

RESEARCH ARTICLE

OPEN ACCESS

Helio1D modeling of temporal variation of solar wind: Interfacing between MULTI-VP and 1D MHD for future operational forecasting at L1

R. Kieokaew^{1,*}, R.F. Pinto^{2,1}, E. Samara^{3,4}, C. Tao⁵, M. Indurain¹, B. Lavraud^{6,1}, A. Brunet⁷, V. Génot¹, A. Rouillard¹, N. André¹, S. Bourdarie⁷, C. Katsavrias⁸, F. Darrouzet⁹, B. Grison¹⁰ and I. Daglis^{8,11}

¹ Institut de Recherche en Astrophysique et Planétologie, UPS, CNES, CNRS, Toulouse, France

² Département d'Astrophysique/AIM, CEA/IRFU, CNRS/INSU, Univ. Paris-Saclay & Univ. de Paris, 91191 Gif-sur-Yvette, France

³ CmPA, KU Leuven, Leuven, Belgium

⁴ NASA, Goddard Space Flight Center, Greenbelt, MD, USA

⁵ National Institute of Information and Communications Technology (NICT), Koganei, Tokyo 184-8795, Japan

⁶ Laboratoire d'astrophysique de Bordeaux, Univ. Bordeaux, CNRS, Pessac, France

⁷ ONERA/DPHY, Université de Toulouse, F-31055 Toulouse, France

⁸ Department of Physics, National and Kapodistrian University of Athens, 15784 Athens, Greece

⁹ Royal Belgian Institute for Space Aeronomy (IASB-BIRA), 1180 Brussels, Belgium

¹⁰ Institute of Atmospheric Physics CAS, Dept of Space Physics, 14100 Prague, Czech Republic

¹¹ Hellenic Space Center, Athens, Greece

Received 15 March 2023 / Accepted 22 May 2024

Abstract—Developing an automated pipeline for solar-wind condition prediction upstream of Earth is an important step for transitioning from space weather research to operation. We develop a prototype pipeline called “Helio1D” to model ambient solar wind conditions comprising temporal profiles of wind speed, density, temperature, and tangential magnetic field at L1 up to 4 days in advance. The prototype pipeline connects the MULTI-VP coronal model that provides daily predictions of the solar wind at 0.14 AU and a 1D magnetohydrodynamics (MHD) model that propagates the solar wind to 1 AU. As a part of development towards a better-performing operational pipeline in the future, our present work focuses on the proof-of-concept, initial implementation, and validation of Helio1D. Here, we first benchmark Helio1D using the synoptic magnetograms provided by Wilcox Space Observatory as inputs to the coronal part of MULTI-VP for the intervals in 2004–2013 and 2017–2018. Using the classic point-to-point metrics, it is found that Helio1D underperforms the 27-day recurrence model for all time intervals while outperforming the 4-day persistence model in the late declining phase of the solar cycle. As a complementary analysis, we evaluate the time and magnitude differences between Helio1D and the observations by exploiting the Fast Dynamic Time Warping technique, which allows us to discuss Helio1D caveats and address calibration to improve the Helio1D performance. Furthermore, we model several solar wind conditions in parallel, for a total of 21 profiles corresponding to various virtual targets to provide uncertainties. Although our prototype pipeline shows less satisfactory results compared to existing works, it is fully automated and computationally fast, both of which are desirable qualities for operational forecasting. Our strategies for future improvements towards better-performing pipeline are addressed.

Keywords: Space weather / Solar wind / Corotating interaction region

1 Introduction

The near-Earth space environment is influenced by the continuous streams of solar wind emanating from the Sun.

*Corresponding author: rkieokaew@irap.omp.eu

Developing a system or a framework to robustly forecast space weather, in particular, the solar wind, becomes imperative as human society increasingly relies on space technologies and modern ground facilities that can be impacted by space weather conditions. Earth-directed coronal mass ejections (CMEs) caused by solar eruptions are the major drivers that can have

profound effects. Stream interaction regions, formed when a fast solar-wind stream takes over a slower stream, are another important driver. Since these regions co-rotate with the Sun, they are also called corotating interaction regions (CIRs), and thus they impact the Earth more regularly, i.e., every solar rotation. Modeling the solar wind consists of two parts: the ambient solar wind and the solar transients; the former comprises CIRs and the latter usually consists of an ad-hoc modeling of CMEs. Hereafter, we focus mainly on modeling and forecasting of the ambient solar wind comprising CIRs. CIRs cause low- to intermediate-strength geomagnetic storms and occasionally lead to strong geomagnetic storms (e.g. Richardson et al., 2001; Borovsky and Denton, 2006; Alves et al., 2006; Tsurutani et al., 2006; Zhang et al., 2008; Chi et al., 2018; Grandin et al., 2019). CIRs can drive radiation belt electron variability (e.g. Lam et al., 2009; Hudson et al., 2021), leading to electron flux enhancements (e.g. Blake et al., 1997; Li et al., 1997); electron acceleration especially in the presence of strong southward interplanetary magnetic field (IMF Bz) component (e.g. Blake et al., 1997), in which in certain cases reaching relativistic level (Paulikas and Blake, 1979; Pinto et al., 2018; Hudson et al., 2021). The high-speed streams (HSSs), originating from the coronal holes, that follow CIRs are rather efficient in producing multi-MeV electron enhancements (Home et al., 2018; Katsavrias et al., 2019). This is due to the fact that they produce intervals of combined southward IMF and solar wind velocity over 500 km/s, which lead to an enhanced magnetic reconnection rate at the dayside magnetopause (Borovsky and Denton, 2006; Miyoshi et al., 2013). A comprehensive review of the current physical understanding and effects of CIRs can be found in Cranmer et al. (2017), Kilpua et al. (2017), Vršnak et al. (2017), Richardson (2018) and Yermolaev et al. (2018).

Communication satellites in the geostationary orbit and medium and low Earth orbits are in the vicinity of the Earth's radiation belts—the regions encircling the near-Earth space and containing significant fluxes of high-energy electrons and ions. Since the beginning of the space technology era, there have been a number of reports of spacecraft anomalies and even failures, for example, at the geostationary orbit due to the elevated fluxes of several MeV electrons that persisted for several days following CIRs (Lanzerotti, 2007). To better mitigate the effects on those satellites, we need to develop an automated pipeline to predict the solar wind conditions. With the real-time monitoring of the solar wind at the Lagrangian L1 point, e.g. by the Advanced Composition Explorer (ACE), the continuous, real-time prediction of the radiation belt conditions via modeling is limited to 1 h in advance or less as the solar wind takes about 40–60 min from the ACE spacecraft to arrive at the magnetosphere. Using the Kp-index, a 3-h magnetic activity index indicative of geomagnetic perturbations at mid-latitude, the SPACECAST (Home et al., 2013) model, for instance, was used to provide a forecast up to 3 h in advance using ACE. To further improve our current capability, we need continuous solar wind prediction with a significant lead time.

Several efforts have been made in continuous solar wind predictions upstream of Earth. The majority of models consist of two main parts: (1) a solar wind emergence modeling at the base of the solar corona and (2) a solar wind propagation from the outer corona to Earth (and beyond). The former part is initiated by a reconstruction of the coronal magnetic fields based on synoptic or synchronic maps, which are composite

maps of the photospheric magnetic field observed over a solar rotation by ground- or space-based telescopes. Reconstructions are often achieved by employing a simplified analytical model with a magnetostatic potential field source surface (PFSS) extrapolation (Wang and Sheeley, 1992). The PFSS extrapolation combined with the Wang-Sheeley-Argé (WSA) model is a widely used predictor of the solar wind in the corona/low heliosphere (Argé and Pizzo, 2000; Argé et al., 2004). In combination with the global 3-D MHD “ENLIL”—the physics-based solar wind model [Odstrčil et al., 1996; Odstrčil and Pizzo, 1999; Odstrčil et al., 2004]—the WSA-ENLIL architecture has been used to provide daily forecasts of solar-wind streams. The WSA-ENLIL tool has also been extended to include CMEs using an ad-hoc hydrodynamic cone model or the spheromak model (Taktakishvili et al., 2009, 2011). The Space Weather Prediction Centre of the National Oceanic and Atmospheric Administration (SWPC/NOAA) has put the WSA-ENLIL into operation to provide daily solar wind predictions with a lead time of 1–4 days (Pizzo et al., 2011). Alternative to the WSA model is the Magnetohydrodynamic Algorithm outside a Sphere (MAS) which is a global, time-dependent model based on the resistive MHD equations (Linker et al., 1999; Riley et al., 2012). The WSA-ENLIL and the MAS-ENLIL tools have been extensively benchmarked by Owens et al. (2005), Lee et al. (2009), MacNeice (2009a, b), Jian et al. (2011a, b, 2015). For the solar-wind propagation part, several less numerically intensive options are also available, though they are not fully put into operation at the time of this writing. A few models include Heliospheric Upwind eXtrapolation (HUX) (Riley and Lionello, 2011), ballistic propagation (Dósa et al., 2018), and Tunable HUX (Reiss et al., 2020). MacNeice et al. (2018) and Reiss et al. (2019, 2020) compiled several combinations of the existing models and established frameworks for cross-validation. Additionally, other modeling frameworks also exist such as the SWMF (Space Weather Modeling Framework; Tóth et al., 2005; Merkin et al., 2016).

A growing effort has been made to combine and benchmark European-based models to develop operational solar-wind forecasting services. Multiple Flux-tube Solar Wind Model, alias “MULTI-VP”, is a coronal model developed by Pinto & Rouillard (2017) that computes multiple solutions of 1D solar wind flux tubes in sub-domains of interest. This approach allows a much faster calculation compared to global models with better accuracy at regions of interest, e.g. at the sub-Earth point, with a lead time of 1–3 days (Rouillard et al., 2020). MULTI-VP has been coupled with EUHFORIA (European Heliospheric Forecasting Information Asset; Pomoell and Poedts, 2018) and validated for HSS modeling in comparison with WSA-EUHFORIA (Samara et al., 2021). More recently, the HelioCast—a global MHD model comprising both coronal and heliospheric parts—has been developed by Réville et al. (2023) to provide a daily ambient solar wind forecast with a lead time of up to 5 days. The HelioCast model has been integrated within OFRAME (Organisation Française de Recherche Appliquative en Météorologie de l'Espace)¹, a national organization aiming to promote the space weather community and to bridge between the space weather research and potential users of space weather services in France. Moreover, an empirical

¹ French Organisation for Applied Research in Space Weather, see <http://www.meteo-espace.fr/oframe>.

solar wind forecast (ESWF; Vršnak et al. 2007; Milošić et al., 2023) model for predicting the temporal evolution of the solar wind speed with a lead time of 4 days has been developed using an empirical relation found between the areas of coronal holes as observed in extreme ultraviolet data and the solar wind speed at 1 AU (Robbins et al. 2006; Vršnak et al. 2007). This model has been put into operation by the European Space Agency.

In this work, we consider the coupling of MULTI-VP and a 1D MHD model (Tao et al., 2005) for the first time to develop an automated prototype pipeline named “Helio1D” for predicting the temporal evolution of the solar wind (including CIRs and HSSs). The 1D MHD model was originally developed for modeling solar wind conditions at Jupiter using the in situ observations at Earth as the inner boundary; it was later extended to provide operational service for solar wind conditions at other solar-system planets (André et al., 2018), see HelioPropa.² In essence, we develop a prototype pipeline for solar wind modeling upstream of Earth for physical parameters including the solar wind speed, tangential magnetic field, plasma density, and temperature, as well as their uncertainties. Here, CMEs are not included as they need a dedicated ad-hoc approach. The prototype pipeline is integrated into a prototype service for the operational forecasting of radiation belt environmental indicators for the safety of space assets as a part of the European Union Horizon 2020 SafeSpace³ project.

The evaluation or validation of modeling results is crucial to investigate model performance. Several works considered standard metrics such as root mean square, mean square error, and Pearson correlation coefficient to compare modeling results to the observations. With such metrics, a skill score is also computed to assess whether the model performs better than a baseline model. For CIR/HSS and CME modeling, some works also considered event-based verification that allow the characterization of a true positive (predicted and observed), a false positive (predicted but not observed), a false negative (not predicted but observed) to construct contingency tables (Woodcock, 1976) and compute the probability of detection (e.g. Reiss et al., 2016). More recently, a complementary approach using the Dynamic Time Warping (DTW) algorithm, a technique commonly employed in time-series analysis, was applied to characterize the performance of solar wind modeling (Samara et al., 2022). We consider these metrics and approaches for verification of the prototype pipeline’s results. Particularly, we limit our focus to the validation of the continuous solar-wind stream modeling without considering the event-based verifications of CIRs/HSSs at this stage.

The paper is organized as follows. Firstly, we introduce MULTI-VP and 1D MHD models and then describe the interfacing between them for the Helio1D prototype pipeline. Secondly, we describe metrics and techniques for quantifying the performance of Helio1D. Thirdly, we provide results from the prototype pipeline with the assessments of the results using the classic metrics. Fourthly, we discuss the prototype pipeline calibration with the DTW algorithm. Finally, we discuss our prototype pipeline advantages and disadvantages, as well as future prospects for operational forecasting, and provide conclusions.

2 Helio1D prototype pipeline

Here, we briefly introduce MULTI-VP and associated data and the 1D MHD model. We then describe the interfacing of the two models for the automated Helio1D prototype pipeline.

2.1 MULTI-VP

MULTI-VP (Pinto and Rouillard, 2017) is a coronal model that covers the heliocentric distances over which all solar wind streams are formed and accelerated: between 1 and about 30 solar radii ($R_{\odot}$). The main advantages are that it is data-driven and computationally fast, while taking into account the thermodynamics of the wind flows across the highly stratified low solar atmosphere, and hence it produces physically correct and realistic estimations of the state of the solar wind across its domain. MULTI-VP determines a full set of physical quantities consisting of the solar-wind speed (V), density (n), temperature (T), magnetic field ($\mathbf{B}$), and secondary quantities without requiring empirical scaling laws. In addition, MULTI-VP actually computes a set of individual wind streams from surface to high corona that can then be put together to build the full three-dimensional solar wind solutions from multiple 1D solar-wind flux tubes. This brings an advantage with respect to more traditional 3D MHD models in terms of required computing time, and also by not being subject to the strong diffusion of the gradients in the transverse directions. In practice, it allows us to compute solar wind solutions restricted to certain regions of interest.

For the 1D MHD model, MULTI-VP is required to produce a time series of the solar wind properties at the sub-Earth point at 0.14 AU (at $30 R_{\odot}$) along the Sun-Earth line. Internally, this output is defined via a data object that lists which magnetic field lines (and hence individual wind streams) are sampled, as well as their coordinates and geometry, and properties of the input magnetogram. Each time a forecast is performed, we select a number of positions around the sub-Earth point (i.e. at 0.14 AU in the Sun-Earth line) stretching up to 15 degrees in azimuth and 15 degrees in latitude. The spread in latitude is aimed to cover positional errors propagated from the magnetic field reconstruction. This approach is similar to Owens & Riley (2017), and references therein, where the near-sun solar wind speed was sampled at a range of latitudes around the sub-Earth point to provide the uncertainty quantification. Additionally, the spread in longitude is aimed to cover the azimuthal range that corresponds to an elemental time-series three days of solar rotation with respect to Earth, which translates to one day behind and two days ahead of the modeling at sub-Earth point. The next forecast proceeds in the same way, producing another elemental time series that partially overlaps the previous in time. This procedure is repeated to continuously build a long-term time-series that can be used as an input to the 1D MHD model.

For the simulations discussed in this work, MULTI-VP was set up to run as a sequence of scalar computational jobs managed by a Message Passing Interface master job. The duration of a single (elemental) run depends on the specific properties of the corresponding individual solar wind stream. The typical runtime of the daily solar wind simulations consisting of 21 solar wind elements (see Section 6) ranges between 2 and 3 h in a few-core CPUs.

² <http://heliopropa.irap.omp.eu/>.

³ <https://www.safespace-h2020.eu/>.

To evaluate the prototype pipeline performance with regard to various phases of the solar cycle, we need sufficiently long datasets. The longest, continuous historical magnetogram series available is from the Wilcox Solar Observatory (WSO; <http://wso.stanford.edu/synopticl.html>). Here, we obtain mainly two continuous datasets from MULTI-VP processed using the WSO magnetograms. The first set extends from Carrington Rotation (CR) 2024 to CR 2139, corresponding to data from December 2004 to August 2013. This 8-year and 7-month-long interval covers the declining phase (2004–2008) and the solar minimum (2009–2010) of solar cycle 23 as well as the ascending phase (2011–2013) of solar cycle 24. The second set extends from CR 2192 to CR 2210 and corresponds to data from June 2017 to November 2018. These two MULTI-VP datasets for the two distinct periods are fed into the 1D MHD model. These data are sufficiently long for the prototype pipeline validation as they cover most phases of the solar cycle. Further evaluation of the pipeline performance with more data should be considered in future work.

2.2 1D MHD model

From 0.14 AU ($30 R_{\odot}$) onwards, the solar wind is propagated through the heliosphere using a 1D MHD model (Tao et al., 2005) that takes the solar wind time-series as time-varying boundary conditions, the solar wind parameters—mainly the radial plasma velocity and tangential velocity—vary as a function of time (see more below). The 1D MHD model propagates the solar wind using an ideal MHD fluid approximation to the target position while taking into account the interaction between fast and slow streams in the radial direction. The code solves the ideal MHD equations under the influence of solar gravity in a one-dimensional spherically symmetric coordinate system. The 1D MHD equations are solved using the Coordinated Astronomical Numerical Software (CANS).⁴ The 1D MHD code was developed by Tao et al. (2005) to propagate the solar wind observations at Earth to upstream of Jupiter and was further extended by the French plasma physics data center⁵ to cover other solar-system planets or target spacecraft positions (André et al. 2018). The code is robust and widely used for solar wind propagation and CIR modeling in the heliosphere (e.g. Palmerio et al., 2021; Nilsson et al., 2022).

The MULTI-VP outputs at $30 R_{\odot}$ were used as boundary conditions for the initiation of the 1D-MHD code, which then propagated these results up to the L1 point at 1 AU. The coordinate system in the 1D MHD code is equivalent to the Radial-Tangential-Normal (RTN) system, where the X -axis (R) is pointing radially outward from the Sun in the equatorial plane to the Earth, the Z -axis (N) is the solar rotation axis, and the Y -axis (T) completes the orthonormal system. The outer boundary is set to 1.4 AU where the derivatives of all physical parameters diminish. The grid spacing and the time step are chosen to be $(1.4 - 0.14)/400$ AU and 150 s, respectively. To adopt the 1D simulation, we keep the tangential magnetic field B_y component to physical values while the B_x is fixed to 0.001 nT to meet the solenoidal criterion and the B_z is set to

zero (see Tao et al., 2005). We retrieve MULTI-VP data at a 30-min cadence and resample them to a 1-h cadence using linear interpolation; these hourly-averaged data are then linearly interpolated to meet the CFL (Courant-Friedrichs-Lewy) condition for every time step. In terms of computational resources, it takes less than a few minutes to run a 1-month long of solar wind sequence with a single-core CPU.

2.3 Helio1D – Interfacing MULTI-VP and 1D MHD

To automate the interfacing between the two models, several steps are taken. MULTI-VP provides solar wind time series with a full set of physical parameters including plasma number density, velocity, temperature, and magnetic field in the RTN coordinates. We first format the time series to fit the input format of the 1D MHD model. Here, we take the input at a 1-h time resolution and output the propagated time series at L1 with the same 1-h resolution. To successfully run the 1D MHD code, two criteria are required to be met: (1) the data length must cover at least a solar rotation, i.e., 27.25 days, in order to produce a consistent CIR formation, and (2) the parameter values must be physical (e.g. non-negative plasma number density). In certain conditions, MULTI-VP may provide unphysical solar wind values due to the poor quality of the magnetograms. We apply a set of criteria to remove unphysical values (see Table A1 in Appendix) and fill gaps using linear interpolation between two available data points. These aspects are further described in Section 6.

We set up the Helio1D prototype pipeline with two modes for running (a) long-term historical data consisting of several Carrington rotations and (b) short-term data of 1-month long for daily, operational forecasting as shown in Figure 1. These two modes are set to provide their respective time-series outputs in ASCII or CDF data formats. The mode-(a) allows us to evaluate the performance of the prototype pipeline as well as identify calibrations that may improve its performance for future operational forecasting. Technically, the interfacing between MULTI-VP and the 1D MHD code has been done for the prototype pipeline such that it can perform solar wind prediction daily with minimal manual handling. Specifically, all the processes (Fig. 1) including fetching the data, filling the data gaps, correcting unphysical values, and modeling the coronal wind and solar wind propagation have been programmed where they can be launched with a single command line through a batch script. We first benchmark the prototype pipeline using long-term MULTI-VP data in Section 4.

3 Measuring the performance of Helio1D

In this section, we first introduce the metrics and methodology for characterizing the prototype pipeline performance, consisting of the standard metrics and the Dynamic Time Warping technique. We then provide a description of the observed data and the recurrence and persistence models.

3.1 Standard metrics

We choose three standard metrics – root mean square error (RMSE), mean square error (MAE), and Pearson correlation coefficient (Pcc) – to measure the performance of the prototype

⁴ Developed by T. Yokoyama at the University of Tokyo and collaborators.

⁵ Centre de Données de Physique des Plasmas, France. See <http://heliopropa.irap.omp.eu/>.

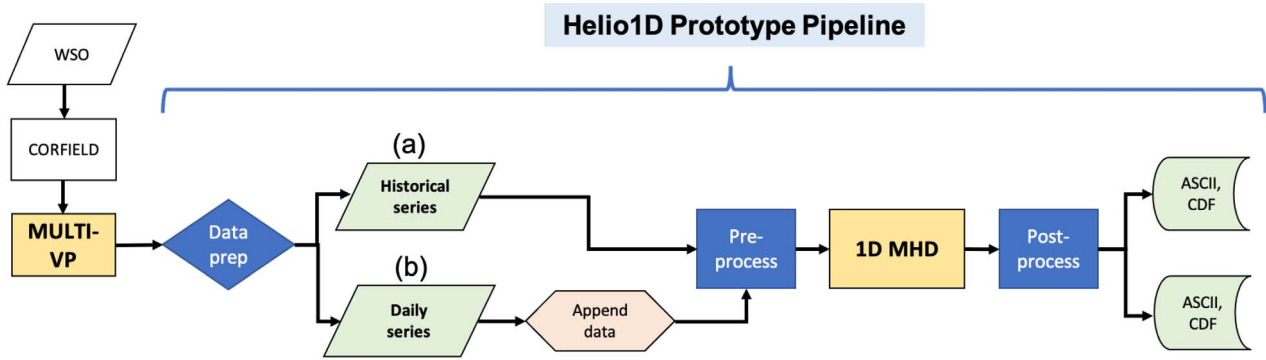


Figure 1. The Helio1D prototype pipeline: interfacing between MULTI-VP and the 1D MHD code. From left to right, MULTI-VP takes WSO synoptic magnetograms where the coronal field reconstruction (CORFIELD) was built upon as inputs. MULTI-VP then generates sequences containing the temporal profiles of solar wind parameters at 0.14 AU along the sub-Earth line (labeled as Data prep). Two types of data are generated: (a) historical series and (b) daily series. The data from (b) must be appended first to create a long-time series required by the 1D MHD code. Next, the pre-process step checks validity of the physical values and replaces unphysical values, if necessary. The 1D MHD code then propagates the near-sun solar wind to 1 AU. The post-process step finally produces final outputs as ASCII and CDF files for historical series or daily series.

pipeline. RMSE is a classic point-by-point metric that compares the modeled value and observed value for a given point. It is defined as

$$\text{RMSE} = \sqrt{\frac{1}{T} \sum_{t=1}^T [V_m(t) - V_o(t)]^2}, \quad (1)$$

where $V_m(t)$ and $V_o(t)$ are the modeled and observed values, respectively, at a given time t while T is the number of time elements. MAE is an alternative metric that measures a similar property. It is defined as

$$\text{MAE} = \frac{1}{T} \sum_{t=1}^T |V_m(t) - V_o(t)|, \quad (2)$$

where the symbols have the same meanings as defined for the RMSE. The units of the RMSE and MAE are the same as the units of parameter values $V_m(t)$ and $V_o(t)$. Smaller values of RMSE and MAE indicate better model prediction when compared to observations. Zero values of both metrics indicate a perfect model prediction.

We also introduce the Pcc to roughly quantify the similarities between the two-time series. In principle, Pcc measures the linear relationship between two datasets. Using the observed and modeled time series ($V_o(t)$ and $V_m(t)$, respectively), the formula for Pcc is

$$\text{Pcc} = \frac{\sum_{t=1}^T (V_m(t) - \langle V_m(t) \rangle) (V_o(t) - \langle V_o(t) \rangle)}{\sqrt{\sum_{t=1}^T (V_m(t) - \langle V_m(t) \rangle)^2} \sqrt{\sum_{t=1}^T (V_o(t) - \langle V_o(t) \rangle)^2}}. \quad (3)$$

The Pcc-value ranges from -1.0 to 1.0 , with a value of 1.0 being a perfect positive linear correlation between the two-time series and -1.0 being a perfect negative linear correlation between the two.

Additionally, to evaluate the model performance in a more concrete way using the above metrics, we introduce the skill of forecast as

$$\text{Skill} = 1 - \frac{\text{MSE}_{\text{pred}}}{\text{MSE}_{\text{ref}}}, \quad (4)$$

where MSE_{pred} is the mean squared error of the prediction compared to the observation and MSE_{ref} is the mean squared error of the prediction compared to a reference baseline. We describe data for the baseline in [Section 3.3](#).

3.2 Dynamic time warping technique

Although the classic metrics—RMSE, MAE, Skill, and Pcc—provide quantification of the performance of the model, they do not provide in-depth information such as time lags, i.e., the difference in arrival times between the modeled and observed stream interfaces, that are crucial for model evaluation. For this reason, we also consider an alternative technique called Dynamic Time Warping (DTW). The DTW technique compares two sequences by finding an optimal alignment by which one sequence may be stretched or compressed (hence, “warped”) in the temporal domain to match the other under certain constraints (e.g., [Müller, 2007](#)). The technique was first introduced by [Bellman & Kalaba \(1959\)](#) in adaptive control processes and later found several applications notably in speech and pattern recognition (e.g., [Sakoe and Chiba, 1978](#); [Myers et al., 1980](#)). More recently, DTW has been applied in space weather in particular for comparing modeled geomagnetic indices and solar wind data to observations (e.g., [Laperre et al. 2020](#); [Owens and Nichols, 2021](#); [Bunting and Morgan, 2022](#); [Nilsson et al., 2022](#); [Samara et al. 2022](#)). We first introduce its formulation and then describe the methodology for our application.

The DTW technique finds the optimal alignment between two sequences, e.g., $X = \{x_1, x_2, \dots, x_{|X|}\}$ and $Y = \{y_1, y_2, \dots, y_{|Y|}\}$, through the DTW cost matrix by choosing the path that minimizes the total cumulative cost compared to all other paths ([Keogh and Pazzani, 2001](#); [Ratanamahatana and Keogh, 2004](#); [Müller, 2007](#)). This is called the warping path, which

characterizes the mapping between the two sequences. The DTW cost matrix is defined as:

$$D(i, j) = \delta(i, j) + \min\{D(i-1, j-1), D(i-1, j), D(i, j-1)\}, \quad (5)$$

where i, j are the indices of X and Y , respectively, and $\delta(i, j) = |x_i - y_j|$ is the Euclidean distance between the element x_i of series X and the element y_j of series Y . Here, the optimal warping path can be found via back-tracing in the DTW cost matrix (e.g., Berndt and Clifford, 1994; Keogh and Pazzani, 2001; Górecki and Łuczak, 2013). A main disadvantage of DTW is that an element x_i may be mapped to several elements of Y —this problem is called “pathological mapping” or “singularities” (e.g., Sakoe and Chiba, 1978). To alleviate this issue, certain constraints have been added such as the so-called windowing, slope weighting, and step patterns (see Keogh and Pazzani, 2001, and references therein). Various variations of the DTW technique have also been developed to optimize the alignments between the sequence points (e.g. Keogh and Pazzani, 2001; Chu et al., 2002; Keogh and Ratanamahatana, 2005; Salvador and Chan, 2007; Efrat et al., 2007; Furtună, 2008; Zhu et al., 2012; Yadav and Alam, 2018, and references therein).

Choosing an appropriate DTW technique is crucial as different algorithms can lead to different results. Here, we choose the FastDTW algorithm developed by Salvador & Chan (2007) due to its multi-level approach. First, the time series are sampled down to a very coarse resolution. A warping path is found in this coarse resolution and projected onto a higher resolution. The warping path is then refined and projected again to a higher resolution. The process of projecting and refining is repeated until a warping path is found for the original resolution time series. Most importantly, since the technique initially finds a warping path in the coarse resolution, it is particularly appropriate to capture large-scale features in the time series, notably “CIRs” with large speed gradients.

3.3 Observation data, recurrence, and persistence models

The observations of solar wind and IMF data at L1 are taken from the High Resolution OMNIWeb database (King and Papitashvili, 2005). Here we use the hourly merged data products, processed from in situ measurements made by satellites including ACE at L1. These solar wind and IMF monitoring data have already shifted to the Earth’s bow shock nose. To evaluate the performance of Helio1D, we compare it to two reference baseline models as follows. First, the observations from the previous Carrington rotation, i.e., 27 days before, are taken as a ‘recurrence’ model. This recurrence model thus assumes that the solar wind structures do not change with respect to the previous solar rotation; this baseline has been used to assess the performance of physics-based numerical models (Owens et al., 2013). Second, the observations from 4 days before are taken as a ‘persistence’ model. Both baseline models were used as reference baselines in Reiss et al. (2016). We will measure the RMSE, MAE, Pcc, and skill score against the 27-day recurrence and 4-day persistence models, in addition to the real observations, in Section 4.1.

4 Results

4.1 Performance of Helio1D using long-term data

Figure 2 shows a comparison between the Helio1D solar wind speed at 1 AU (magenta) and the observations (black). The Helio1D data are shown for all the available intervals, from January 2005 to July 2013 in panels (a)–(i) and then from July 2017 to November 2018 in panels (j)–(k). From these long-term results, we find that the large-scale solar wind fluctuations including the modulations of CIRs are correctly produced by Helio1D. For instance, the amplitudes of CIRs, e.g. in 2006 (Fig. 2b) and 2009 (Fig. 2c), are mostly reproduced correctly. On a shorter timescale, however, there are some mismatches in the amplitudes of the fluctuations. In particular, peaks in the velocity are often missed, e.g. in 2010 (Fig. 2f). This is partly because we do not have CMEs in our model. When the CME occurrence is low, e.g. between July and December 2007 (Fig. 2c), our model produces repeated CIR and HSS structures similar to those of the observations although the timing and amplitude often do not match. We will quantify these differences in detail in Section 4.2.

To quantitatively assess the model performance, we split these data into individual 6-month intervals, except for the last interval from the latter half of 2018 which covers about 5 months. Table 1 lists these intervals and shows the classic metrics described in Section 3.1. In particular, these metrics are shown in comparison with the observed data (labeled as OMNI), the 27-day recurrence (labeled as 27d), and the 4-day persistence (labeled as 4d) models. We find that the RMSE and MAE values between the Helio1D (labeled as 1D) and OMNI are worse than those with the 27-day recurrence model. In comparison to the 4-day persistence model, some of the RMSE and MAE values for Helio1D are better, especially between 2007 and 2008. The Pcc values between the Helio1D and OMNI show rather low values, indicating that there is no linear relationship between the two data unlike those from the 27-day recurrence model. In comparison to the 27-day recurrence model, the Helio1D skill values show all negative values, indicating that it performs worse than the 27-day recurrence model. Nevertheless, in comparison with the 4-day persistence model, it shows some positive values; this indicates that Helio1D performs better than the 4-day persistence model in certain cases.

Among all the intervals in Table 1, we find that the intervals between July 2008 and June 2010 have relatively low RMSE and MAE values, with the interval of July–Dec 2009 displaying the lowest values for all three metrics. This interval corresponds to Figure 2e for the latter half. Here, the CIR and HSS structures modeled by Helio1D reproduce the large-scale structure of the solar wind, despite some dissimilarities in timing and differences in the smaller-scale structures. To further investigate the Helio1D performance quantitatively, we next consider these metrics over various phases of the solar cycle as expressed by the sunspot number.

Figure 3 shows the performance of Helio1D characterized by the RMSE and MAE for the intervals shown in Table 1 (left vertical-axis, black) compared to the average sunspot numbers of the corresponding intervals (right vertical-axis, red). The sunspot number data were obtained as the monthly mean total

Comparison between solar wind speeds [Vx in km/s] from Helio1D (magenta) and OMNI (black) data for all intervals

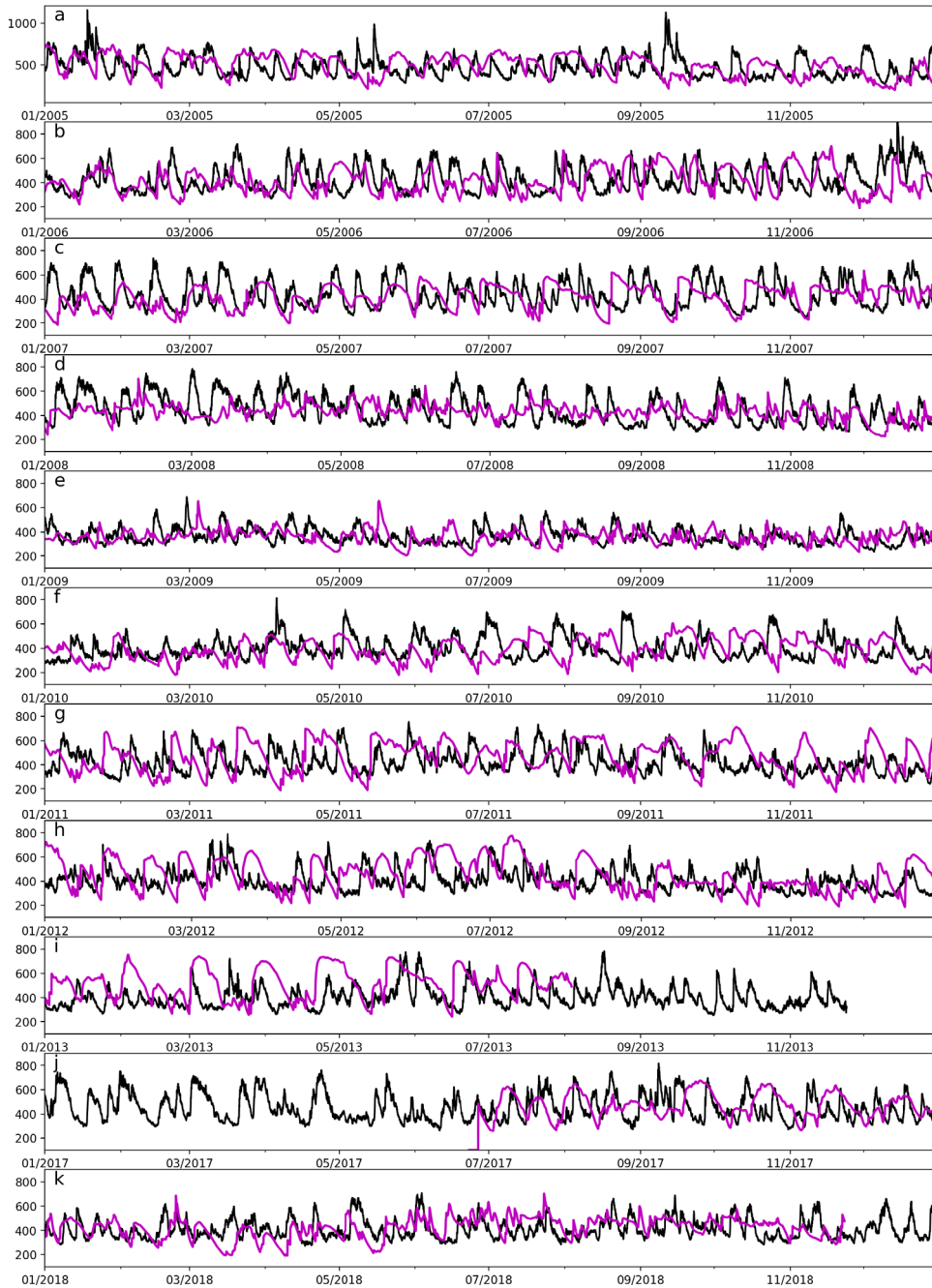


Figure 2. Comparison between the Helio1D solar wind speed (magenta) and the observations (black) for all the intervals. Panels (a)–(k) correspond to the intervals in Table 1 where the Helio1D data are available, from January 2005 to June 2013 and from July 2017 to November 2018.

sunspot number from the Royal Observatory of Belgium.⁶ The data in 2005–2013 correspond to the solar cycle 23 and the data in 2017–2018 correspond to the solar cycle 24. Here, we have more complete data for solar cycle 23 than the solar cycle 24. Based on the average sunspot numbers, we infer that our data for the solar cycle 23 comprise the declining phase from 2005

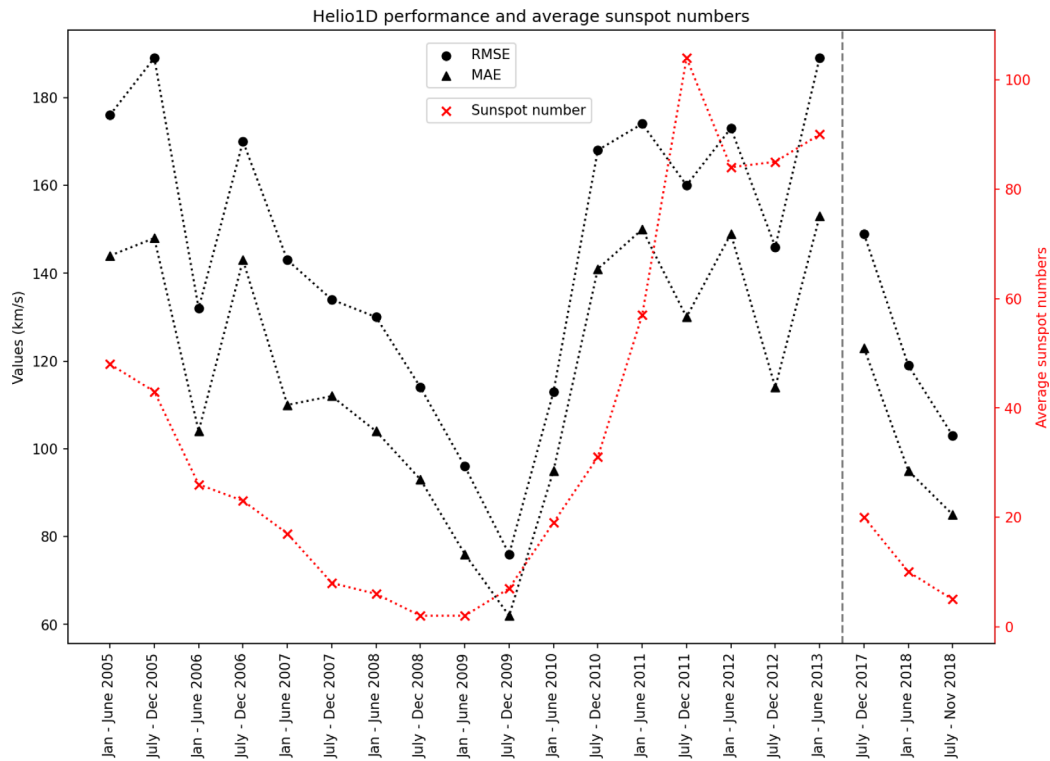
to 2007, the solar minimum from 2008 to mid-2009, the ascending phase from mid-2009 to 2011, and the solar maximum from 2012 to 2013. The rest of our data from mid-2017 to late 2018 correspond to the declining phase of the solar cycle 24.

We find that all the metric values (shown in black) vary with the average sunspot numbers (shown in red). During the declining phase of the solar cycle 23, all the metric values proportionally decrease with the average sunspot numbers. During the ascending phase of the solar cycle 23, all the metric values

⁶ Source: WDC-SILSO, Royal Observatory of Belgium, Brussels (<https://www.sidc.be/silso/datafiles>).

Table 1. List of the Helio1D long-series intervals and the point-by-point metrics calculated using the Helio1D series in comparison with OMNI data, the 27-day recurrence model (27d), and the 4-day persistence model (4d) for the same intervals.

Metrics	RMSE	RMSE	RMSE	MAE	MAE	MAE	Pcc	Pcc	Pcc	Skill	Skill
	(1D, OMNI)	(27d)	(4d)	(1D, OMNI)	(27d)	(4d)	(1D, OMNI)	(27d)	(4d)	(1D, 27d)	(1D, 4d)
Intervals/Units	km s ⁻¹	km s ⁻¹	km s ⁻¹	km s ⁻¹	km s ⁻¹	km s ⁻¹	–	–	–	–	–
Jan–Jun 2005	176	154	180	144	107	153	0.00	0.20	−0.08	−0.31	0.05
Jul–Dec 2005	189	117	181	148	81	144	−0.17	0.55	−0.08	−1.61	−0.08
Jan–Jun 2006	132	83	141	104	61	112	0.03	0.67	0.08	−1.52	0.13
Jul–Dec 2006	170	104	151	143	74	126	−0.14	0.51	0.05	−1.70	−0.27
Jan–Jun 2007	143	107	183	110	74	153	0.20	0.63	−0.26	−0.79	0.39
Jul–Dec 2007	134	85	149	112	62	122	0.17	0.70	0.12	−1.48	0.18
Jan–June 2008	130	91	162	104	69	133	0.04	0.66	−0.09	−1.03	0.36
Jul–Dec 2008	114	70	160	93	50	129	0.16	0.80	−0.16	−1.60	0.49
Jan–Jun 2009	96	67	94	76	50	72	0.01	0.51	0.02	−1.01	−0.03
Jul–Dec 2009	76	67	87	62	52	68	0.10	0.43	0.03	−0.29	0.24
Jan–Jun 2010	113	84	108	95	63	84	0.20	0.49	0.12	−0.83	−0.11
Jul–Dec 2010	168	110	126	141	84	96	−0.45	0.38	0.19	−1.35	−0.78
Jan–Jun 2011	174	112	138	150	87	109	−0.14	0.30	−0.09	−1.42	−0.58
Jul–Dec 2011	160	110	117	130	83	91	0.06	0.18	0.07	−1.09	−0.87
Jan–Jun 2012	173	113	127	149	81	93	−0.01	0.22	0.03	−1.35	−0.85
Jul–Dec 2012	146	100	98	114	76	78	0.01	0.28	0.25	−1.13	−1.22
Jan–Jun 2013	189	98	125	153	72	95	0.14	0.43	0.08	−2.76	−1.29
Jul–Dec 2017	149	95	157	123	71	130	−0.04	0.60	−0.12	−1.45	0.11
Jan–Jun 2018	119	88	120	95	67	91	0.10	0.44	−0.06	−0.83	0.02
Jul–Nov 2018	103	83	110	85	63	88	0.12	0.44	0.04	−0.56	0.12

**Figure 3.** The average RMSE (dot) and MAE (triangle) for a 6-month interval. The values of the three metrics are shown on the left Y-axis (black), in comparison to the average sunspot numbers (cross) on the right Y-axis (red) for the same intervals. The data for 2005–2013 (solar cycle 23) and 2017–2018 (solar cycle 24) are separated by a vertical grey dashed line. The individual metric values for each solar cycle are connected by dotted lines.

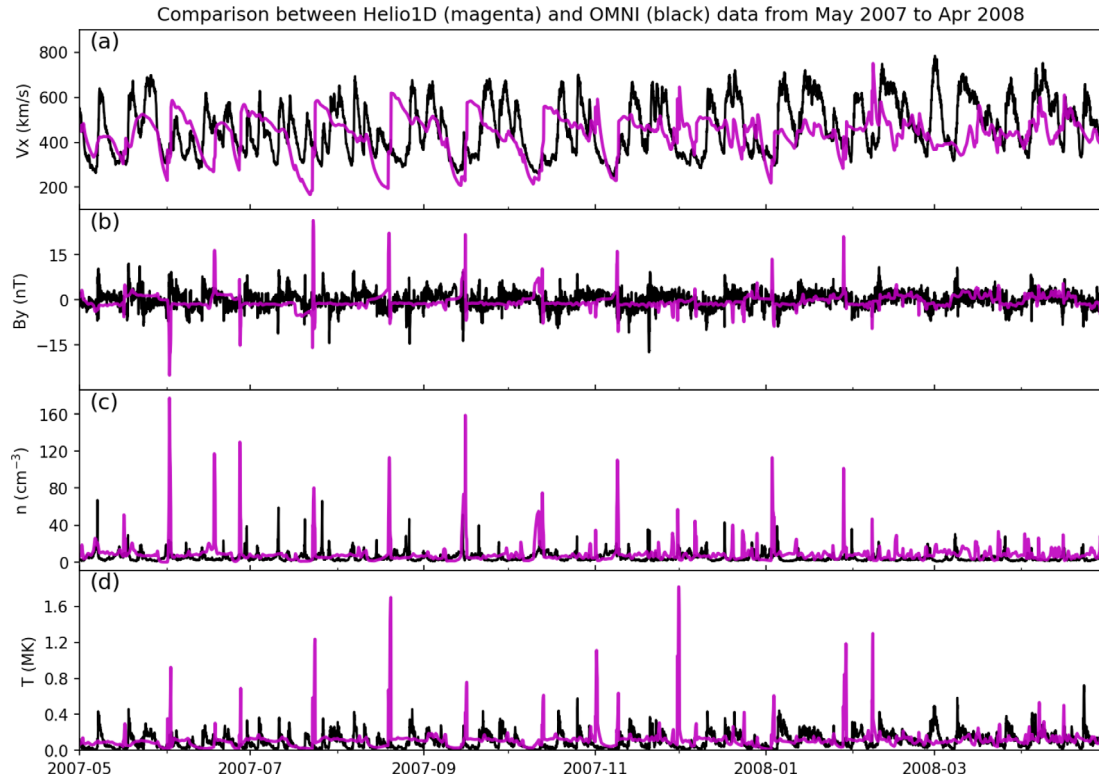


Figure 4. Comparison between temporal variations of the solar wind parameters from Helio1D (magenta) and the hourly in situ measurements (black) taken from OMNI between May 2007 and April 2008—(a) radial bulk flow velocity (V_x); (b) tangential magnetic field component (B_y); (c) ion number density (n); and (d) ion temperature (T).

roughly increase with the average sunspot numbers. The most striking features are during the solar minimum and the solar maximum. The metric values reach their lowest values during the solar minimum and/or the late declining phase and early ascending phase. Here we find the minimum RMSE and MAE in Jan–June 2009, i.e., six months after the lowest average sunspot number. The highest errors are found during the solar maximum, which is not surprising as the solar magnetic fields often undergo changes and they are thus less predictable. During the declining phase of the solar cycle 24, we find that the errors proportionally decrease following the average sunspot numbers. These results demonstrate that the performance of the Helio1D varies with the solar cycle—it performs best during the solar minimum and it performs worst during the solar maximum. This finding is consistent with the performance of solar wind modeling by other existing models (e.g. [Hinterreiter et al., 2019](#)). We further discuss these results in [Section 7](#).

To understand how Helio1D qualitatively performs, we next visualize an example of the model output against the observations at L1. [Figure 4](#) shows the Helio1D output in magenta and the OMNI data in black. In [Figure 4a](#), the bulk flow velocity displays a recurring pattern of the stream interfaces of CIRs characterized by the transition from slow to fast wind, clearly noticeable between August and November 2007. The stream interfaces also coincide with the polarity change in B_y , shown in [Figure 4b](#), and the enhancement of density and temperature in [Figure 4c](#) and [Figure 4d](#), respectively. Comparing to the observations, we find that the modeled solar wind speed agrees qualitatively well and that the model correctly produces consistent CIR formation.

Nevertheless, the number density and temperature in [Figure 4c](#) and [Figure 4d](#) at the stream interfaces are generally higher than the observed values. These higher peaks are produced as a consequence of over-compression at the stream interfaces due to the ideal MHD plasma assumption in the 1D MHD code, which limits dissipation. Despite some over-compression and temporal uncertainties, we conclude that the interfacing between MULTI-VP and 1D MHD works reasonably well. We further perform qualitative assessments of the prototype pipeline performance and discuss calibration in [Section 5](#).

4.2 Performance of Helio1D with the FastDTW technique

Since the classic metrics only indicate crude information about the model performance when compared to real data, we consider applying the FastDTW algorithm to complement our analysis. [Figure 5a](#) shows an example of the FastDTW mapping of the OMNI data to Helio1D for the velocity during 27 December 2005 and 14 February 2006. The OMNI data (black) show two clear stream interfaces and two corresponding HSSs. It can be seen that the Helio1D misses the first HSS while correctly produces the second stream interface along with the adjacent HSS compared to the observation despite some time delay and differences in the finer-scale structures. Qualitatively, [Figure 5a](#) shows the FastDTW alignments (cyan) that map large-scale features comprising the stream interfaces and HSS. Here, the alignments can be used for extracting detailed information indicative of the model performance (e.g., if a modeled

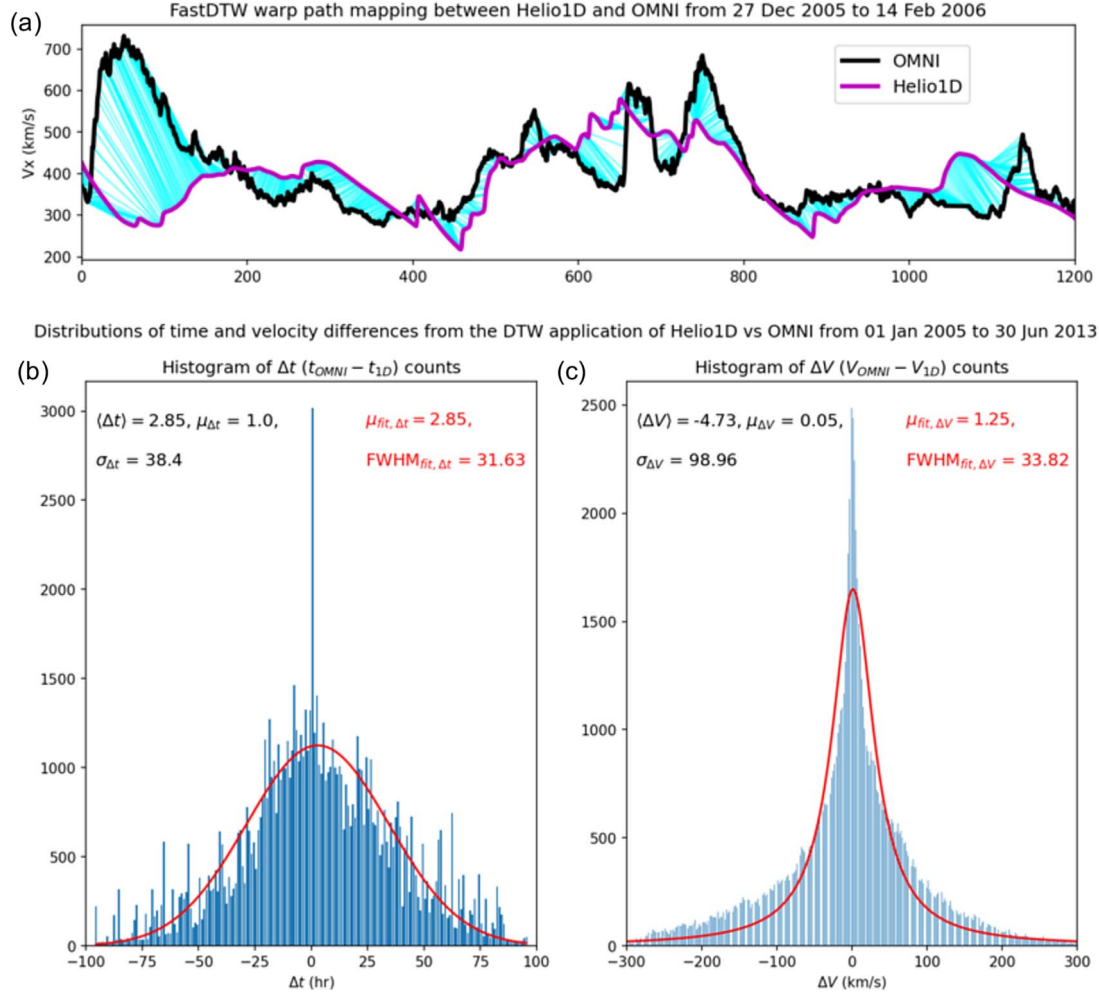


Figure 5. Example of the FastDTW applications. (a) The FastDTW alignments (cyan) between the OMNI (black) and Helio1D (magenta) sequences for the solar wind speed, from 27 December 2005 to 14 February 2006. Histograms of (b) time difference (Δt) and (c) velocity difference (ΔV) obtained from the FastDTW alignments from January 2005 to June 2013 with the corresponding statistics (see text). The Gaussian (b) and the Cauchy-Lorentzian (c) distribution fits (red solid lines) are added for comparisons with the best-fit parameters in red applied for the Δt and ΔV distributions, respectively.

stream interface arrives before or after, and if the modeled speed is higher or lower compared to observations). The statistical information from the alignments is useful for investigation of the model performance in terms of time and magnitude differences. For a given alignment, we may define the time difference as $\Delta t = t_{OMNI} - t_{1D}$ and the velocity difference as $\Delta V = V_{OMNI} - V_{1D}$, where the subscript OMNI refers to the observations and the subscript 1D refers to the Helio1D results. For each alignment, $\Delta t > 0$ ($\Delta t < 0$) means the Helio1D sequence arrives earlier (later) than the observation. Similarly, $\Delta V > 0$ ($\Delta V < 0$) means the Helio1D sequence underestimates (overestimates) the observed speed. We evaluate the Δt and ΔV statistically by considering their distributions next.

Figure 5b and 5c show histograms of the Δt and ΔV based on the FastDTW alignments on the velocity time series from January 1, 2005, to June 30, 2013. We note that the Δt and ΔV presented in the histograms are not only relevant to the time and amplitude differences of the points that are aligned uniquely between them but to the singularity points as well (see Section 3.2). In Figure 5b, the mean Δt ($\langle \Delta t \rangle$) is 2.85 h, while

the median ($\mu_{\Delta t}$) is 1 h, which is rather low. The standard deviation ($\sigma_{\Delta t}$) is 38.4 h. For a comparison, the best-fit Gaussian distribution (red solid line) is added with $\mu_{fit, \Delta t}$ of 2.85 h and the full width at half maximum (FWHM) of 31.63 h. The true distribution of Δt is not strictly Gaussian as can be seen visually. In particular, there are several spikes of Δt counts, especially near zero. This is likely because the FastDTW technique maps one point or one section (e.g., a peak or a stream interface) of the first sequence to several neighboring points in the second sequence (see Fig. 5a), leading to a repetition of similar Δt . Using the direct statistics, we can conclude that the Δt is biased towards positive values up to a few hours, with the standard deviation of 38.4 h. In other words, the velocity structures of Helio1D are on average ahead of those of the observations, as can also be seen qualitatively in Figs. 2 and 4.

Figure 5c shows the histogram of ΔV . The mean ΔV is -4.73 km/s while the median ΔV is 0.05 km/s, which is quite low. The standard deviation is rather large with $\sigma_{\Delta V} = 98.96$ km/s. In addition, we add the best-fit Cauchy-Lorentzian distribution (red solid line) to the ΔV distribution to extract

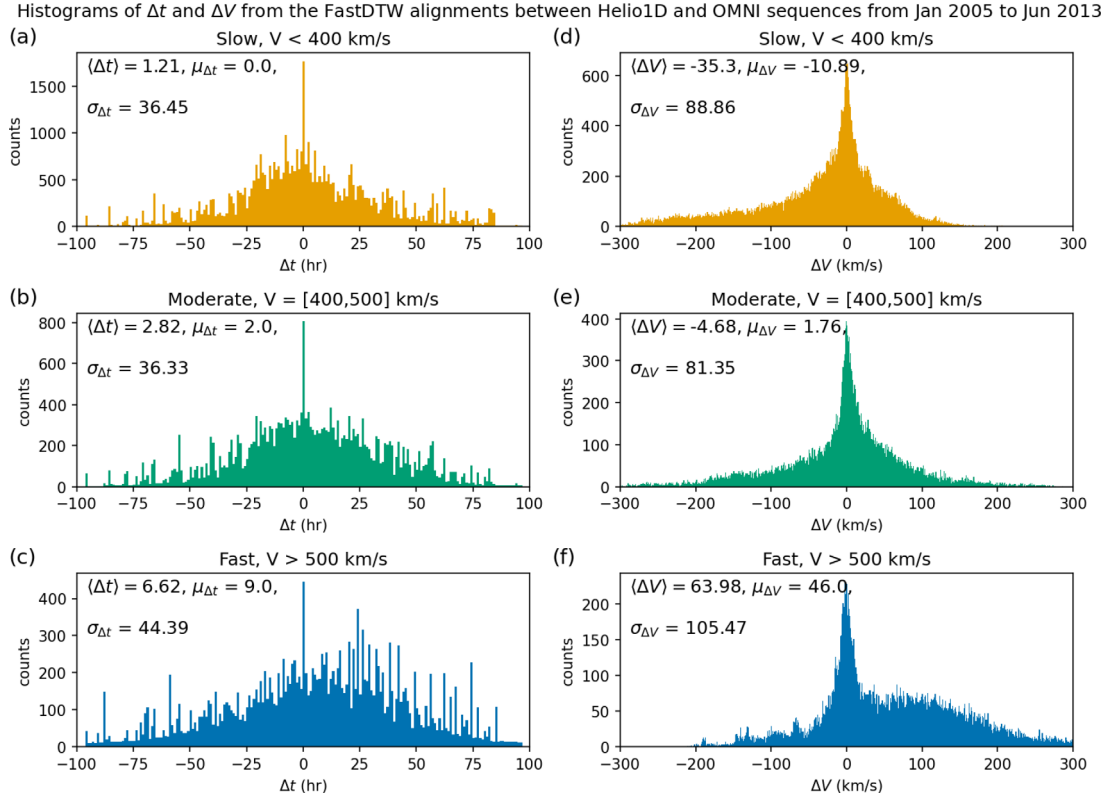


Figure 6. Histograms of (a–c) time difference and (d–f) velocity difference obtained from the FastDTW alignments between Helio1D and OMNI sequences between January 2005 and June 2013 for the (a, d) slow, (b, e) moderate, and (c, f) fast winds, along with their statistics.

their statistical properties to compare with the direct statistics. The best fit is found with $\mu_{fit} = 1.25$ km/s and FWHM of 33.82 km/s. We note that the Cauchy-Lorentzian is a symmetric distribution (around the median). Visually, we can see that the actual ΔV distribution is biased towards negative values. In other words, Helio1D on average produces the solar-wind speed that slightly overestimates the observed speed, with $\langle \Delta V \rangle = -4.73$ and a large $\sigma_{\Delta V}$ of 98.96 km/s.

Using the FastDTW alignment information, we also explore the model performance for different types of solar wind. Specifically, we categorize the solar wind schemes into three types: (1) slow wind with $V < 400$ km/s, (2) moderate wind with $400 \leq V \leq 500$ km/s, and (3) fast wind with $V > 500$ km/s. This categorization is based on the observed solar wind. Figure 6 shows distributions of Δt and ΔV for the slow (a, d), moderate (b, e), and fast (c, f) wind speeds.

We now discuss our direct statistical results from the histograms of Δt and ΔV obtained from the FastDTW alignment information in Figure 6. The mean Δt and median Δt (panels a, b, and c) all show positive values with greater values for higher wind speeds. These indicate that the solar wind structures modeled by Helio1D arrive ahead of the observations, especially for the fast wind where $\langle \Delta t \rangle = 6.6$ h and $\mu_{\Delta t} = 9$ h. On the Δt histograms, there again appear several spikes for the reason discussed above. Considering the ΔV histograms (panels d, e, and f), we find that the three wind types have rather different distributions. The ΔV distribution of the slow wind (Fig. 6d) is biased towards negative values and has a negative $\langle \Delta V \rangle$ of -35.3 km/s. This indicates that Helio1D overestimates

the slow wind speed. Meanwhile, the ΔV distribution of the fast wind (Fig. 6f) is biased towards positive values and has a very large, positive $\langle \Delta V \rangle$ of 63.4 km/s. This indicates that Helio1D significantly underestimates the fast wind speed. For the moderate wind (Fig. 6e), this $\langle \Delta V \rangle$ is moderate (-4.7 km/s). These results demonstrate that Helio1D indeed performs differently for various solar wind speed ranges. These insights are useful for the discussion of Helio1D caveats and future improvements as discussed in Section 7.

5 Prototype pipeline calibration

As mentioned in Section 4.1, Helio1D usually overestimates the number density and temperature of plasma at stream interaction regions as a consequence of over-compression at the stream interfaces. To alleviate this issue, we consider a post-calibration of the Helio1D data to lower the extreme peaks in the density and temperature. As shown in Section 4.2, the FastDTW technique provides the alignments between the modeled and observed solar wind speeds. Since the FastDTW algorithm maps similar structures within the two-time series, we create 2D histograms or heatmaps for the mapped values between the Helio1D and OMNI data of all the intervals (Fig. 2). Once mapped using the solar wind speed (e.g., Fig. 5a), we retrieve the alignment information and then obtain the mapping of other physical parameters. Figure 7 shows a kernel density estimation plot of the mapped OMNI (horizontal axis) and the Helio1D data (vertical axis), which shows their smoothed 2D density

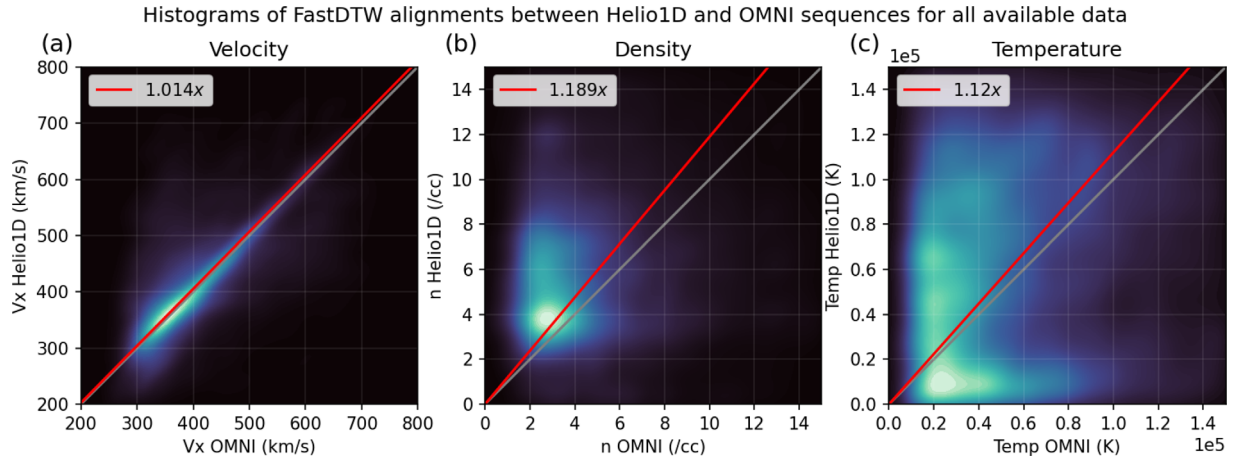


Figure 7. Heatmap of the FastDTW alignments between Helio1D (vertical axis) and OMNI (horizontal axis) sequences for (a) solar wind speed, (b) number density, and (c) temperature produced using all the data. Brighter regions have higher probability density than darker regions. Red solid lines show best linear fits with intercepts at zero. Grey solid lines show the lines with slope equal to one.

probability histogram. The heatmaps are shown for (a) velocity, (b) density, and (c) temperature. The brighter color in the heatmaps corresponds to a higher probability density (i.e. more data points). In addition, we add the line of the slope of 1 in each plot as a grey solid line. In the absence of systematic biases from the model, most data points should align with this line of slope of 1. Furthermore, we add a best linear fit with the intercept at zero for each parameter, shown as a red solid line with the corresponding equation. The velocity data in Figure 7a has a high probability density that mostly follows the line of slope of 1. The slope of the linear fit on the velocity is also close to one, showing statistically that there is almost no systematic bias for this parameter. The density in Figure 7b, in contrast, shows that Helio1D mostly overestimates the observed density; a similar conclusion can be drawn in Figure 7c for the temperature. The best linear fit is found to be $y = 1.189x$ for the density and $y = 1.12x$ for the temperature. To the first order, these linear functions may be used to lessen the overestimations of the density and the temperature from Helio1D. We apply such the correction to the Helio1D density and temperature for the prototype pipeline for forecasting as explained next.

6 Prototype pipeline for operational forecasting

We now shift our focus to the development of a stable “prototype” service for solar wind forecasting with Helio1D. To continuously predict solar wind conditions, several technical aspects as mentioned in Section 2.3 are addressed. First and foremost, characterization of model uncertainties is critical to assess the modeling results in order to plan for mitigation of plausible effects of extreme events. We provide model uncertainties using ensemble modeling in Section 6.1. Secondly, the 1D MHD model requires that the length of the input time series covers at least one solar rotation in order to produce consistent CIR formation. MULTI-VP has a setup that provides for the daily solar wind emergence comprising 72 h of solar wind conditions covering from the present day (D) to the next two

days ($D + 2$). Therefore, the solar wind emergence time series from MULTI-VP must be concatenated to produce a continuous time series with a minimum length of 27.25 days. Lastly, there can be data gaps and/or unphysical values arising from either the inputs for MultiVP (i.e., magnetogram), due to the lack of data or the presence of invalid data points, or numerical errors. The implementation of automatic procedures to tackle these issues is detailed in Section 6.2.

6.1 Helio1D ensemble modeling with 21 virtual targets

We perform an ensemble forecasting of Helio1D by considering a range of heliospheric latitudinal and longitudinal uncertainties. For example, a solar wind structure that is ahead or behind in time compared to the observation can be accounted for by considering the time series at some nearby heliospheric longitudes. Also, the magnitude of the solar wind speed profiles is different at different heliospheric latitudes as a function of the warping of the heliospheric current sheet and nearby velocity gradients. We note that this approach is partly similar to the one developed by Owens & Riley (2017). While Owens & Riley (2017) sampled at a range of latitudes, our approach implements a star-like grid with spacing along both latitudinal and longitudinal directions (see Fig. A1 in Appendix). Using daily magnetograms from WSO, we set up MULTI-VP to automatically generate daily one-dimensional solar wind profiles that cover from D to $D + 2$ at the sub-Earth point and the surrounding virtual targets. The surrounding virtual targets are set to spread up to 15° from the sub-Earth point in latitude and longitude; they are set at 5° apart from each other in latitude and longitude; these chosen points constitute the star-grid pattern with a total of 21 points including the sub-Earth point. The spatial cuts through these 21 points on the 2D map of solar wind emergence from MULTI-VP along the ecliptic are then translated to temporal solar wind profiles. These 21 time-series inputs are automatically fed into the 1D MHD model to propagate them in parallel from 0.14 AU to 1 AU. The solar wind propagation using the 1D MHD model is rather computationally inexpensive; it takes 2 min per profile to run on a single-core CPU.

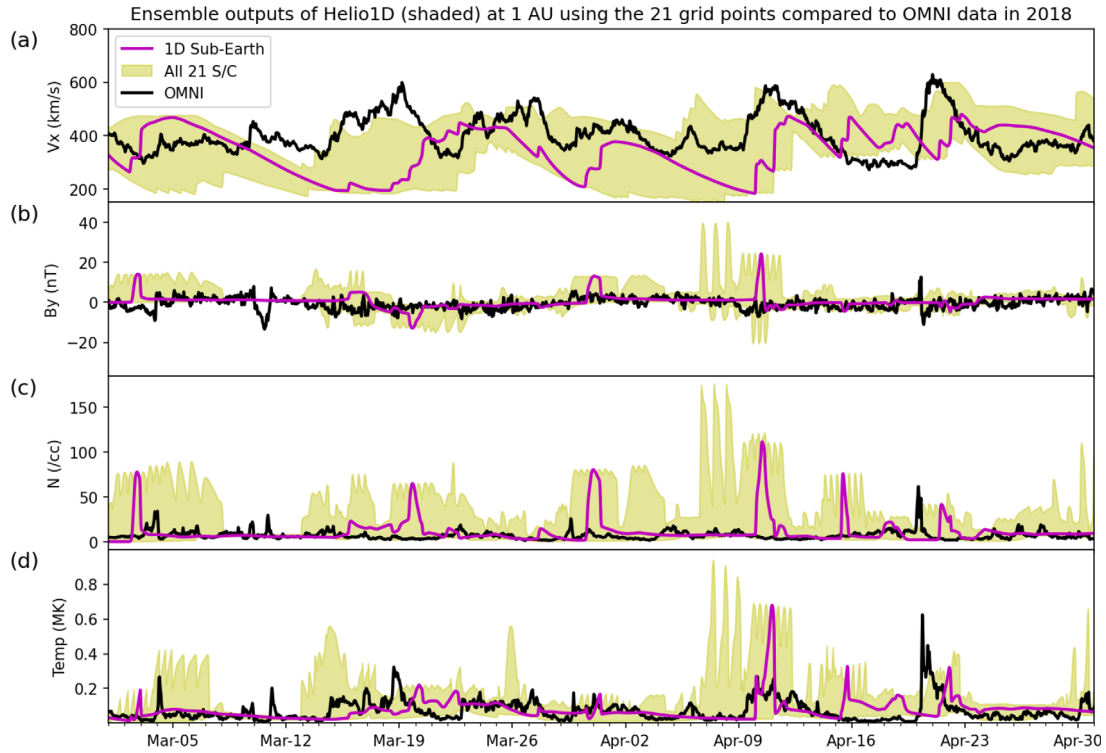


Figure 8. The ensemble Helio1D outputs (yellow-green shade) using the 21 virtual spacecraft in comparison to the OMNI data (black) during March and April 2018, shown for (a) the solar wind speed, (b) the tangential magnetic field, (c) the plasma number density, and (d) the plasma temperature. The Helio1D output at the sub-Earth point is displayed in magenta.

Figure 8 shows results from Helio1D for the 21 virtual targets for the interval from March to April 2018. The yellow-green shade highlights the spread of values between the minimum and maximum among the 21 virtual targets at each hour. We note that the post-calibration with the linear functions for number density and temperature (found in Section 5) has been applied. Here, the ensemble spread in a yellow-green shade provides an error bar. In this example, we find that the stream interfaces modeled for the sub-Earth point appear to lag behind the observations for about two days. Nevertheless, the timing uncertainties from the values at the virtual points indeed cover the timing of the arrivals of observed stream interfaces. Overall, we find that most of the observed data points fall within the error spread for both magnitude and timing, although there are some underestimations of the speed of the high-speed stream (consistent with our findings in Section 4.2). The Helio1D modeling using the 21 virtual targets may thus be used to provide error bars. Our future work will consider an exploitation of these multiple solar-wind solutions to provide a more robust estimation of the model uncertainties.

6.2 Solar wind data concatenation & data gaps

The main technical problems deal with the automated interfacing between MULTI-VP and 1D MHD model. Unlike the model benchmarking and case-by-case analysis, we rely on automatic modeling for both MULTI-VP and 1D MHD models. Here, the Helio1D prototype pipeline is scheduled to run in-house every day. First, MULTI-VP is set up to produce a daily forecast with a data length of 3 days, covering from day D to

$D + 2$. We first concatenate these MULTI-VP data for each virtual target by averaging the data from the day $D - 30$ to the present day D to make the time series input sufficiently long. After the concatenation process, this 1-month solar wind profile at 0.14 AU is subsequently propagated by the 1D MHD model to provide the solar wind profile at 1 AU. Since the solar wind takes time to propagate to the Earth, we gain an extra lead time of 2–7 days depending on the solar wind speed. To provide a total lead time of 4 days from Helio1D, we limit the extra time gained from the 1D MHD propagation modeling to 2 days. This setup has been done to automatically interface MULTI-VP and 1D MHD to provide daily solar wind nowcast and forecast from day D to $D + 4$ (see branch (b) in Fig. 1). In the absence of the magnetogram data, the magnetogram of the previous solar rotation is programmed to be fetched. This procedure is implemented based on the assumption that the coronal structures remain unchanged compared to the previous solar rotation (i.e. when the consistent, recurring CIRs are expected).

Furthermore, one of the models, or both, may terminate earlier than expected and give incomplete outputs. This is because, in reality, there can be data gaps arising from the lack of daily data (e.g., magnetogram data), or unphysical values in the data. The latter can come from numerical errors from one of the codes or poor quality of the raw inputs. To prevent the Helio1D prototype pipeline from early termination, we perform an automatic search for data gaps and unphysical values (after performing the fetching of the previous magnetogram when the present-day magnetogram is unavailable as mentioned above). The data gaps are then filled using linear interpolation between two available data points. The unphysical values identified using

the criteria (see Table A1 in the Appendix) are subsequently removed. The resulting data gaps are then filled using linear interpolation. To remove unphysical features (e.g., artificial or sharp discontinuities) that may arise from these procedures, we perform a running average on the data using a window of ± 3 h centered around the data point. Finally, when either the input or daily output is shorter than expected, we fill the data with latest available values to complete the length of the expected output. For practical purposes, we provide daily solar wind time series extending from day $D - 3$ to $D + 4$ for a given day D .

Plots of the daily data from Helio1D are publicly available via <http://swift.irap.omp.eu/>. Since the time of this development, we have changed the website to display the Helio1D data produced using ADAPT (Air Force Data Assimilative Photospheric flux Transport; Hickmann et al., 2015) magnetograms retrieved from GONG [Global Oscillation Network Group; Harvey et al., 1996], which provide generally better results. The error bars are produced from the standard deviation of the 21 solar-wind solutions. The modeling results are shown in comparison to the hourly-updated, real-time observations made by ACE and DSCOVR. The Helio1D data produced using WSO magnetograms remain available (see ‘Options’ under the menu bar).

7 Discussion

We developed Helio1D for solar wind modeling by interfacing MULTI-VP for the coronal part and 1D MHD for the interplanetary part. Regarding the interfacing, Helio1D is a modular and upgradable framework. Each step of the modeling pipeline—magnetogram, coronal field reconstruction, coronal solar wind, and heliospheric solar wind—can be switched over to a new one when necessary, and different combinations can be run simultaneously. The initial implementation of the pipeline, including all the model and data interfaces in a setup that runs autonomously, implies an intensive long-haul development roadmap. In addition, the methodology for its future upgrades from the current prototype to a more precise forecaster was made rather simple. This work reports on the initial implementation and validation results of the prototype pipeline. The interfacing has been done for the first time as an effort to connect two developed models for future operational forecasting purposes. Using the long-term MULTI-VP data produced with the WSO magnetograms in 2005–2013 and 2017–2018, we evaluated the prototype pipeline performance in various phases of the solar cycles 23 and 24. In addition to classic metrics such as RMSE, MAE, and skill score, we also complement our analysis by exploiting the FastDTW algorithm. We discuss the main findings and limitations as follows.

1. The performance of Helio1D in the implementation presented in this work is rather poor. One of the contributing factors is the quality of the source inputs. As a first step, we use synoptic WSO magnetograms as inputs for the coronal field reconstruction. Solar wind modeling using WSO magnetograms was shown to yield poorer results when compared to using GONG-ADAPT or HMI (the Helioseismic and Magnetic Imager, onboard SDO spacecraft; Schou et al., 2012) magnetograms (Perri et al., 2023). Specifically, WSO has the lowest resolution

(73×30 ; 5° in Carrington longitude) compared to GONG-ADAPT (360×180 ; 1° in Carrington longitude) and HMI (3600×1440 ; 0.1° in Carrington longitude). Here we focus first on results with WSO magnetograms as they provide the longest available and most continuous data (since 1976) while the others start much later (2006–2007 for GONG-ADAPT and 2010 for HMI). Our results with GONG-ADAPT should be analyzed, validated, and published in future work. Furthermore, our high uncertainties may be attributed to the underlying simplified assumptions used in our physics-based models (MULTI-VP, 1D MHD), as well as the extrapolation performed when processing the magnetograms. Identification of error sources, as well as the propagation of errors within the models, should be investigated in future work.

2. The performance of the Helio1D prototype pipeline varies with the phases of the solar cycle. Using the RMSE and MAE, measured for each 6-month interval, we find that their values positively correlate with the average sunspot numbers of the previous 6-month interval. The RMSE, for example, reaches a minimum value of 76 km/s during the solar minimum and reaches a maximum value of 189 km/s during the solar maximum. Overall, the Helio1D prototype pipeline provides more satisfactory results during the late declining phase and the solar minimum. These results are expected as the coronal structures remain stable over several solar rotations and there are fewer solar eruptions that would modify the coronal structures during such periods. Also, as we do not exclude CMEs in the observation data, the high errors in the solar ascending phase and solar maximum can be partly explained by their presence.
3. The Helio1D prototype pipeline produces consistent CIR formation. For each CIR, the solar wind speed profile at the stream interface qualitatively agrees with the observations (Fig. 4a). Thus, the model sufficiently includes the large-scale physics of the CIRs. However, due to the limited dimensionality and ideal MHD assumption, the compression at stream interfaces (resulting in extreme peaks of number density and temperature) is rather strong compared to the observations. This effect was found to be common to 1D MHD models used for solar wind propagation (e.g., Zieger and Hansen, 2008). To alleviate this effect, we propose a post-calibration of the Helio1D density and temperature to lower the peaks at stream interfaces using linear functions. These linear functions were found using the FastDTW alignments of the solar-wind speed sequences.
4. Using the FastDTW alignments, we evaluated the time and velocity differences between Helio1D and the observation for slow ($V < 400$ km/s), moderate ($400 \leq V \leq 500$ km/s), and fast wind ($V > 500$ km/s) speeds from January 2005 to June 2013 (Fig. 6). We find that the Helio1D sequence arrives ahead of the observations especially for the fast wind. Furthermore, Helio1D underestimates the slow wind speed while overestimating the fast wind speed.

There are limited existing works with which we can compare our results with as they mostly consider only short time intervals. Here we discuss the results where applicable. From Reiss et al. (2016), the ESWF and WSA models for ambient wind

have been benchmarked using data from 2011 to 2014. The ESWF is founded on an empirical relation between the areas of coronal holes as observed by EUV imaging, here taken from the SDO/AIA instrument, and the solar wind speed at 1 AU (Robbins et al., 2006; Vršnak et al., 2007). The WSA model is based on the empirical WSA relation at the source surface, taken from GONG magnetograms, coupled with a simple kinematic model (Arge and Pizzo, 2000). Both models yield only the solar wind speed, while our model provides also other physical parameters. The RMSE for the wind speed is found to be 108.2 km/s for the former and 99.5 km/s for the latter for the considered period. Our model yields the RMSE of 168 km/s for January 2011–June 2013, which is much higher. From Hinterreiter et al. (2019), the modeling of ambient wind from EUHFORIA using GONG-ADAPT magnetograms was validated for 2008 and 2012. The RMSE for both years is found to be about 125 km/s. Here, our RMSE values are 122 km/s for 2008 and 160 km/s for 2012. Based on these comparisons, our modeling results are not better than the existing physics-based models. This is likely because of uncertainties in the source inputs and the extrapolation employed for the coronal field reconstruction, the simplified assumptions employed in our models, as well as error propagation in the prototype pipeline.

There are a few works that benchmarked their results using large data sets. Owens & Riley (2017) utilized solar wind emergence at 0.14 AU from MAS based on Carrington maps (e.g., WSO) of the photospheric magnetic field. Particularly, they used a large ensemble of solar wind time series from MAS, where the ensemble members are produced from sampling solar wind solutions within a range of latitudes about the sub-Earth point. The solar wind flows at 0.14 AU are then propagated using the HUX model (Riley and Lionello, 2011) which based on the fluid momentum equation (e.g., Pizzo, 1978) and takes into account a residual solar wind acceleration (Schwenn, 1990). Using a long interval from 1996 to 2016, they find that the median RMSE of the upwind ensemble propagation is 107 km/s. Reiss et al. (2020) performed a forecasting of ambient solar wind using the WSA model and the HUX tool as well as the Tunable HUX (THUX; Reiss et al., 2020) to map solar wind flows from near Sun to Earth for the period 2006–2015. Their RMSE ranges from 90 to 122 km/s while their MAE ranges from 72 to 93 km/s. Their skill scores are negative for WSA/HUX and positive for WSA/THUX. Our RMSE and MAE values are in the range 76–189 km/s and 62–153 km/s, respectively, from January 2005 to June 2013. Additionally, the Helio1D skill scores are negative for all considered 6-month intervals when compared to the 27-day recurrence model, while they are positive for some of the intervals when compared to the 4-day persistence model. We emphasize that there are rather limited works that intensively benchmark their solar wind modeling. Also, the metrics used are rather varied. Community-wide efforts to unify the validation of existing solar wind models are critically needed (e.g., Reiss et al., 2023) for a more systematic evaluation and comparison of different frameworks.

Regarding the variation of the Helio1D performance with phases of the solar cycle, we find that our findings are generally consistent with other 1D MHD models. Zieger & Hansen (2008) performed an extensive validation of solar wind propagation using a 1D MHD model. They find that the variation

of the coronal structure on the timescale of a solar rotation, characterized by the recurrence index of solar wind speed, plays an important role in the prediction efficiency. This explains our Helio1D results such that, in an absence of variation, i.e., during the late declining phase, the model generally predicts CIR formation consistent with the observations. In contrast, in a presence of strong variations, i.e., when there are CME emergences during the ascending phase and the solar maximum, the model generally predicts poor results compared to the observations.

We find that there is a bias on the solar wind stream arrival time such that the Helio1D solar wind streams arrive earlier than what is shown by the observations. This bias is stronger for the fast wind compared to the slow wind. This effect may come from the assumption of an ideal MHD plasma in the 1D MHD model that ignores finer-scale physics and/or interactions that could slow down fast wind or accelerate slow wind in the interplanetary space. Moreover, the limited dimensionality of the simulation does not take into account the propagation in other directions; this may result in the time difference. Furthermore, the Helio1D underestimates the speed of the fast wind while overestimates the speed of the slow wind. The bias on the fast and slow wind magnitudes may come from the lack of the physics on solar wind acceleration. In the HUX model, an empirical, ad-hoc solar wind acceleration is incorporated into the model (Schwenn, 1990). Since the 1D MHD code was originally developed for solar wind propagation in the outer heliosphere (Tao et al., 2005), this physics was deemed less important and it was not added to the model. We note that this aspect must be addressed in a future improvement of the 1D MHD model for the inner heliosphere; this aspect is out of the scope of this work which focuses on interfacing the matured models.

Our work exploited the FastDTW technique to provide a complementary assessment of the Helio1D performance. In particular, the FastDTW alignments encoded statistics of time delays and magnitude differences between the modeled stream interfaces and observed structures. Nevertheless, the FastDTW technique is not without caveats. The main constraint of DTW techniques in general is the pathological mapping or the singularities. Therefore, the obtained Δt and ΔV from the DTW applications are not unique but instead depend on the constraints that have been added. Future works should include tests the robustness of various DTW algorithms.

Finally, we highlight the work that needs to be taken to develop an automated prototype solar wind forecasting pipeline of Helio1D which we have attempted for the SafeSpace project. In addition to the prototype pipeline calibration to alleviate the pipeline caveats, there are other aspects that must be addressed. This comprises (1) providing model uncertainties and (2) dealing with data gaps and bad data. The initial Helio1D modeling of an ensemble of 21 solar wind solutions was shown to be a reasonable method to provide a worst-case scenario, especially for the timing of stream interface arrivals. We note that the ensemble members can be further exploited to obtain an optimized forecast; this aspect is left for future work. Most importantly, to develop a reliable service, we implemented strategies to tackle data gaps and bad data points so that the prototype pipeline can automatically run and provide continuous nowcast and forecast. This aspect should be further tested in order to

evaluate their effects on the accuracy, performance, and stability of the future operational forecasting pipeline. Such a task is of particular importance; it requires dedicated tests on this prototype pipeline in various situations before employing it for operational services.

8 Summary

We developed a prototype pipeline “Helio1D” that automatically interfaces the coronal model “MULTI-VP” and the solar wind propagation model “1D MHD” to model the ambient solar wind comprising CIRs upstream of Earth. Our current work is focused on the proof-of-concept, initial implementation, and validation of Helio1D in light of the plans in place, e.g., using other magnetogram sources and including more physics in future efforts and improvements, as we develop toward a better-performing operational pipeline in the future.

Here, we benchmarked and tested the Helio1D prototype pipeline using data 2005–2013 and 2017–2018. As a first step, we tested the pipeline using synoptic WSO magnetograms for the coronal field reconstruction. We evaluated the performance of the prototype pipeline using classic metrics including RMSE, MAE, and skill score. We find that Helio1D performs best during the declining phase and solar minimum. In addition, we exploited the FastDTW technique to map solar wind stream structures to provide a complementary analysis. For instance, we characterized the statistical information on time and magnitude differences for velocity profiles of solar wind streams. We find that Helio1D underestimates the speed of the fast wind while overestimating the speed of the slow wind. Furthermore, the solar wind structures modeled by Helio1D often arrive earlier than what is shown by the observations, especially for the fast wind. These caveats plausibly arise from the simplistic assumptions in the 1D MHD model, comprising the limited dimensionality, the lack of dissipation, and the solar wind acceleration physics, for example. The Helio1D prototype pipeline is fully automated; it models consistent CIR formation while remaining computationally light, both of which are desirable for operational forecasting. Overall, the Helio1D results for the initial implementation presented in this work are less satisfactory compared to other existing physics-based models. We aim to further benchmark and improve Helio1D in the future by considering using other sources of magnetograms with higher resolution and including more physics.

To transition from case-by-case benchmarking to operational solar wind forecasting, we implemented (1) the prototype pipeline calibration to alleviate the over-compression at stream interface due to the ideal MHD assumption, (2) the ensemble modeling of 21 solar wind solutions at virtual targets including the sub-Earth point to provide timing and magnitude uncertainties, and (3) the procedures to remove unphysical data points and automatically fill data gaps. The latter two procedures were discussed to be crucial in providing continuous, reliable service while providing forecasting uncertainties. Alternatively, the Helio1D prototype pipeline can be adapted to use other magnetogram sources. For instance, the synchronic product from HMI, which uses synoptic magnetograms and replaces the portion within 30° of the central meridian with HMI observations of the day could be used. We emphasized that the implementation of procedures for making the pipeline fully operational,

in addition to the pipeline benchmarking, are critical and they must be further tested in several situations before employing the pipeline for real applications.

Acknowledgements

The SafeSpace project has received funding from the European Union’s Horizon 2020 research and innovation program under grant agreement No 870437. The 1D MHD model of solar wind propagation initially developed by Tao et al. (2005) and used for Helio1D is duplicated from the Heliopropa service provided by CDP (http://heliopropa.irap.omp.eu) and developed during the Planetary Space Weather Services (PSWS) Virtual Activity of the Europlanet H2020 Research Infrastructure funded by the European Union’s Horizon 2020 research and innovation program under grant agreement No 654208 and extended during the Sun Planet Interactions Digital Environment on Request (SPIDER) Virtual Activity of the Europlanet H2024 Research Infrastructure funded by the European Union’s Horizon 2020 research and innovation programme under grant agreement No 871149. The FastDTW algorithm is open-source and it can be accessed at <https://pypi.org/project/fastdtw/>. Work at IRAP is supported by the French National Centre for Scientific Research (CNRS), Centre National d’Études Spatiales (CNES), and the University of Toulouse III (UPS). We acknowledge the use of NASA/GSFC’s Space Physics Data Facilities (<http://doi.org/10.17616/R3P301>): OMNIWeb (<http://doi.org/10.17616/R3TH0D>), CDAWeb (<http://doi.org/10.17616/R39H0R>) and OMNI data. The editor thanks three anonymous reviewers for their assistance in evaluating this paper.

References

- Alves MV, Echer E, Gonzalez WD. 2006. Geoeffectiveness of corotating interaction regions as measured by Dst index. *J Geophys Res* **111**(A7): A07S05. <https://doi.org/10.1029/2005JA011379>.
- André N, Grande M, Achilleos N, Barthélémy M, Bouchemit M, et al. 2018. Virtual planetary space weather services offered by the europlanet H2020 research infrastructure. *Planet Space Sci* **150**: 50–59. <https://doi.org/10.1016/j.pss.2017.04.020>.
- Arge CN, Luhmann JG, Odstrcil D, Schrijver CJ, Li Y. 2004. Stream structure and coronal sources of the solar wind during the May 12th, 1997 CME. *J Atm Solar-Terr Phys* **66**(15–16): 1295–1309. <https://doi.org/10.1016/j.jastp.2004.03.018>.
- Arge CN, Pizzo VJ. 2000. Improvement in the prediction of solar wind conditions using near-real time solar magnetic field updates. *J Geophys Res* **105**(A5): 10465–10480. <https://doi.org/10.1029/1999JA000262>.
- Bellman R, Kalaba R. 1959. On adaptive control processes. *IRE Trans Autom Control* **4**(2): 1–9. <https://doi.org/10.1109/TAC.1959.1104847>.
- Berndt D, Clifford J. 1994. Using dynamic time warping to find patterns in time series. In: “Workshop on Knowledge Discovery in Databases”, AAAI Technical Report WS-94-03.
- Blake JB, Baker DN, Turner N, Ogilvie KW, Lepping RP. 1997. Correlation of changes in the outer-zone relativistic-electron population with upstream solar wind and magnetic field measurements. *Geophys Res Lett* **24**(8): 927–929. <https://doi.org/10.1029/97GL00859>.
- Borovsky JE, Denton MH. 2006. Differences between CME-driven storms and CIR-driven storms. *J Geophys Res* **111**(A7): A07S08. <https://doi.org/10.1029/2005JA011447>.
- Bunting KA, Morgan H. 2022. An inner boundary condition for solar wind models based on coronal density. *J Space Weather Space Clim* **12**: 30. <https://doi.org/10.1051/swsc/2022026>.

- Chi Y, Shen C, Luo B, Wang Y, Xu M. 2018. Geoeffectiveness of stream interaction regions from 1995 to 2016. *Space Weather* **16**(12): 1960–1971. <https://doi.org/10.1029/2018SW001894>.
- Chu S, Keogh E, Hart D, Pazzani M. 2002. Iterative deepening dynamic time warping for time series. In: Proceedings of the 2002 SIAM International Conference on Data Mining, Eds. Robert Grossman, Jiawei Han, Vipin Kumar, Heikki Mannila, and Rajeev Motwani. <https://doi.org/10.1137/1.9781611972726.12>.
- Cranmer SR, Gibson SE, Riley P. 2017. Origins of the ambient solar wind: implications for space weather. *Space Sci Rev* **212**(3–4): 1345–1384. <https://doi.org/10.1007/s11214-017-0416-y>.
- Dósa M, Opitz A, Dálya Z, Szegő K. 2018. Magnetic Lasso: a new kinematic solar wind propagation method. *Sol Phys* **293**(9): 127. <https://doi.org/10.1007/s11207-018-1340-3>.
- Efrat A, Fan Q, Venkatasubramanian S. 2007. Curve matching, time warping, and light fields: new algorithms for computing similarity between curves. *J Math Imaging Vis* **27**(3): 203–216. <https://doi.org/10.1007/s10851-006-0647-0>.
- Furtună TF. 2008. Dynamic programming algorithms in speech recognition. *Rev Inform Econ* **2**(46): 94.
- Górecki T, Łuczak M. 2013. Using derivatives in time series classification. *Data Mining Knowledge Discov* **26**(2): 310–331. <https://doi.org/10.1007/s10618-012-0251-4>.
- Grandin M, Aikio AT, Kozlovsky A. 2019. Properties and geoeffectiveness of solar wind high-speed streams and stream interaction regions during solar cycles 23 and 24. *J Geophys Res* **124**(6): 3871–3892. <https://doi.org/10.1029/2018JA026396>.
- Harvey JW, Hill F, Hubbard RP, Kennedy JR, Leibacher JW, et al. 1996. The global oscillation network group (GONG) project. *Science* **272**(5266): 1284–1286. <https://doi.org/10.1126/science.272.5266.1284>.
- Hickmann KS, Godinez HC, Henney CJ, Arge CN. 2015. Data assimilation in the ADAPT photospheric flux transport model. *Sol Phys* **290**(4): 1105–1118. <https://doi.org/10.1007/s11207-015-0666-3>.
- Hinterreiter J, Magdalenic J, Temmer M, Verbeke C, Jebaraj IC, et al. 2019. Assessing the performance of EUHFORIA modeling the background solar wind. *Sol Phys* **294**(12): 170. <https://doi.org/10.1007/s11207-019-1558-8>.
- Horne RB, Glauert SA, Meredith NP, Koskinen H, Vainio R, et al. 2013. Forecasting the Earth's radiation belts and modelling solar energetic particle events: recent results from SPACECAST. *J Space Weather Space Clim* **3**: A20. <https://doi.org/10.1051/swsc/2013042>.
- Horne RB, Phillips MW, Glauert SA, Meredith NP, Hands ADP, Ryden KA, Li W. 2018. Realistic worst case for a severe space weather event driven by a fast solar wind stream. *Space Weather* **16**(9): 1202–1215. <https://doi.org/10.1029/2018SW001948>.
- Hudson MK, Elkington SR, Li Z, Patel M, Pham K, Sorathia K, Boyd A, Jaynes A, Leali A. 2021. MHD-test particles simulations of moderate CME and CIR-driven geomagnetic storms at solar minimum. *Space Weather* **19**(12): e02882. <https://doi.org/10.1029/2021SW002882>.
- Jian LK, MacNeice PJ, Taktakishvili A, Odstrcil D, Jackson B, Yu HS, Riley P, Sokolov IV, Evans RM. 2015. Validation for solar wind prediction at Earth: comparison of coronal and heliospheric models installed at the CCMC. *Space Weather* **13**(5): 316–338. <https://doi.org/10.1002/2015SW001174>.
- Jian LK, Russell CT, Luhmann JG. 2011a. Comparing solar minimum 23/24 with historical solar wind records at 1 AU. *Sol Phys* **274**(1–2): 321–344. <https://doi.org/10.1007/s11207-011-9737-2>.
- Jian LK, Russell CT, Luhmann JG, MacNeice PJ, Odstrcil D, Riley P, Linker JA, Skoug RM, Steinberg JT. 2011b. Comparison of observations at ACE and Ulysses with Enlil model results: stream interaction regions during Carrington rotations 2016–2018. *Sol Phys* **273**(1): 179–203. <https://doi.org/10.1007/s11207-011-9858-7>.
- Katsavrias C, Sandberg I, Li W, Podladchikova O, Daglis I, Papadimitriou C, Tsironis C, Aminiargia-Giamini S. 2019. Highly relativistic electron flux enhancement during the weak geomagnetic storm of April–May 2017. *J Geophys Res* **124**(6): 4402–4413. <https://doi.org/10.1029/2019JA026743>.
- Keogh E, Ratanamahatana CA. 2005. Exact indexing of dynamic time warping. *Knowl Inf Syst* **7**(3): 358–386. <https://doi.org/10.1007/s10115-004-0154-9>.
- Keogh EJ, Pazzani MJ. 2001. Derivative dynamic time warping. In: Proceedings of the 2001 SIAM International Conference on Data Mining, Eds. Vipin Kumar and Robert Grossman, <https://doi.org/10.1137/1.9781611972719.1>.
- Kilpua EKJ, Balogh A, von Steiger R, Liu YD. 2017. Geoeffective properties of solar transients and stream interaction regions. *Space Sci Rev* **212**(3): 1271–1314. <https://doi.org/10.1007/s11214-017-0411-3>.
- King JH, Papitashvili NE. 2005. Solar wind spatial scales in and comparisons of hourly Wind and ACE plasma and magnetic field data. *J Geophys Res* **110**(A2): A02104. <https://doi.org/10.1029/2004JA010649>.
- Lam MM, Horne RB, Meredith NP, Glauert SA. 2009. Radiation belt electron flux variability during three CIR-driven geomagnetic storms. *J Atm Solar-Terr Phys* **71**(10–11): 1145–1156. <https://doi.org/10.1016/j.jastp.2008.06.007>.
- Lanzerotti LJ. 2007. Space weather effects on communications. In: *Space weather physics and effects*. Springer Praxis Books, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-34578-7_9.
- Laperre B, Amaya J, Lapenta G. 2020. Dynamic time warping as a new evaluation for Dst forecast with machine learning. *Front Astron Space Sci* **7**: 39. <https://doi.org/10.3389/fspas.2020.00039>.
- Lee CO, Luhmann JG, Odstrcil D, MacNeice PJ, de Pater I, Riley P, Arge CN. 2009. The solar wind at 1 AU during the declining phase of solar cycle 23: comparison of 3D numerical model results with observations. *Sol Phys* **254**(1): 155–183. <https://doi.org/10.1007/s11207-008-9280-y>.
- Li X, Baker DN, Temerin M, Larson D, Lin RP, Reeves GD, Looper M, Kanekal SG, Mewaldt RA. 1997. Are energetic electrons in the solar wind the source of the outer radiation belt? *Geophys Res Lett* **24**(8): 923–926. <https://doi.org/10.1029/97GL00543>.
- Linker JA, Mikić Z, Biesecker DA, Forsyth RJ, Gibson SE, Lazarus AJ, Lecinski A, Riley P, Szabo A, Thompson BJ. 1999. Magnetohydrodynamic modeling of the solar corona during Whole Sun Month. *J Geophys Res* **104**(A5): 9809–9830. <https://doi.org/10.1029/1998JA900159>.
- MacNeice P. 2009a. Validation of community models: 2. Development of a baseline using the Wang-Sheeley-Arge model. *Space Weather* **7**(12): S12002. <https://doi.org/10.1029/2009SW000489>.
- MacNeice P. 2009b. Validation of community models: Identifying events in space weather model timelines. *Space Weather* **7**(6): S06004. <https://doi.org/10.1029/2009SW000463>.
- MacNeice P, Jian LK, Antiochos SK, Arge CN, Bussy-Virat CD, et al. 2018. Assessing the quality of models of the ambient solar wind. *Space Weather* **16**(11): 1644–1667. <https://doi.org/10.1029/2018SW002040>.
- Merkin VG, Lionello R, Lyon JG, Linker J, Török T, Downs C. 2016. Coupling of coronal and heliospheric magnetohydrodynamic models: solution comparisons and verification. *Astrophys J* **831**(1): 23. <https://doi.org/10.3847/0004-637X/831/1/23>.
- Milošić D, Temmer M, Heinemann SG, Podladchikova T, Veronig A, Vršnak B. 2023. Improving the empirical solar wind forecast (ESWF). *Sol Phys* **298**(3): 45. <https://doi.org/10.1007/s11207-022-02102-5>.

- Miyoshi Y, Kataoka R, Kasahara Y, Kumamoto A, Nagai T, Thomsen MF. 2013. High-speed solar wind with southward interplanetary magnetic field causes relativistic electron flux enhancement of the outer radiation belt via enhanced condition of whistler waves. *Geophys Res Lett* **40**(17): 4520–4525. <https://doi.org/10.1002/grl.50916>.
- Müller M. 2007. Dynamic time warping. In: *Information retrieval for music and motion*. Springer, Berlin Heidelberg. pp. 69–84. ISBN 978-3-540-74048-3. https://doi.org/10.1007/978-3-540-74048-3_4.
- Myers C, Rabiner L, Rosenberg A. 1980. Performance tradeoffs in dynamic time warping algorithms for isolated word recognition. *IEEE Trans Acoust Speech* **28**(6): 623–635. <https://doi.org/10.1109/TASSP.1980.1163491>.
- Nilsson H, Moeslinger A, Williamson HN, Bergman S, Gunell H, Stenberg Wieser G, Futaana Y, Karlsson T, Behar E, Holmström M. 2022. Upstream solar wind speed at comet 67P. Reconstruction method, model comparison, and results. *Astron Astrophys* **659**: A18. <https://doi.org/10.1051/0004-6361/202142867>.
- Odstreil D, Pizzo VJ. 1999. Distortion of the interplanetary magnetic field by three-dimensional propagation of coronal mass ejections in a structured solar wind. *J Geophys Res* **104**(A12): 28225–28240. <https://doi.org/10.1029/1999JA000319>.
- Odstreil D, Pizzo VJ, Linker JA, Riley P, Lionello R, Mikic Z. 2004. Initial coupling of coronal and heliospheric numerical magneto-hydrodynamic codes. *J Atm Solar-Terr Phys* **66**(15–16): 1311–1320. <https://doi.org/10.1016/j.jastp.2004.04.007>.
- Odstreil D, Dryer M, Smith Z. 1996. Propagation of an interplanetary shock along the heliospheric plasma sheet. *J Geophys Res* **101**(A9): 19973–19986. <https://doi.org/10.1029/96JA00479>.
- Owens MJ, Arge CN, Spence HE, Pembroke A. 2005. An event-based approach to validating solar wind speed predictions: High-speed enhancements in the Wang-Sheeley-Arge model. *J Geophys Res* **110**(A12): A12105. <https://doi.org/10.1029/2005JA011343>.
- Owens MJ, Challen R, Methven J, Henley E, Jackson DR. 2013. A 27 day persistence model of near-Earth solar wind conditions: A long lead-time forecast and a benchmark for dynamical models. *Space Weather* **11**(5): 225–236. <https://doi.org/10.1002/swe.20040>.
- Owens MJ, Nichols JD. 2021. Using in situ solar-wind observations to generate inner-boundary conditions to outer-heliosphere simulations – I. Dynamic time warping applied to synthetic observations. *Mon Notices Royal Astron Soc* **508**(2): 2575–2582. <https://doi.org/10.1093/mnras/stab2512>.
- Owens MJ, Riley P. 2017. Probabilistic solar wind forecasting using large ensembles of near-Sun conditions with a simple one-dimensional “upwind” scheme. *Space Weather* **15**(11): 1461–1474. <https://doi.org/10.1002/2017SW001679>.
- Palmerio E, Nieves-Chinchilla T, Kilpua EKJ, Barnes D, Zhukov AN, et al. 2021. Magnetic structure and propagation of two interacting CMEs from the Sun to Saturn. *J Geophys Res* **126**(11): e2021JA029770. <https://doi.org/10.1029/2021JA029770>.
- Paulikas GA, Blake JB. 1979. Effects of the solar wind on magnetospheric dynamics: Energetic electrons at the synchronous orbit. In: *Quantitative modeling of magnetospheric processes*, Ed. W.P. Olson. <https://doi.org/10.1029/GM021p0180>.
- Perri B, Kuźma B, Brčnelova M, Baratashvili T, Zhang F, Leitner P, Lani A, Poedts S. 2023. COCONUT, a novel fast-converging MHD model for solar corona simulations. II. Assessing the impact of the input magnetic map on space-weather forecasting at minimum of activity. *Astrophys J* **943**(2): 124. <https://doi.org/10.3847/1538-4357/ac9799>.
- Pinto RF, Rouillard AP. 2017. A multiple flux-tube solar wind model. *Astrophys J* **838**(2): 89. <https://doi.org/10.3847/1538-4357/aa6398>.
- Pinto VA, Kim H-J, Lyons LR, Bortnik J. 2018. Interplanetary parameters leading to relativistic electron enhancement and persistent depletion events at geosynchronous orbit and potential for prediction. *J Geophys Res* **123**(2): 1134–1145. <https://doi.org/10.1002/2017JA024902>.
- Pizzo V. 1978. A three-dimensional model of corotating streams in the solar wind 1. Theoretical foundations. *J Geophys Res* **83**(A12): 5563–5572. <https://doi.org/10.1029/JA083iA12p05563>.
- Pizzo V, Millward G, Parsons A, Biesecker D, Hill S, Odstreil D. 2011. Wang-Sheeley-Arge-Enlil Cone model transitions to operations. *Space Weather* **9**(3): 03004. <https://doi.org/10.1029/2011SW000663>.
- Pomoell J, Poedts S. 2018. EUHFORIA: European heliospheric forecasting information asset. *J Space Weather Space Clim* **8**: A35. <https://doi.org/10.1051/swsc/2018020>.
- Ratanamahatana CA, Keogh E. 2004. Making time-series classification more accurate using learned constraints. In: *Proceedings of the 2004 SIAM international conference on data mining (SDM)* Eds. Berry Michael W, Dayal U, Kamath C, Skillicorn D, pp. 11–22. <https://doi.org/10.1137/1.9781611972740.2>.
- Reiss MA, MacNeice PJ, Mays LM, Arge CN, Möstl C, Nikolic L, Amerstorfer T. 2019. Forecasting the ambient solar wind with numerical models. I. On the implementation of an operational framework. *Astrophys J* **240**(2): 35. <https://doi.org/10.3847/1538-4365/aaf8b3>.
- Reiss MA, MacNeice PJ, Muglach K, Arge CN, Möstl C, et al. 2020. Forecasting the ambient solar wind with numerical models. II. An adaptive prediction system for specifying solar wind speed near the Sun. *Astrophys J* **891**(2): 165. <https://doi.org/10.3847/1538-4357/ab78a0>.
- Reiss MA, Muglach K, Mullinix R, Kuznetsova MM, Wiegand C, et al. 2023. Unifying the validation of ambient solar wind models. *Adv Space Res* **72**(12): 5275–5286. <https://doi.org/10.1016/j.asr.2022.05.026>.
- Reiss MA, Temmer M, Veronig AM, Nikolic L, Vennerstrom S, Schöngassner F, Hofmeister SJ. 2016. Verification of high-speed solar wind stream forecasts using operational solar wind models. *Space Weather* **14**(7): 495–510. <https://doi.org/10.1002/2016SW001390>.
- Réville V, Poirier N, Kouloumvakos A, Rouillard AP, Ferreira Pinto R, et al. 2023. HelioCast: heliospheric forecasting based on white-light observations of the solar corona. *J Space Weather Space Clim* **13**: 11. <https://doi.org/10.1051/swsc/2023008>.
- Richardson IG. 2018. Solar wind stream interaction regions throughout the heliosphere. *Living Rev Sol Phys* **15**(1): 1. <https://doi.org/10.1007/s41116-017-0011-z>.
- Richardson IG, Cliver EW, Cane HV. 2001. Sources of geomagnetic storms for solar minimum and maximum conditions during 1972–2000. *Geophys Res Lett* **28**(13): 2569–2572. <https://doi.org/10.1029/2001GL013052>.
- Riley P, Linker JA, Lionello R, Mikic Z. 2012. Corotating interaction regions during the recent solar minimum: The power and limitations of global MHD modeling. *J Atm Solar-Terr Phys* **83**: 1–10. <https://doi.org/10.1016/j.jastp.2011.12.013>.
- Riley P, Lionello R. 2011. Mapping solar wind streams from the Sun to 1 AU: a comparison of techniques. *Sol Phys* **270**(2): 575–592. <https://doi.org/10.1007/s11207-011-9766-x>.
- Robbins S, Henney CJ, Harvey JW. 2006. Solar wind forecasting with coronal holes. *Sol Phys* **233**(2): 265–276. <https://doi.org/10.1007/s11207-006-0064-y>.
- Rouillard AP, Pinto RF, Vourlidis A, De Groof A, Thompson WT, et al. 2020. Models and data analysis tools for the Solar Orbiter mission. *Astron Astrophys* **642**: A2. <https://doi.org/10.1051/0004-6361/201935305>.

- Sakoe H, Chiba S. 1978. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Trans Acoust Speech* **26**(1): 43–49. <https://doi.org/10.1109/TASSP.1978.1163055>.
- Salvador S, Chan P. 2007. Toward accurate dynamic time warping in linear time and space. *Intell Data Anal* **11**(5): 561–580. <https://doi.org/10.3233/IDA-2007-11508>.
- Samara E, Laperre B, Kieokaew R, Temmer M, Verbeke C, Rodriguez L, Magdalenic J, Poedts S. 2022. Dynamic time warping as a means of assessing solar wind time series. *Astrophys J* **927**(2): 187. <https://doi.org/10.3847/1538-4357/ac4af6>.
- Samara E, Pinto RF, Magdalenic J, Wijsen N, Jerčic V, Scolini C, Jebaraj IC, Rodriguez L, Poedts S. 2021. Implementing the MULTI-VP coronal model in EUHFORIA: Test case results and comparisons with the WSA coronal model. *Astron Astrophys* **648**: A35. <https://doi.org/10.1051/0004-6361/202039325>.
- Schou J, Scherrer PH, Bush RI, Wachter R, Couvidat S, et al. 2012. Design and ground calibration of the helioseismic and magnetic imager (HMI) instrument on the solar dynamics observatory (SDO). *Sol Phys* **275**(1–2): 229–259. <https://doi.org/10.1007/s11207-011-9842-2>.
- Schwenn R. 1990. Large-scale structure of the interplanetary medium. In: Schwenn, R, Marsch E. (Eds.) *Physics of the inner heliosphere I. Physics and chemistry in space, space and solar physics*, vol 20. Springer, Berlin, Heidelberg. ISBN 978-3-642-75361-9. https://doi.org/10.1007/978-3-642-75361-9_3.
- Taktakishvili A, Kuznetsova M, MacNeice P, Hesse M, Rastätter L, Pulkkinen A, Chulaki A, Odstreil D. 2009. Validation of the coronal mass ejection predictions at the Earth orbit estimated by ENLIL heliosphere cone model. *Space Weather* **7**(3): S03004. <https://doi.org/10.1029/2008SW000448>.
- Taktakishvili A, Pulkkinen A, MacNeice P, Kuznetsova M, Hesse M, Odstreil D. 2011. Modeling of coronal mass ejections that caused particularly large geomagnetic storms using ENLIL heliosphere cone model. *Space Weather* **9**(6): 06002. <https://doi.org/10.1029/2010SW000642>.
- Tao C, Kataoka R, Fukunishi H, Takahashi Y, Yokoyama T. 2005. Magnetic field variations in the Jovian magnetotail induced by solar wind dynamic pressure enhancements. *J Geophys Res* **110**(A11): A11208. <https://doi.org/10.1029/2004JA010959>.
- Tóth G, Sokolov IV, Gombosi TI, Chesney DR, Clauer CR, et al. 2005. Space Weather Modeling Framework: A new tool for the space science community. *J Geophys Res* **110**(A12): A12226. <https://doi.org/10.1029/2005JA011126>.
- Tsurutani BT, Gonzalez WD, Gonzalez ALC, Guarnieri FL, Gopalswamy N, et al. 2006. Corotating solar wind streams and recurrent geomagnetic activity: a review. *J Geophys Res* **111**(A7): A07S01. <https://doi.org/10.1029/2005JA011273>.
- Vršnak B, Dumbović M, Čalogović J, Verbanac G, Poljančič Beljan I. 2017. Geomagnetic effects of corotating interaction regions. *Sol Phys* **292**(9): 140. <https://doi.org/10.1007/s11207-017-1165-5>.
- Vršnak B, Temmer M, Veronig AM. 2007. Coronal holes and solar wind high-speed streams: I. Forecasting the solar wind parameters. *Sol Phys* **240**(2): 315–330. <https://doi.org/10.1007/s11207-007-0285-8>.
- Wang YM, Sheeley NR. 1992. On potential field models of the solar corona. *Astrophys J* **392**: 310. <https://doi.org/10.1086/171430>.
- Woodcock F. 1976. The evaluation of yes/no forecasts for scientific and administrative purposes. *Mon Weather Rev* **104**(10): 1209–1214.
- Yadav M, Alam MA. 2018. Dynamic time warping (dtw) algorithm in speech: a review. *Int J Res Electr Comp Eng* **6**(1): 524–528.
- Yermolaev YI, Lodkina IG, Nikolaeva NS, Yermolaev MY, Riazantseva MO, Rakhmanova LS. 2018. Statistic study of the geoeffectiveness of compression regions CIRs and Sheaths. *J Atm Solar-Terr Phys* **180**: 52–59. <https://doi.org/10.1016/j.jastp.2018.01.027>.
- Zhang Y, Sun W, Feng XS, Deehr CS, Fry CD, Dryer M. 2008. Statistical analysis of corotating interaction regions and their geoeffectiveness during solar cycle 23. *J Geophys Res* **113**(A8): A08106. <https://doi.org/10.1029/2008JA013095>.
- Zhu Q, Batista G, Rakthanmanon T, Keogh E. 2012. A novel approximation to dynamic time warping allows anytime clustering of massive time series datasets. In: *Proceedings of the 2012 SIAM international conference on data mining (SDM)*. Eds. Ghosh J, Liu H, Davidson I, Domeniconi C, Kamath C, pp. 999–1010. <https://doi.org/10.1137/1.9781611972825.86>.
- Zieger B, Hansen KC. 2008. Statistical validation of a solar wind propagation model from 1 to 10 AU. *J Geophys Res* **113**: A8. <https://doi.org/10.1029/2008JA013046>.

Appendix

Table A1. Criteria for unphysical values. The data points that are outside of these minimum and maximum are removed.

Parameters	Solar wind speed (km/s)	Number density (cm ⁻³)	Temperature (K)	By (nT)
Minimum	150	200	0.2	−100
Maximum	800	1500	1e6	100

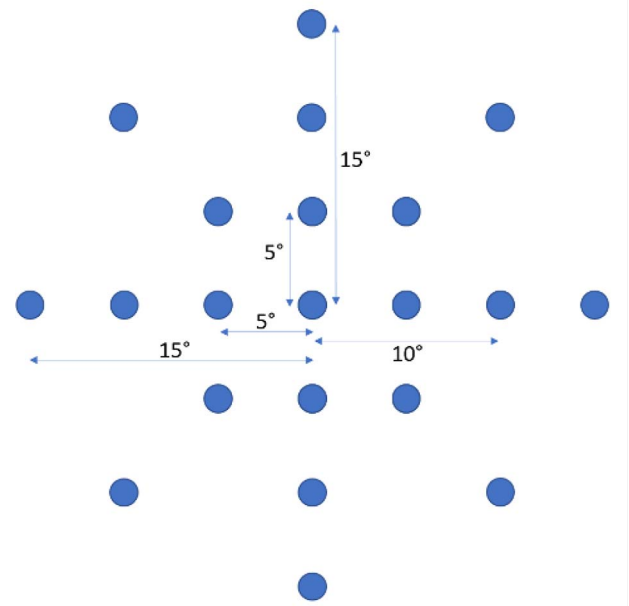


Figure A1. The grid-pattern showing positions of the 21 targets for 1-D solar wind solutions. The Sub-Earth point is at the center.