The still images shown here are the first frames of the movie from 2018 October 1, 2019 March 1, and 2019 August 1 for encounters 1–3 respectively.

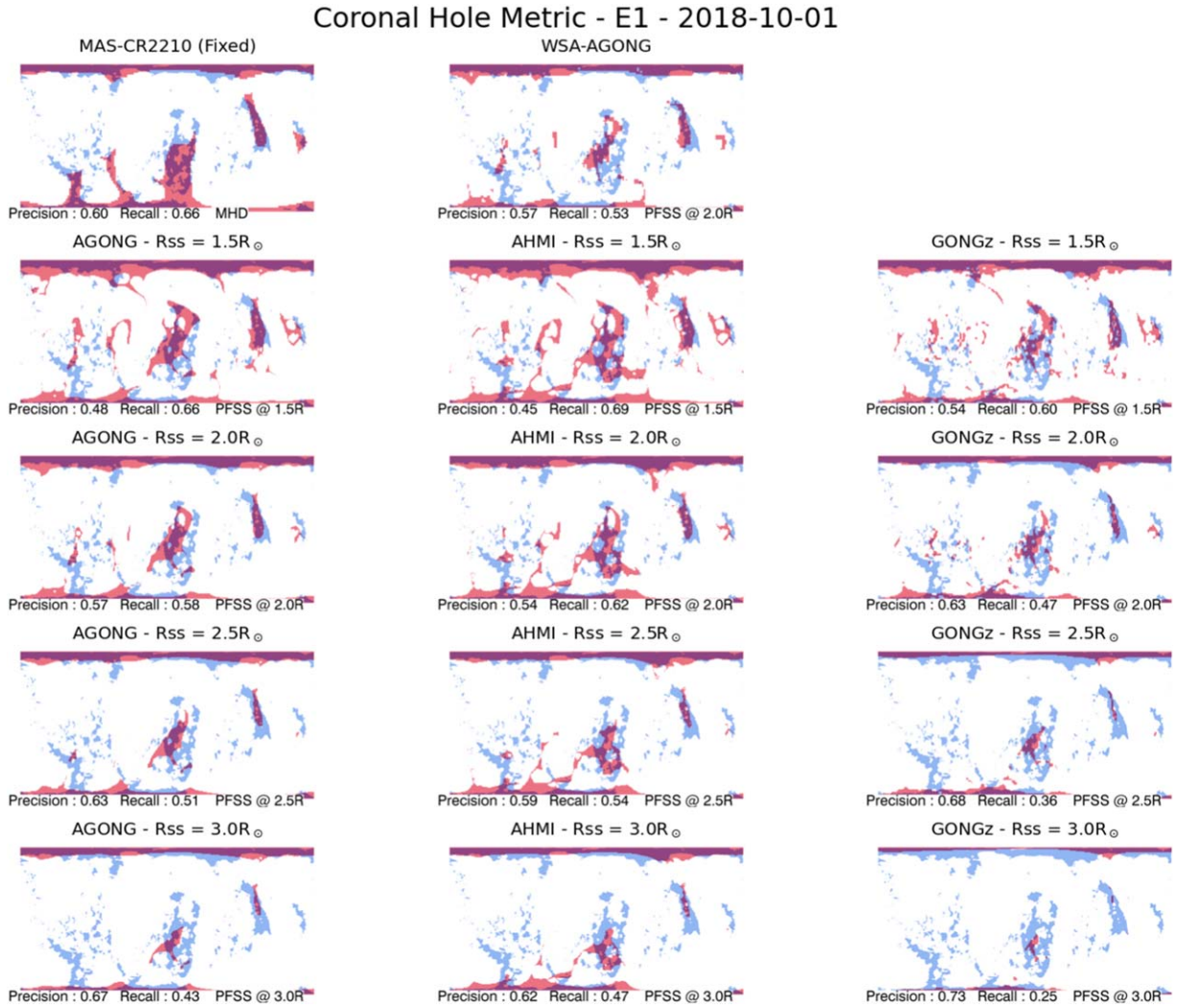


Figure 8. Metric 1 (Coronal Hole) model-observation comparison for the first PSP encounter. The panels depict the binary classification scheme used to define our coronal hole metric, as detailed in Section 4.1 and illustrated in the bottom panel of Figure 1. Blue shading indicates the coronal hole area determined by applying the EZSEG algorithm to EUV synoptic maps; red shading indicates the open field area determined from the particular coronal model run. Where the model and observations agree, the pixels are shaded purple. The titles of each panel indicate the model and model parameters. The top row shows MAS and WSA results, while the bottom four rows show PFSS results with the columns differentiating the model and the rows differentiating the source surface height. Inset text in each panel records the associated precision and recall scores. In the online material, this figure is animated as a function of time for the encounter from 2018 October 1 to 2018 November 30 at a daily cadence. Although the background EUV map is fixed, the model coronal hole distributions evolve except for the MAS MHD results (top-left panel), which are based on a single magnetogram for this encounter.

(An animation of this figure is available.)

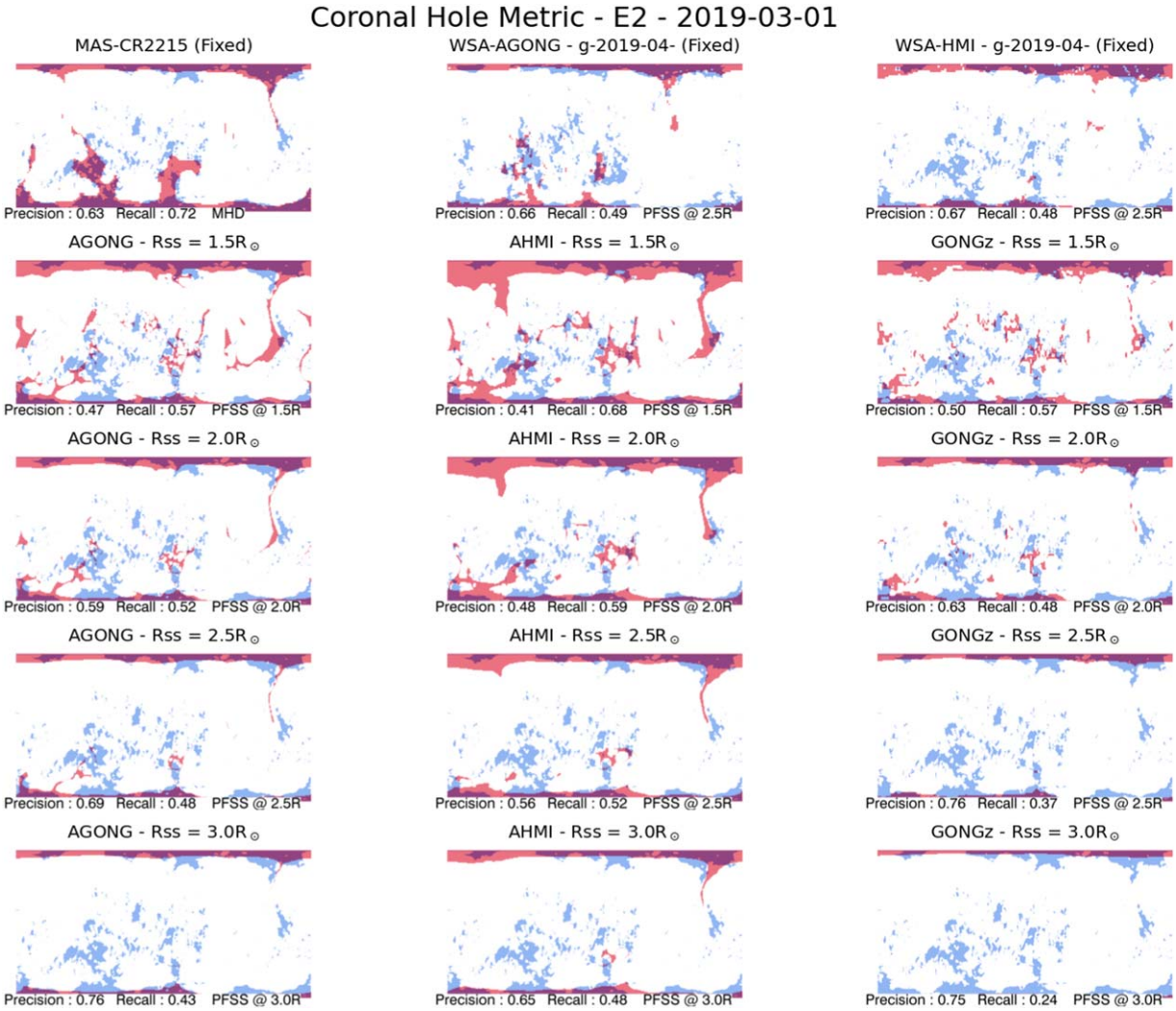


Figure 9. Metric 1 (Coronal Hole) model-observation comparison for the second PSP encounter. The panels are organized as in Figure 8. The online animation shows daily evolution from 2019 March 1 to 2019 April 30 except for the MAS and WSA models (top row), which are based on a single magnetogram for this encounter. (An animation of this figure is available.)

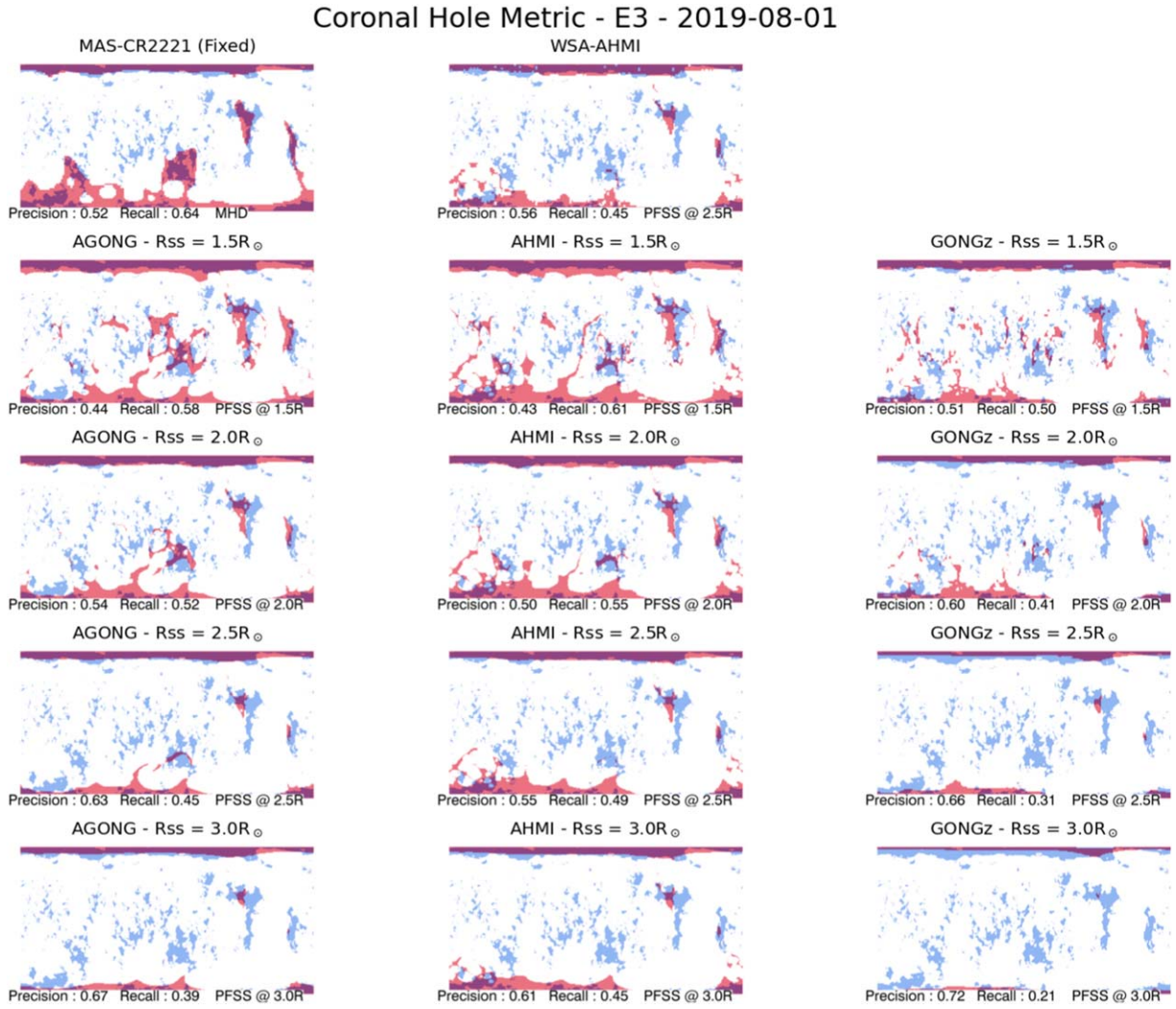


Figure 10. Metric 1 (Coronal Hole) model-observation comparison for the third PSP encounter. The panels are organized as in Figure 8. The online animation shows daily evolution from 2019 August 1 to 2019 September 30 except for the MAS model (top-left panel), which is based on a single magnetogram for this encounter. (An animation of this figure is available.)

A.2. Metric 2: Neutral Line—Streamer Belt Comparison

Following the format of Section A.1, we next show schematically the comparison between the modeled and measured streamer belt topology. Again we show a multipanel figure for each encounter (Figures 11–13) with panels organized by model and source surface height. Each panel follows the same format as the bottom panel of Figure 3: the background image, black contour, and dashed centroid curve show the WL Carrington map and extracted streamer belt (SMB line), while the red curve shows the model HCS, and the blue lines show the distances between the curves that contribute to the metric score, as detailed in Section 4.2 and Poirier et al. (2021). In the online version of this article, these figures are movies for each encounters over the 60 day intervals (in which both elements evolve with time for the 60 day interval around perihelion, except where noted in the figure captions). The static versions of these figures show the first frame of each movie from dates 2018 October 1, 2019 March 1, and 2019 August 1 for encounters 1–3 respectively.

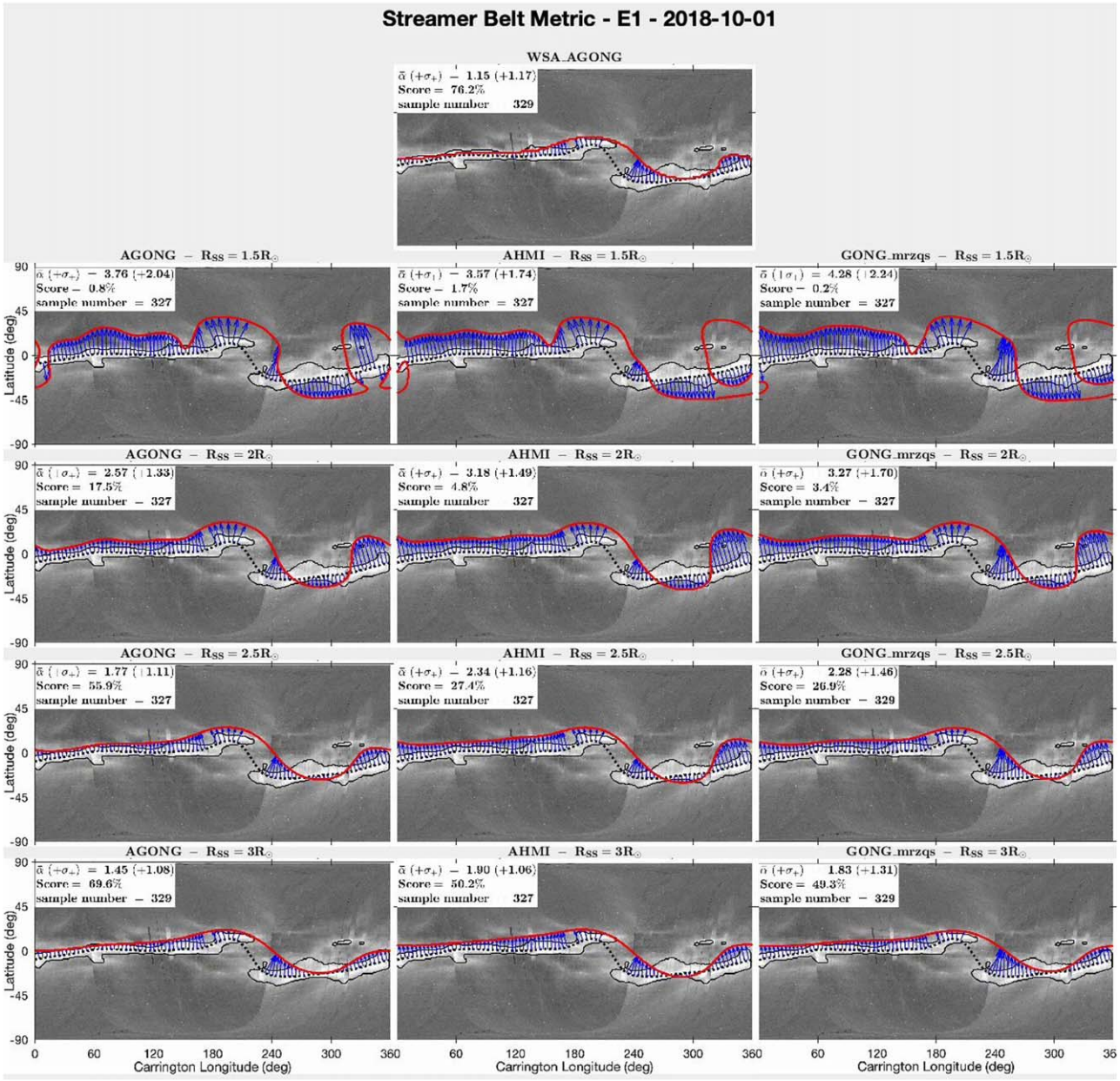


Figure 11. Metric 2 (Streamer Belt) model-observation comparison for the first PSP encounter. The panels show the comparison between the extracted streamer belt line and the model HCS, as detailed in Section 4.2 and illustrated in the bottom panel of Figure 2. The streamer belt score as well as the intermediate $\bar{\alpha}$ and $\sigma_{\pm}$ quantities are given in the top-left corner of each panel. In the online version of this article, this figure is animated at a daily cadence from 2018 October 1 to 2018 November 30, showing how the model neutral lines and background Carrington white-light (WL) maps evolve as a function of time (except the MAS model, top-left panel, for which only the WL map evolves in this encounter).

(An animation of this figure is available.)

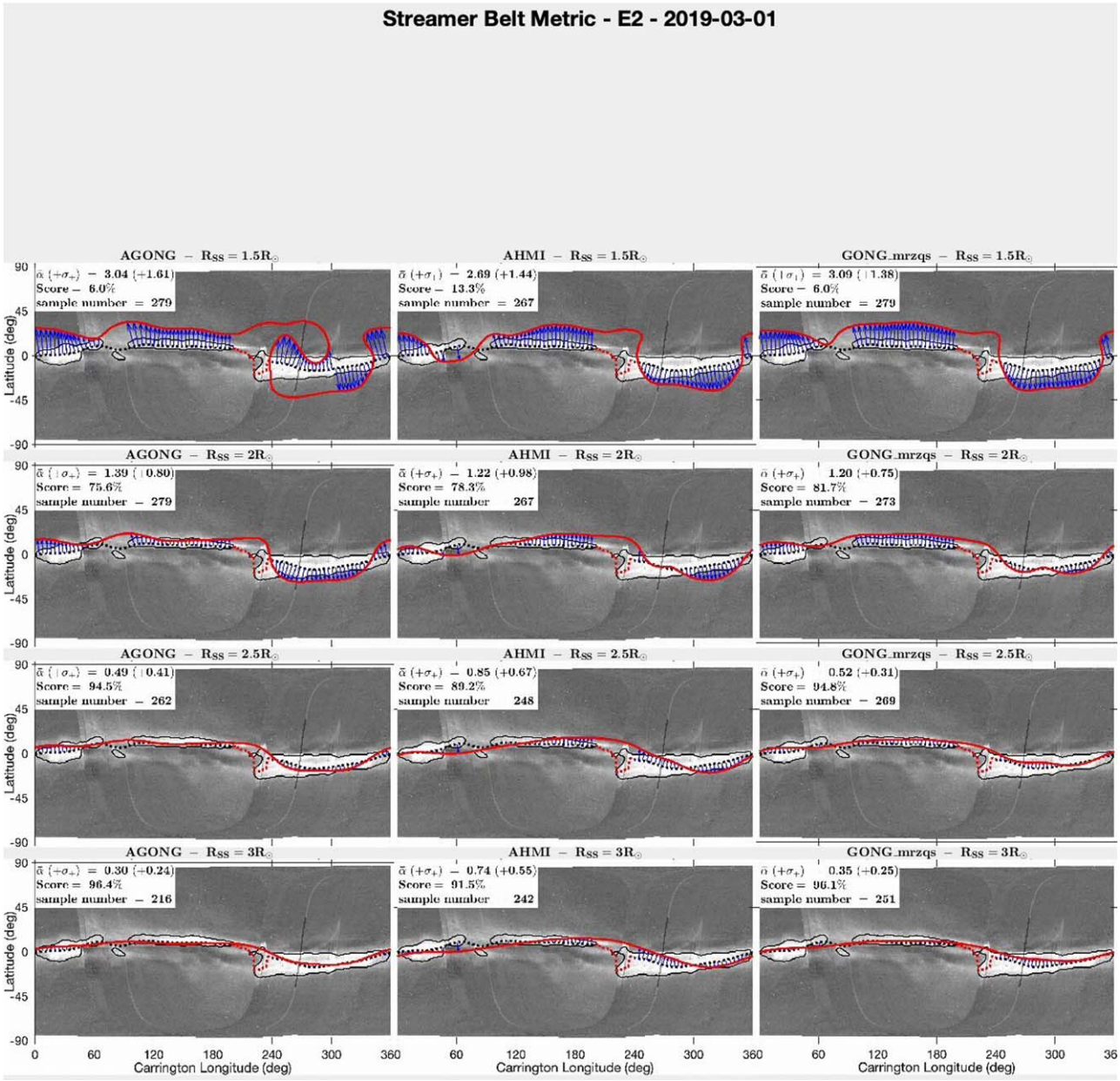


Figure 12. Metric 2 (Streamer Belt) model-observation comparison for the second PSP encounter. The panels are organized as in Figure 11. In the online version of this article, this figure is animated at a daily cadence from 2019 March 1 to 2019 April 30, except for the MAS and WSA models (top row) for which only the WL map evolves.

(An animation of this figure is available.)

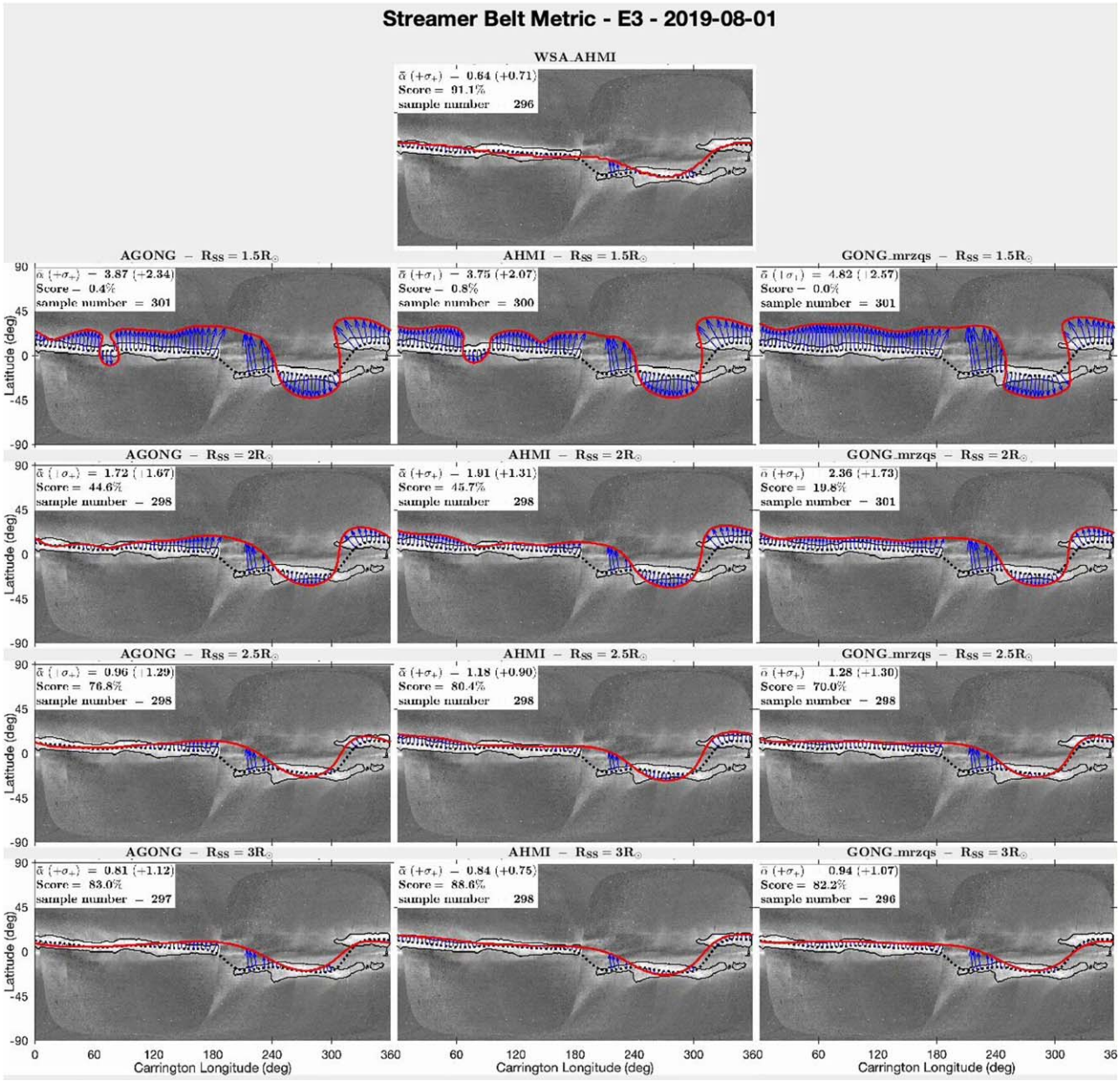


Figure 13. Metric 2 (Streamer Belt) model-observation comparison for the third PSP perihelion. The panels are organized as in Figure 11. In the online version of this article, this figure is animated at a daily cadence from 2019 August 1 to 2019 September 30, except for the MAS models (top right panel) for which only the WL map evolves.

(An animation of this figure is available.)

A.3. Metric 3: In situ Polarity—HCS Comparison

Finally, in Figures 14–16, we summarize model-observation comparisons for the third metric studied (in situ polarity). These figures are corollaries to Figure 7, which already displayed in a compact way all model/data comparisons for this metric as function of time. Here we plot the measured data in the format of the middle panel of the illustrative Figure 3, showing the measured polarity time-series data plotted as the spacecraft trajectory ballistically mapped across the model's outer boundary. This is to allow the reader to see visually how the change in the model's HCS topology affects agreement with the in situ data.

As in Sections A.1 and A.2, here we have an individual multipanel plot for each encounter for all models and model input parameters studied in this work. In the online version of this article, these figures are animated over the relevant 60 day interval for each encounter, showing the movement of the HCS as the models evolve (except where noted in the figure captions) as the spacecraft whose in situ data we use move in the heliographic frame. The still images shown here are the first frames of these movies, snapshotting 2018 October 1, 2019 March 1, and 2019 August 1 for encounters 1–3 respectively.

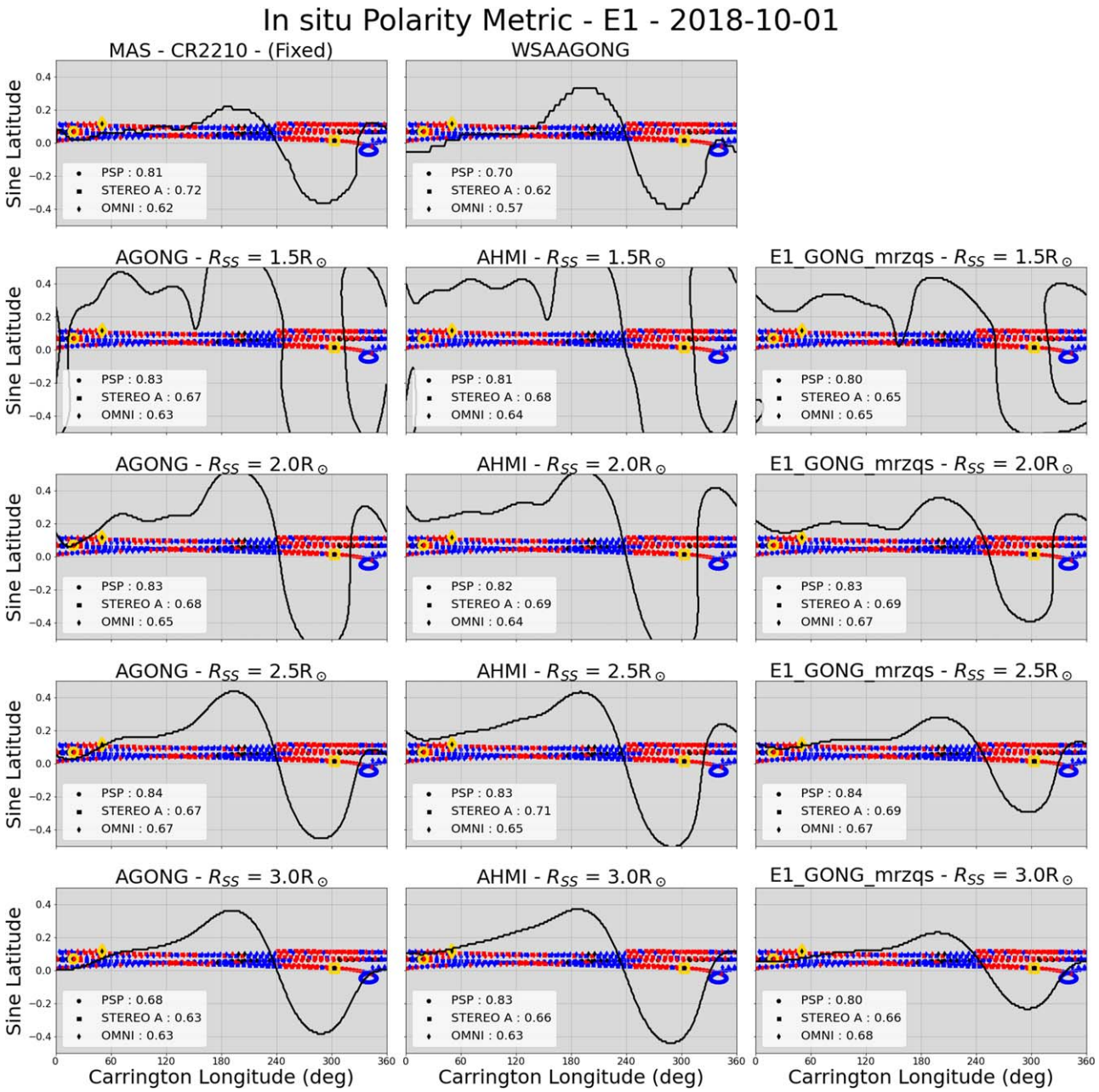


Figure 14. In situ polarity model-observation comparison for models for the first PSP encounter. The panels schematically show the extent to which the in situ measured polarity (blue/red markers) corresponds to the modeled heliospheric current sheet (black contour), as detailed in Section 4.1 and illustrated in the middle panel of Figure 3. Model-data agreement occurs where the in situ data is taken northwards of a HCS contour *and* is red (positive polarity) or where the data is taken southwards of the HCS contour *and* is blue (negative polarity). The panel titles indicate the model parameters. The top row shows MAS and WSA results, while the bottom four rows show PFSS results with the columns differentiating the model and the rows differentiating the source surface height. In the online version, this figure is animated for the 60 day interval with each frame advancing in time one day at a time from 2018 October 1 to 2018 November 30. As time evolves, the model HCSs evolve for each model (except for the MAS model, top-left panel that only has a single model run for E1). Simultaneously, the spacecraft (yellow symbols) moves in the solar corotating reference frame. The static version of the figure is the first frame for the video, showing the metrics for the date 2018 October 1.

(An animation of this figure is available.)

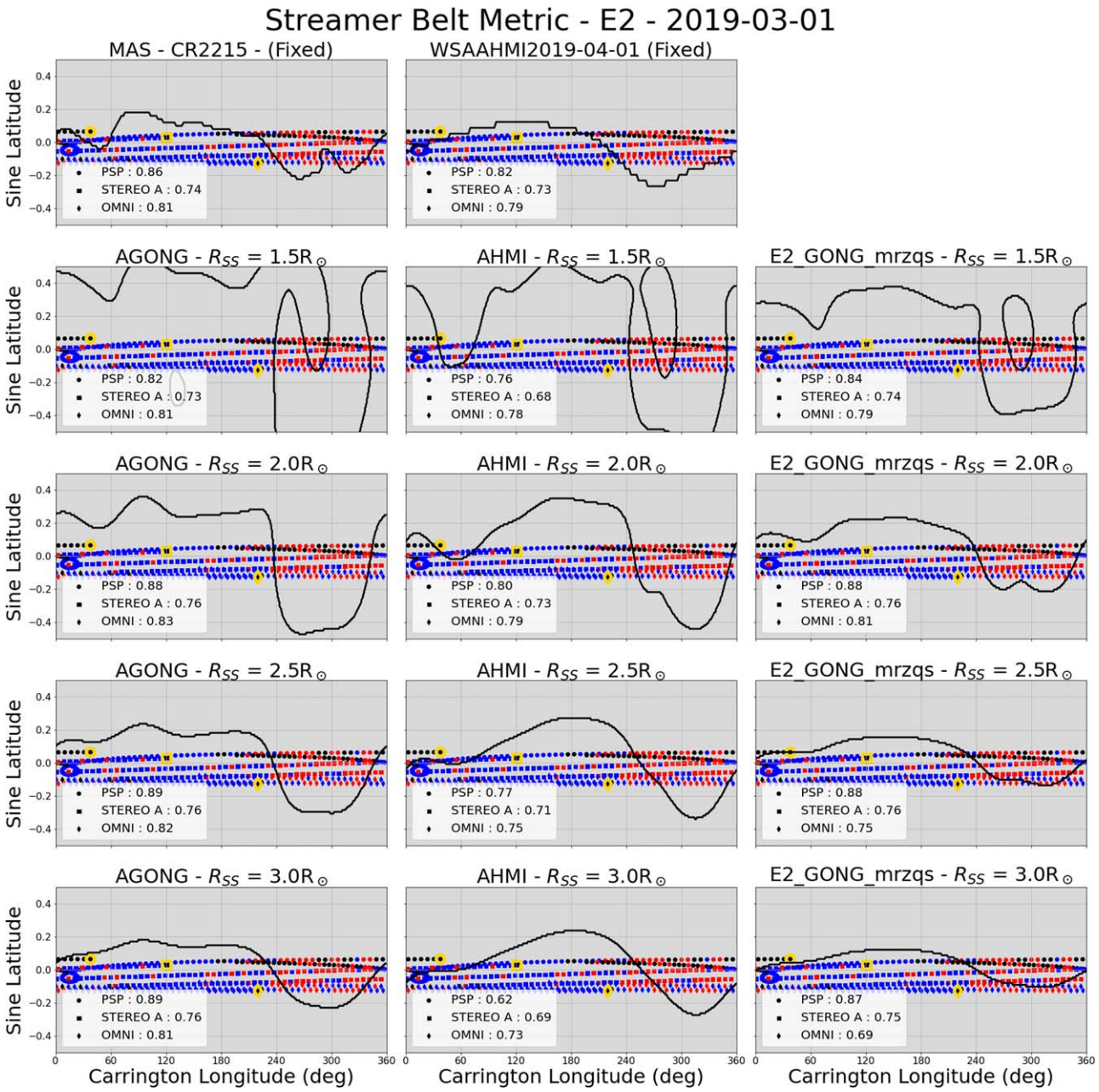


Figure 15. In situ polarity model-observation comparison for models for the second PSP encounter. The panels are organized as in Figure 14. In the online version of this article, this figure is animated for the 60 day interval for encounter 2 from 2019 March 1 to 2019 April 30 showing relative motion of the model HCS and spacecraft positions (projected to the source surface). Here the model HCS from MAS and WSA (top row) are frozen in time. The static version of the figure is the first frame for the video, showing the metrics for the date 2019 March 1.

(An animation of this figure is available.)

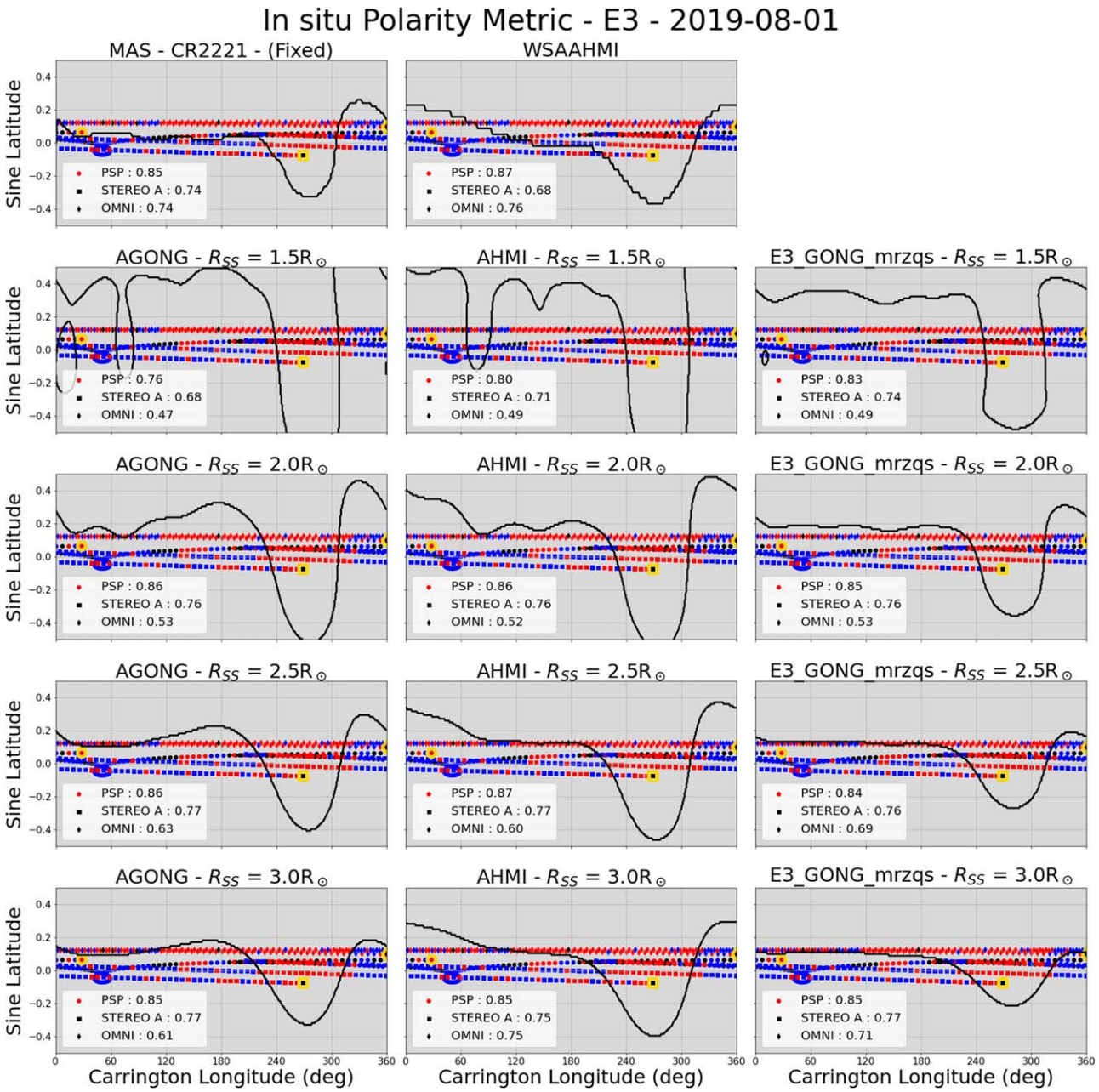


Figure 16. In situ polarity model-observation comparison for models for the third PSP encounter. The panels are organized as in Figure 14. In the online version of this article, this figure is animated for the 60 day interval for encounter 3 from 2019 August 1 to 2019 September 30, showing relative motion of the model HCS and spacecraft positions (projected to the source surface). Here the model HCS from MAS (top-left panel) is frozen in time. The static version of the figure is the first frame for the video, showing the metrics for the date 2019 August 1.

(An animation of this figure is available.)

ORCID iDs

Samuel T. Badman https://orcid.org/0000-0002-6145-436X
David H. Brooks https://orcid.org/0000-0002-2189-9313
Nicolas Poirier https://orcid.org/0000-0002-1814-4673
Harry P. Warren https://orcid.org/0000-0001-6102-6851
Gordon Petrie https://orcid.org/0000-0001-8462-9161
Alexis P. Rouillard https://orcid.org/0000-0003-4039-5767
C. Nick Arge https://orcid.org/0000-0001-9326-3448
Stuart D. Bale https://orcid.org/0000-0002-1989-3596
Diego de Pablos Agüero https://orcid.org/0000-0002-5469-2378
Louise Harra https://orcid.org/0000-0001-9457-6200
Shaela I. Jones https://orcid.org/0000-0001-9498-460X
Athanasios Kouloumvakos https://orcid.org/0000-0001-6589-4509
Pete Riley https://orcid.org/0000-0002-1859-456X
Olga Panasenco https://orcid.org/0000-0002-4440-7166
Marco Velli https://orcid.org/0000-0002-2381-3106
Samantha Wallace https://orcid.org/0000-0002-1091-4688

References

Altschuler, M. D., & Newkirk, G. 1969, SoPh, 9, 131
Arden, W. M., Norton, A. A., & Sun, X. 2014, JGRA, 119, 1476
Arge, C., Luhmann, J., Odstrcil, D., Schrijver, C., & Li, Y. 2004, JASTP, 66, 1295
Arge, C. N., Henney, C. J., Hernandez, I. G., et al. 2013, in AIP Conf. Proc. 1539, Solar Wind 13, ed. G. P. Zank et al. (Melville, NY: AIP), 11
Arge, C. N., Henney, C. J., Koller, J., et al. 2010, in AIP Conf. Proc. 1216, Twelfth Int. Solar Wind Conf. (Melville, NY: AIP), 343
Arge, C. N., Henney, C. J., Koller, J., et al. 2011, in ASP Conf. Ser. 444, 5th Int. Conf. of Numerical Modeling of Space Plasma Flows, ed. N. V. Pogorelov, E. Audit, & G. P. Audit (San Francisco, CA: ASP), 99
Arge, C. N., Odstrcil, D., Pizzo, V. J., & Mayer, L. R. 2003, in AIP Conf. Proc. 679, Solar Wind Ten (Melville, NY: AIP), 190
Arge, C. N., & Pizzo, V. J. 2000, JGRA, 105, 10465
Badman, S. T., Bale, S. D., Oliveros, J. C. M., et al. 2020, ApJS, 246, 23
Bale, S. D., Goetz, K., Harvey, P. R., et al. 2016, SSRv, 204, 49
Boe, B., Habbal, S., & Druckmüller, M. 2020, ApJ, 895, 123
Brooks, D. H., Ugarte-Urra, I., & Warren, H. P. 2015, NatCo, 6, 5947
Brueckner, G. E., Howard, R. A., Koomen, M. J., et al. 1995, SoPh, 162, 357
Caplan, R. M., Downs, C., & Linker, J. A. 2016, ApJ, 823, 53
Caplan, R. M., Downs, C., Linker, J. A., & Mikic, Z. 2021, ApJ, 915, 44
Cattell, C., Glesener, L., Leiran, B., et al. 2021, A&A, 650, A6
de Patoul, J., Foullon, C., & Riley, P. 2015, ApJ, 814, 68
Domingo, V., Fleck, B., & Poland, A. I. 1995, SoPh, 162, 1
Druckmuller, M. 2008, S&T, 116, 102
Feng, X. 2020, Magnetohydrodynamic Modeling of the Solar Corona and Heliosphere (Berlin: Springer)
Fox, N. J., Velli, M. C., Bale, S. D., et al. 2016, SSRv, 204, 7
Freeland, S. L., & Handy, B. N. 1998, SoPh, 182, 497
Garton, T. M., Gallagher, P. T., & Murray, S. A. 2018, JSWSC, 8, A02
Hansen, S. F., Sawyer, C., & Hansen, R. T. 1974, GeoRL, 1, 13
Harra, L., Brooks, D. H., Bale, S. D., et al. 2021, A&A, 650, A7
Harvey, J. W., Hill, F., Hubbard, R. P., et al. 1996, Sci, 272, 1284
Harvey, K. L., & Recely, F. 2002, SoPh, 211, 31
Hickmann, K. S., Godinez, H. C., Henney, C. J., & Arge, C. N. 2015, SoPh, 290, 1105
Hill, F. 2018, SpWea, 16, 1488
Hoeksema, J. T. 1984, PhD thesis, Stanford Univ.
Hoeksema, J. T., Wilcox, J. M., & Scherrer, P. H. 1983, JGR, 88, 9910
Howard, R. A., & Koomen, M. J. 1974, SoPh, 37, 469
Kaiser, M. L., Kucera, T. A., Davila, J. M., et al. 2008, SSRv, 136, 5
Krista, L. D., & Gallagher, P. T. 2009, SoPh, 256, 87
Kruse, M., Heidrich-Meisner, V., & Wimmer-Schweingruber, R. F. 2021, A&A, 645, A83
Kruse, M., Heidrich-Meisner, V., Wimmer-Schweingruber, R. F., & Hauptmann, M. 2020, A&A, 638, A109
Lee, C. O., Luhmann, J. G., Hoeksema, J. T., et al. 2011, SoPh, 269, 367
Lemen, J. R., Title, A. M., Akin, D. J., et al. 2012, SoPh, 275, 17
Lionello, R., Linker, J. A., & Mikić, Z. 2001, ApJ, 546, 542
Lionello, R., Linker, J. A., & Mikić, Z. 2009, ApJ, 690, 902
Luhmann, J. G., Curtis, D. W., Schroeder, P., et al. 2008, SSRv, 136, 117
Macneil, A. R., Owens, M. J., Finley, A. J., & Matt, S. P. 2021, MNRAS, 509, 2390
Malanushenko, O. V., & Jones, H. P. 2005, SoPh, 226, 3
Mignone, A., Bodo, G., Massaglia, S., et al. 2007, ApJS, 170, 228
Mikić, Z., Downs, C., Linker, J. A., et al. 2018, NatAs, 2, 913
Mikic, Z., & Linker, J. A. 1994, ApJ, 430, 898
Mikić, Z., & Linker, J. A. 1996, in AIP Conf. Proc. 382, Proc. Eighth Int. Solar Wind Conf.: Solar Wind Eight, ed. D. Winterhalter et al. (Melville, NY: AIP), 104
Mikić, Z., Linker, J. A., Schnack, D. D., Lionello, R., & Tarditi, A. 1999, PhPl, 6, 2217
Morgan, H., & Cook, A. C. 2020, ApJ, 893, 57
Mumford, S., Christe, S., Mayer, F., et al. 2019, sunpy/sunpy: v0.9.6, doi:10.5281/zenodo.2551710
Nolte, J. T., & Roelof, E. C. 1973, SoPh, 33, 241
Panasenco, O., Velli, M., D'Amicis, R., et al. 2020, ApJS, 246, 54
Pesnell, W. D., Thompson, B. J., & Chamberlin, P. C. 2012, SoPh, 275, 3
Pinto, R. F., & Rouillard, A. P. 2017, ApJ, 838, 89
Pogorelov, N. V., Borovikov, S. N., Bedford, M. C., et al. 2013, in ASP Conf. Ser. 474, Numerical Modeling of Space Plasma Flows (ASTRONUM2012), ed. N. V. Pogorelov, E. Audit, & G. P. Zank (San Francisco, CA: ASP), 165
Poirier, N., Rouillard, A. P., Kouloumvakos, A., et al. 2021, FrASS, 8, 84
Pomoell, J., & Poedts, S. 2018, JSWSC, 8, A35
Pulupa, M., Bale, S. D., Badman, S. T., et al. 2020, ApJS, 246, 49
Réville, V., & Brun, A. S. 2017, ApJ, 850, 45
Rice, O. E. K., & Yeates, A. R. 2021, ApJ, 923, 57
Riley, P., Downs, C., Linker, J. A., et al. 2019, ApJ, 874, L15
Riley, P., Linker, J. A., & Mikić, Z. 2001, JGR, 106, 15889
Riley, P., Linker, J. A., Mikić, Z., et al. 2006, ApJ, 653, 1510
Riley, P., Lionello, R., Caplan, R. M., et al. 2021, A&A, 650, A19
Rouillard, A. P., Kouloumvakos, A., Vourlidas, A., et al. 2020, ApJS, 246, 37
Sasso, C., Pinto, R. F., Andretta, V., et al. 2019, A&A, 627, A9
Schatten, K. H. 1972, in NASA Special Publication 308, Solar Wind, ed. C. P. Sonett, P. J. Coleman, & J. M. Wilcox (Washington, DC: NASA), 44
Schatten, K. H., Wilcox, J. M., & Ness, N. F. 1969, SoPh, 6, 442
Scherrer, P. H., Schou, J., Bush, R. I., et al. 2012, SoPh, 275, 207
Sokolov, I. V., van der Holst, B., Oran, R., et al. 2013, ApJ, 764, 23
Stansby, D., Horbury, T. S., Wallace, S., & Arge, C. N. 2019, RNAAS, 3, 57
Stansby, D., Yeates, A., & Badman, S. 2020, JOSS, 5, 2732
SunPy Community, T., Mumford, S. J., Christe, S., et al. 2015, CS&D, 8, 014009
van der Holst, B., Sokolov, I. V., Meng, X., et al. 2014, ApJ, 782, 81
Wagner, A., Asvestari, E., Temmer, M., Heinemann, S. G., & Pomoell, J. 2022, A&A, 657, A117
Wang, Y. M., & Sheeley Jr., N. R. 1992, ApJ, 392, 310
Wang, Y. M., Sheeley, N. R., & Rich, N. B, J. 2007, ApJ, 658, 1340
Wang, Y. M., Sheeley, N. R. J., Walters, J. H., et al. 1998, ApJL, 498, L165
Wang, Y. M., Ulrich, R. K., & Harvey, J. W. 2022, ApJ, 926, 113
Warren, H. P., & Brooks, D. H. 2009, ApJ, 700, 762
Wiegelmann, T., & Sakurai, T. 2021, LRSP, 18, 1