University of Reading

Improving solar wind forecasting using data assimilation

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open Access

Lang, M. ORCID: https://orcid.org/0000-0002-1904-3700, Witherington, J., Turner, H. ORCID: https://orcid.org/0000-0002-4012-8004, Owens, M. J. ORCID: https://orcid.org/0000-0003-2061-2453 and Riley, P. ORCID: https://orcid.org/0000-0002-1859-456X (2021) Improving solar wind forecasting using data assimilation. Space Weather, 19 (7). ISSN 1542-7390 doi: 10.1029/2020SW002698 Available at https://centaur.reading.ac.uk/98785/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1029/2020SW002698

Publisher: American Geophysical Union

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

www.reading.ac.uk/centaur

University of Reading

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Check for updates

Space Weather

RESEARCH ARTICLE
10.1029/2020SW002698

Improving Solar Wind Forecasting Using Data Assimilation

Matthew Lang[1], Jake Witherington[1], Harriet Turner[1], Mathew J. Owens[1], and Pete Riley[2]

[1]Department of Meteorology, University of Reading, Reading, Berkshire, UK, [2]Predictive Science Inc., San Diego, CA, USA

Key Points:

- Data assimilation (DA) is shown to improve solar wind speed forecasts over the interval 2007–2014
- Initializing forecasts with DA leads to a 31.4% reduction in solar wind speed root mean square error (RMSE) compared to forecasts with no DA
- DA can remove significant positive correlations between forecast RMSE and latitude of STEREO-B

Correspondence to:
M. Lang,
matthew.lang@reading.ac.uk

Citation:
Lang, M., Witherington, J., Turner, H., Owens, M. J., & Riley, P. (2021). Improving solar wind forecasting using data assimilation. *Space Weather*, *19*, e2020SW002698. https://doi.org/10.1029/2020SW002698

Received 10 DEC 2020
Accepted 7 JUN 2021

Abstract Data Assimilation (DA) has enabled huge improvements in the skill of terrestrial operational weather forecasting. In this study, we use a variational DA scheme with a computationally efficient solar wind model and in situ observations from STEREO-A, STEREO-B and ACE. This scheme enables solar-wind observations far from the Sun, such as at 1 AU, to update and improve the inner boundary conditions of the solar wind model (at 30 solar radii). In this way, observational information can be used to improve estimates of the near-Earth solar wind, even when the observations are not directly downstream of the Earth. This allows improved initial conditions of the solar wind to be passed into forecasting models. To this effect, we employ the HUXt solar wind model to produce 27-day forecasts of the solar wind during the operational lifetime of STEREO-B (November 01, 2007–September 30, 2014). In near-Earth space, we compare the accuracy of these DA forecasts with both non-DA forecasts and simple corotation of STEREO-B observations. We find that 27-day root mean square error (RMSE) for STEREO-B corotation and DA forecasts are comparable and both are significantly lower than non-DA forecasts. However, the DA forecast is shown to improve solar wind forecasts when STEREO-B's latitude is offset from the Earth, which is an issue for corotation forecasts. And the DA scheme enables the representation of the solar wind in the whole model domain between the Sun and the Earth to be improved, which will enable improved forecasting of CME arrival time and speed.

Plain Language Summary Forecasting the adverse effects of space weather at Earth requires accurate modeling of the solar wind, the continuous outflow of matter from the Sun's atmosphere into the solar system. At present, computer simulations of the solar wind are generated from telescopic observations of the Sun's surface. The simulations then propagate structures to the Earth with no further input from observations. However, spacecraft also take direct measurements of the solar wind near Earth's orbit. Using a technique called “data assimilation” (DA) we are able to combine these direct measurements with solar wind models to provide improved forecasts of the solar wind. To quantify the ability of DA to improve forecasts, we generate 27-day forecasts for the solar wind using DA over the lifetime of the STEREO-B spacecraft (2007–2014) and find that, on average over the 7-year period, DA can reduce forecast errors by around a third. Furthermore, we see that DA is able to reduce biases from available spacecraft observations being misaligned with Earth.

1. Introduction

The solar wind is a continuous outflow of plasma and magnetic flux which fills the heliosphere (e.g., Owens & Forsyth, 2013). Variations in the solar wind properties in near-Earth space can lead to adverse effects (Hapgood, 2011), such as disruption to power and communications systems, and health hazards for astronauts and aircrew on flights over the poles. As a result, it is vital to accurately and efficiently forecast solar-wind conditions in near-Earth space. However, forecasting ahead by more than about one hour (the nominal travel time to Earth for solar wind from L1, the gravitational null 0.01 AU ahead of Earth on the Earth-Sun line, where observing spacecraft are routinely positioned) requires prediction of the solar wind conditions near the Sun. This is typically achieved using empirical relations to the coronal magnetic field (Arge et al., 2003; McGregor et al., 2011; Riley et al., 2015), which is in turn determined using extrapolation from the observed photospheric magnetic field (Linker et al., 1999; Mackay & Yeates, 2012). These empirical models approximate the solar wind conditions at ~21–30 solar radii (r_S). From here, the solar wind is propagated to Earth, typically using a numerical magnetohydrodynamic (MHD) model, such as ENLIL

© 2021. The Authors.
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

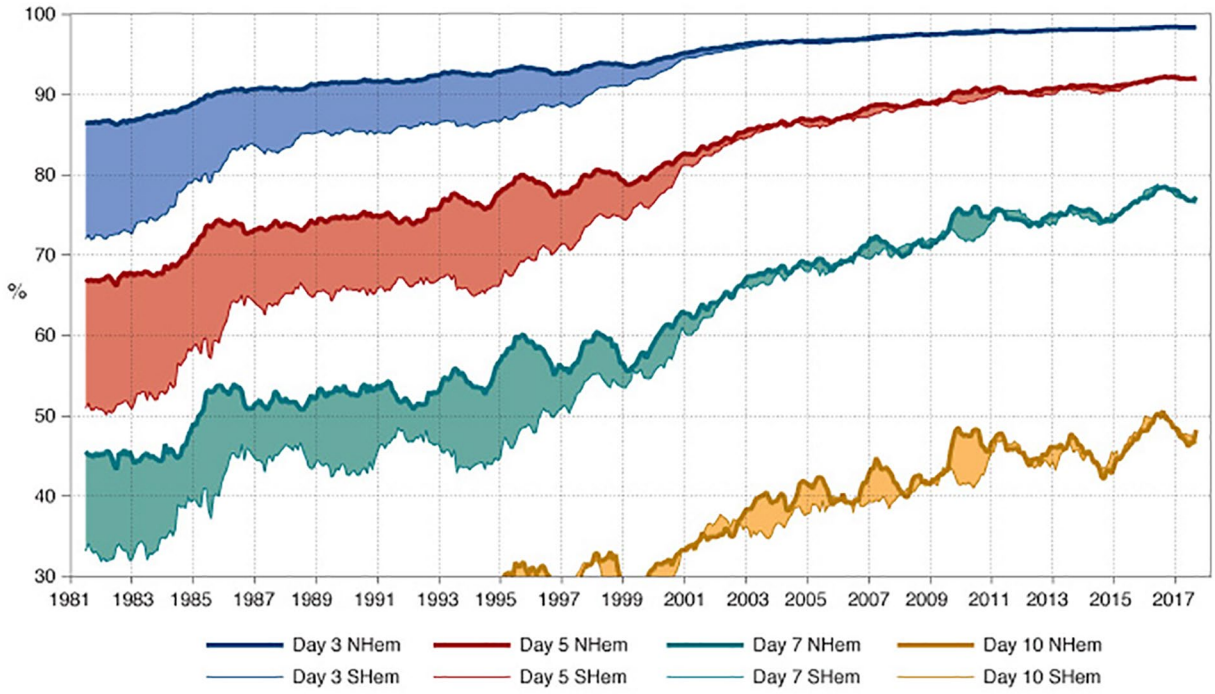


Figure 1. The 12-months running mean of the anamoly correlation for European Center for Medium-range Weather Forecasting (ECMWF's) 500 hPa geopotential height forecasts in the Earth's northern and the southern hemisphere at 3−day (blue), 5−day (red), 7−day (green) and 10−day (yellow) lead-times. The shaded regions show the difference between the northern (top) and southern (bottom) hemispheres, which is largely a result of differing observational coverage. Image reproduced from https://www.ecmwf.int/en/about/media-centre/news/2018/noaa-satellite-launch-20-years-ago-marked-start-new-era.

(Odstrcil, 2003), HelioMAS (Riley et al., 2012), HelioLFM (Merkin et al., 2016) or EUHFORIA (Poedts & Pomoell, 2017), with no further observational constraints.

In this study, however, we apply a further step using Data Assimilation (DA), assimilating available in-situ observations of the solar wind, before forecasting the solar wind. Data assimilation is the process by which observational data is incorporated into numerical models to improve knowledge of the trajectory, and uncertainties, of the model state. In operational meteorological models, data assimilation is commonly used to generate an optimal initial state to be used in subsequent forecasts through the use of prior knowledge of the state (e.g., a previous forecast) and observational data obtained from satellites/weather stations, etc. Data assimilation is considered a vital step in meteorological forecasting, reducing the impact of the "butterfly effect" (where small changes in initial conditions lead to large differences in model evolution), and hence improving the forecast skill. Improvements in numerical weather predictions have gone hand-in-hand with improvements in the implementation of the observation network and data assimilation methods into the models. This has led to huge improvements in the lead-times over which forecasts are accurate (Kalnay, 2003). This can clearly be seen in Figure 1. In the late nineties, a significant improvement was made to data assimilation methodology and satellite observations were included in the data assimilation, vastly increasing observational coverage. The assimilation of satellite observations led to vast improvements in Southern Hemisphere forecasts especially, leading to the closing of the gap in Northern and Southern Hemisphere forecast skill. Space weather forecasting has not yet exploited the full potential available from implementing data assimilation methods in forecasting models.

Previous DA work in space-weather forecasting has focused mostly on the photosphere, such as the Air Force Data Assimilative Photospheric Flux Transport (ADAPT) model (Arge et al., 2010), or the ionosphere (Bust & Mitchell, 2008; Durazo et al., 2017). Solar wind data assimilation is less mature and much work is still required in this area. Lang et al. (2017) connected the EMPIRE (Employing MPI for Researching Ensembles Browne & Wilson, 2015) data assimilation library to the ENLIL solar wind model in the first implementation of complex DA methodologies into a solar wind model. This work highlighted the sharp differences between terrestrial and solar wind DA, specifically that the solar wind is highly driven by the

Sun with relatively little chaotic behavior. Data assimilation updates were swept away if they did not (permanently) update the inner boundary and these updates did not have any influence at Earth unless the observation was downstream of Earth. This means that any future DA methodologies need to be developed to reflect this and allow DA updates to persist in the system and influence near-Earth space. Aiming to update the inner boundary conditions, Lang and Owens (2019) generated a variational data assimilation scheme with a simple solar wind propagation model (see Owens & Riley, 2017; Riley et al., 2012 for more details of the solar wind model) that could map information from in-situ observations at near-Earth orbit to the inner solar wind model boundary. This was shown to give improved and persistent updates to the solar wind over a 27-day period. However, this demonstration of improved solar wind reconstruction was not directly analogous to how the method would work in a forecast situation.

In this study, we will use the data assimilation methodology developed in Lang and Owens (2019), referred to here as the Burger Radial Variational Data Assimilation (BRaVDA) scheme, to investigate how data assimilation of in-situ observations affects solar wind forecast (strictly speaking, hindcast) skill. The BRaVDA scheme will be used to assimilate observations from the twin Solar-Terrestrial Relations Observatory (STEREO; Kaiser et al., 2008) spacecraft, in Earth-like orbits about the Sun, and the Advanced Composition Explorer (ACE; Stone et al., 1998) in near-Earth space. The BRaVDA scheme will be used to generate 27−day reanalysis sets over the operational lifetime of the STEREO-B satellite (November 01, 2007–September 30, 2014). For each 27-day DA analysis, a 27−day solar wind forecast will be produced using the HUXt (Heliospheric Upwind Extrapolation with time dependence) solar wind forecasting model (Owens, Lang, Barnard, et al., 2020), initialized with the results of the BRaVDA scheme. These forecasts will be presented and their accuracy will be verified against in-situ observations at STEREO and ACE. As a benchmark, a simple corotation forecast from the STEREO-B spacecraft (which is behind Earth in its orbit for the time-period of interest) will also be produced at L1 for comparison with ACE observations. To produce this corotation forecast, the solar wind structures will be assumed to be steady and to statically rotate with the Sun. Therefore, the observations at STEREO-B are predicted to be the solar wind conditions at L1 x time later, where x is the time it takes for the Sun to rotate between the longitudes of STEREO-B and L1 (Kohutova et al., 2016; Thomas et al., 2018).

This paper will describe the BRaVDA scheme and the HUXt forecasting model. Numerical experiments will then be presented, analyzing the seven-year BRaVDA reanalyzes and subsequent HUXt forecasts. Statistical analysis shall be performed and the improvements due to data assimilation will be discussed. In the final section, we shall make conclusions and discuss future developments.

2. The BRaVDA Scheme

The BRaVDA scheme is a variational data assimilation scheme (Lang & Owens, 2019), to map information from in-situ observations, typically near 1 AU, back to a solar wind model's inner boundary (in this case, 30 solar radii [r_S]). Figure 2 outlines qualitatively how the BRaVDA scheme works: our current estimate for the solar wind speed at the inner boundary is mapped out to the observation radius using our solar wind propagation model. Information from the observation (including its uncertainty) is then mapped back toward the Sun (Figures 2a–2b), where it is used to update the inner boundary of the solar wind model. The updated inner boundary is propagated back out using the solar wind model (Figure 2c), the results of which can then be used to produce forecasts of the solar wind speed at Earth (Figure 2d).

The solar wind model used by BRaVDA is based upon a solar wind speed propagation model developed by Riley and Lionello (2011), which maps the (two-dimensional) equatorial solar wind speed over the heliocentric domain from $30r_S$ to $236r_S$ from the Sun. Given the 2D nature of this model, we only represent a single heliolatitude. In reality there is typically a latitudinal offset between observing spacecraft and the desired forecast location. For example, the STEREO spacecraft are in the ecliptic plane, which is inclined by 7.25° to the heliographic equator. This introduces a representivity error (an error that arises as a result of representing an observation incorrectly in a model; other examples include an observation moved to the nearest gridpoint or a parameterization estimating a sub-gridscale process) into the system. A solar wind DA scheme in a full 3-dimensional solar wind model could relax this assumption, though the problem of determining the latitudinal representivity error of the observations remains.

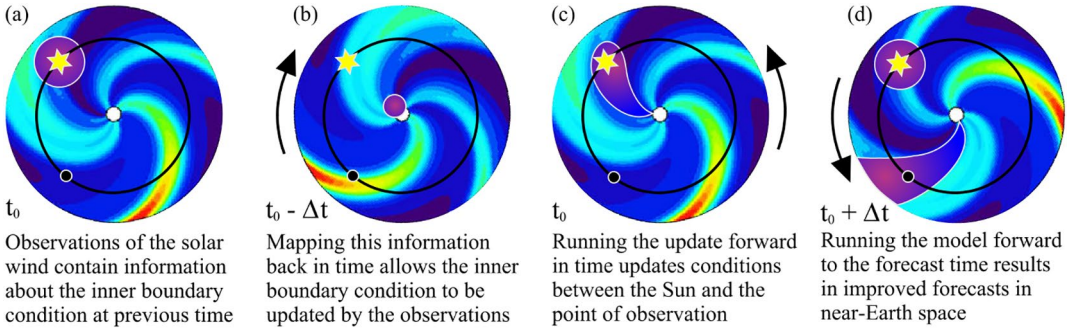


Figure 2. A schematic of how the Burger radial variational data assimilation (BRaVDA) scheme works from (Lang & Owens, 2019). The inner boundary (the white circle) is updated using information from observations from a position behind Earth in its orbit (the yellow star). The updated model conditions (the purple regions) persist in the model domain such that they impact forecasts at Earth's location (the black circle). Updating the inner boundary stops updates from being advected out of the model domain by the supersonic radial solar wind flow. (a) Observations of the solar wind contain information about the inner-boundary condition at previous time. (b) Mapping this information back in time allows the inner boundary to be updated by the observations. (c) Running the update forward in time updates conditions between the Sun and the point of observation. (d) Running the model forward to the forecast time results in updated forecasts in near-Earth space.

In data assimilation, we aim to reconstruct the true state vector, $\mathbf{x^t}$, an N_x-dimensional vector that contains the values of the quantities of interest at all gridpoints, using prior knowledge about the system (e.g., a previous forecast), $\mathbf{x^b}$, and observations, $\mathbf{y}$. The variables within a state vector for meteorological applications can include the precipitation over Africa, the mean sea level pressure over France, wind speed/direction in Llanfairpwllgwyngyllgogerychwyrndrobwllllantysiliogogogoch, etc. depending upon the purpose of the numerical model. In order to do this, the errors in the prior and observational information is typically assumed to be normally distributed and the numerical model is assumed to be linear. This allows a cost function to be defined, based upon the relative uncertainties in the prior errors, $\mathbf{B}$, and the observation errors, $\mathbf{R}$. Through construction, the state that minimizes the cost function maximizes the posterior distribution (the probability distribution of the state given the observations, $p(\mathbf{x}|\mathbf{y})$, which is proportional to the likelihood distribution, $p(\mathbf{y}|\mathbf{x})$, multiplied by the prior distribution, $p(x)$, through Bayes' Theorem (Bayes & Price, 1763), hence giving us the mode of the posterior distribution.

In the BRaVDA system, the state vector that we wish to estimate is the solar wind speed at the inner-boundary at all longitudinal points (i.e., a 128-dimension vector, $\mathbf{v_0}$). The prior error covariance matrix, $\mathbf{B}$, is estimated by an ensemble of inner boundary solar wind speeds generated using the Magnetohydrodynamics Around a Sphere (MAS; Linker et al., 1999) model (further details about how this ensemble is created can be found at Owens & Riley, 2017) as:

$$\mathbf{B} \approx \frac{1}{M-1}\sum_{m=1}^{M}\left[\left(\mathbf{v}_0^{(\mathbf{m})}-\overline{\mathbf{v}_0^{\mathbf{M}}}\right)\left(\mathbf{v}_0^{(\mathbf{m})}-\overline{\mathbf{v}_0^{\mathbf{M}}}\right)^T\right], \tag{1}$$

where $\overline{\mathbf{v}_0^{\mathbf{M}}} = \frac{1}{M}\sum_{m=1}^{M}\left[v_0^{(m)}\right]$ is the MAS ensemble mean and $M = 576$ is the number of ensemble members. The prior state, $\mathbf{v^b}$, is a solar wind speed randomly drawn from the normal distribution, $\mathcal{N}(\overline{\mathbf{v}_0^{\mathbf{M}}}, \mathbf{B})$, defined by the probability distribution function:

$$p(x) = \left(2\pi\right)^{\frac{-N_x}{2}}\det\left(\mathbf{B}\right)^{\frac{-1}{2}}\exp\left(-\frac{1}{2}\left(x-\overline{\mathbf{v}_0^{\mathbf{M}}}\right)^T B^{-1}\left(x-\overline{\mathbf{v}_0^{\mathbf{M}}}\right)\right), \tag{2}$$

where $\det\left(\mathbf{B}\right)$ is the determinant of $\mathbf{B}$.

The BRaVDA scheme also utilizes the Strong Constraint (Howes et al., 2017), which makes an additional assumption that the numerical model is perfect (i.e., the numerical model contains no errors). This typically

produces a poorer result than a system that allows for model error (the Weak Constraint), however incorporating model error is a complex operation (Evensen et al., 1998; Lang et al., 2016), and is beyond the scope of the present study. The BRaVDA Cost Function ($\mathcal{J}\left(\mathbf{v_0}\right)$, a full derivation of which can be found in Appendix A of Lang & Owens 2019), that is to be minimized, is given by:

$$\mathcal{J}\left(\mathbf{v_0}\right) = \frac{1}{2}\left(\mathbf{v_0} - \mathbf{v^b}\right)^T \mathbf{B}^{-1}\left(\mathbf{v_0} - \mathbf{v^b}\right) + \frac{1}{2}\sum_{k=0}^{N_y}\left[\left(\mathbf{y_k} - \mathcal{H}_k(\mathbf{v_0})\right)\mathbf{R_k}^{-1}\left(\mathbf{y_k} - \mathcal{H}_k(\mathbf{v_0})\right)\right], \tag{3}$$

where $\mathbf{y_k}$ is the kth observation, N_y is the number of observations, and $\mathcal{H}_k$ is the observation operator that is a function that maps the state to the observation space (i.e., $\mathcal{H}_k\left(\mathbf{x_0}\right)$ tells us what the numerical model "thinks" the observation should be). The observations used in this study are in-situ observations from spacecraft such as STEREO and ACE, but can also be expanded to include remotely sensed observations such as Interplanetary Scintillation (IPS; e.g., Breen et al., 2006) or Heliospheric Imagers (HI; e.g., Eyles et al., 2009), albeit these require a far more complex observation operator than is currently utilized for in-situ observations in BRaVDA. $\mathbf{R_k}$ is the observation error covariance matrix and approximating it fully is a large research topic in its own right (Owens, Lang, Riley, et al., 2020), as it will contain representivity errors. The BRaVDA scheme currently only supports diagonal $\mathbf{R}$ matrices that are either constant or proportional to the prior state, which implies that all the observations are assumed to be independent of each another and with equal uncertainty leading to potential inaccuracies (the addressing of which is beyond this study).

The $\mathbf{v_0}$ that minimizes the cost function is referred to as the analysis or posterior state and is given the notation $\mathbf{v^a}$ in this manuscript. Obtaining $\mathbf{v^a}$ is nontrivial and can be done by using a finite differences approach or by evaluating the gradient of the cost function directly (Bannister, 2007), however, for higher-dimensional systems, these methods are not practical. A more efficient method of calculating the gradient is to use the adjoint method, a method of sensitivity analysis that efficiently computes the gradient of a function (Errico, 1997). The adjoint equations for the BRaVDA scheme are derived explicitly in Lang and Owens (2019). The BRaVDA scheme is capable of producing a 27-day analysis assimilating STEREO-A, B and ACE data over the period (~128 obs each) in 6 min 36 $s \pm 6.1$ s on a home laptop with 8 GB of RAM and an Intel Core i5-7300HQ 2.5 GHz processor. Thus the scheme is suitable for near-real time forecasting.

This adjoint method is extremely powerful once computed and its implementation enabled the great improvements in forecast skill seen by ECMWF in the 1990s (see Figure 1). However, the adjoint equations are unique for each model, which means that they can be difficult to produce for complex models. Future studies will use DA methodologies that remove the need to use an adjoint and may allow for more nonlinear behavior, such as 4DEnsVar (Amezcua et al., 2017; Goodliff et al., 2017), or a smoother-based ensemble DA method, such as the Iterative Ensemble Kalman Smoother (Bocquet & Sakov, 2014).

3. The HUXt Solar Wind Model

The HUXt Heliospheric Upwind eXtrapolation (HUXt) with time dependence solar-wind model is a reduced-physics solar-wind model based upon Burgers' inviscid equation (Owens, Lang, Barnard, et al., 2020). In order to make the solar-wind model as computationally efficient as possible, a number of physical assumptions and approximations are required, beyond those typically used in three-dimensional magnetohydrodynamics. These are that in the heliosphere, pressure gradient and gravitation terms are small compared with the flow momentum and hence can be neglected. A further simplification is that we only consider variations in the radial direction. See also Riley and Lionello (2011).

Using 40+ years of hindcasts of ambient solar wind, the HUXt model was shown to emulate the full three-dimensional MHD solution using the same $30r_S$ boundary conditions, with solar-wind speeds throughout the simulation domain agreeing to within 6% (Owens, Lang, Barnard, et al., 2020). The biggest difference between the two solutions is that HUXt produces sharper boundaries between slow and fast wind than the MHD solutions of HelioMAS and ENLIL (see Owens, Lang, Barnard, et al., 2020 for more details).

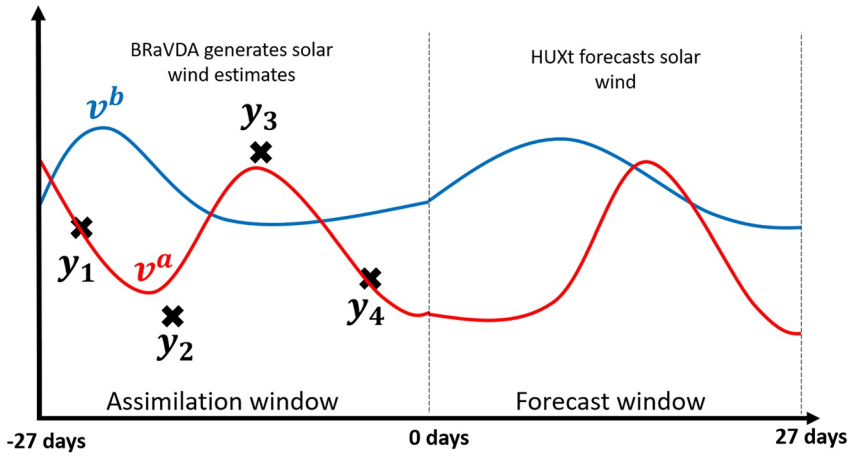


Figure 3. Schematic showing how Burger radial variational data assimilation (BRaVDA) and heliospheric upwind eXtrapolation (HUXt) are used to generate solar-wind estimates and forecasts. The prior solar wind ($\mathbf{v^b}$, blue), is generated using the radial solar-wind propagation scheme defined from Riley and Lionello (2011). The BRaVDA scheme uses Equation 3 and the observations ($\mathbf{y_k}$, black crosses), to generate an analysis ($\mathbf{v^a}$, red), for the 27 days prior to the forecast start-time (the assimilation window). $\mathbf{v^b}$ and $\mathbf{v^a}$ are used to initialize HUXt forecasts of the solar-wind speed for the next 27 days (forecast window).

4. Experimental Setup

For each consequent 27-day period from January 01, 2007 to November 02, 2014, the MAS ensemble (available from http://www.predsci.com/mhdweb/home.php) is constructed for that period and the prior state, $\mathbf{v^b}$, and prior error covariance matrix, $\mathbf{B}$, are approximated as described in Section 2.

The observation error covariance matrix, $\mathbf{R}$, is also generated for each 27-day window. As in Lang and Owens (2019), we specify the observation error covariance matrix as a diagonal matrix with standard deviation approximated as 10% of the mean prior solar-wind speed at the observation radius. This is a somewhat arbitrary value, as the true $\mathbf{R}$ matrix will include a term that varies depending upon the observation's latitudinal offset from Earth and phase of solar cycle, etc., but it provides a starting point from which we can progress and test such assumptions. This gives:

$$\begin{aligned} R_{kk} &= (0.1\overline{\mathbf{v^b_{rk}}})^2 \\ &= (0.1\sum_{i=1}^{N_\phi}\left[v_r^b(\phi_i)\right])^2, \end{aligned} \tag{4}$$

where $0 \le k \le N_y$ is the observation number and $N_\phi = 128$ is the number of longitude points.

The BRaVDA scheme is run with these parameters for consecutive 27-day windows (each approximating 1 solar rotation from Earth's perspective) spanning the seven-year operational window of STEREO-B, assimilating all available observations from STEREO-A, STEREO-B and ACE. This produces initial conditions for forecasting the next 27-day period. The HUXt model is used to produce the forecast for the next 27-day window, with no additional observational constraints. This is summarized in Figure 3.

5. Results

5.1. Forecasting Results Over the Whole 7−Year Period

This section describes the results of the forecasts produced using the data assimilation over the whole seven-year window. As mentioned in the previous sections, 27-day forecasts of the solar wind are produced for the prior state (generated from an ensemble of MAS runs) and the posterior state (the resultant solar wind from the DA). At ACE, we also compared against the STEREO-B corotation forecast. This involves time-shifting STEREO-B observations by Δt, given by:

$$\Delta t = \frac{\Delta \phi_B}{\Omega_{ROT}}, \tag{5}$$

where $\Delta\phi_B$ is the longitude difference between STEREO-B and Earth. See Thomas et al. (2018); Kohutova et al. (2016) for more details.

For each 27-day forecast window, Root Mean Squared Error (RMSE) is computed over the whole 27-day forecast window, that is:

$$RMSE_{27,y} = \sqrt{\frac{1}{N_y}\sum_{i=1}^{N_y}\left[\left(y_i - x_i\right)^2\right]}, \tag{6}$$

where N_y corresponds to the number of STEREO-A, STEREO-B and ACE observations during the 27-day forecast window, y_i correspond to the STEREO-A, STEREO-B or ACE observations, and the x_i are the prior/posterior state, at the relevant satellite's time/location.

This measures the accuracy of the solar wind forecasts over the whole 27-day forecast window. RMSE is a simple point-by-point analysis, which can be problematic for distinguishing between forecasts with insufficient variability and forecasts with small timing errors in events (Owens, 2018). However, in this particular application, it nevertheless provides a useful initial diagnostic of performance. More detailed forecast validation will be the subject of future research.

Figure 4 shows the 27-day forecast RMSEs against time. The posterior forecast RMSEs are lower than the prior forecasts RMSEs for the vast majority of windows, at all spacecraft locations. To be more specific, for the 105 individual 27-day forecasts generated: the posterior RMSEs are lower than the prior's for 95 forecasts at STEREO-A, 92 forecasts at STEREO-B and 97 forecasts at ACE. This result indicates that the improvements made by the DA in the assimilation window are carried into the forecast window by the HUXt model.

In the first half of the seven-year period (i.e., before 2011), we note that the prior state has relatively high RMSEs. This is likely due to the Sun being in the solar minimum phase, meaning that greater representivity errors due to latitudinal offsets are observed (see Owens, Lang, Riley, et al., 2020 for more details). Also the solar wind speed is generally higher and more (spatially) variable at this time (Owens, 2020). We can see that the data assimilation is able to produce more accurate solar wind speed forecasts at all observation locations for the vast majority of this time period.

In the latter half of the seven-year period, the solar maximum phase implies the prior state has lower RMSEs due to reduced latitudinal representivity errors and lower solar-wind speed/variability (Owens, Lang, Riley, et al., 2020). In addition, solar maximum is associated with more transient events. These facts lead to a reduction in the improvement in DA forecast RMSEs from the posterior forecast (as the HUXt forecasts here assume the Sun is in steady-state). Nonetheless, the DA still leads to improvements in the forecast solar-wind speeds in comparison to the prior forecasts.

In the center of the seven-year period (end of 2010/start of 2011), there appears to be regions where the prior, posterior and STEREO-B corotation forecasts produce very similar RMSEs. This is most likely due to the start of solar maximum and STEREO-B being far from Earth (in both longitude and latitude). This corresponds to errors from the longitudinal offsets reducing as we move into solar maximum, but corotation lead time and increased transients/solar-wind complexity compensating for these effects. These effects are difficult to disentangle and are currently being investigated to form part of a future study.

At ACE, we also compare the posterior forecast to the corotation forecast from STEREO-B. The two forecasts are comparable when looking at the 27-day forecast RMSE. This is to be expected somewhat, as the data assimilation analysis incorporates the observation at STEREO-B within it. Nonetheless, the posterior forecast at ACE has lower 27-day RMSEs for 63 out of 105 windows (that is, the posterior forecasts are better for 60% of windows). Note that the STEREO-B forecast does not exist for three windows, due to insufficient observations (but the posterior forecast still exists, as it can be based on ACE and STEREO-A data).

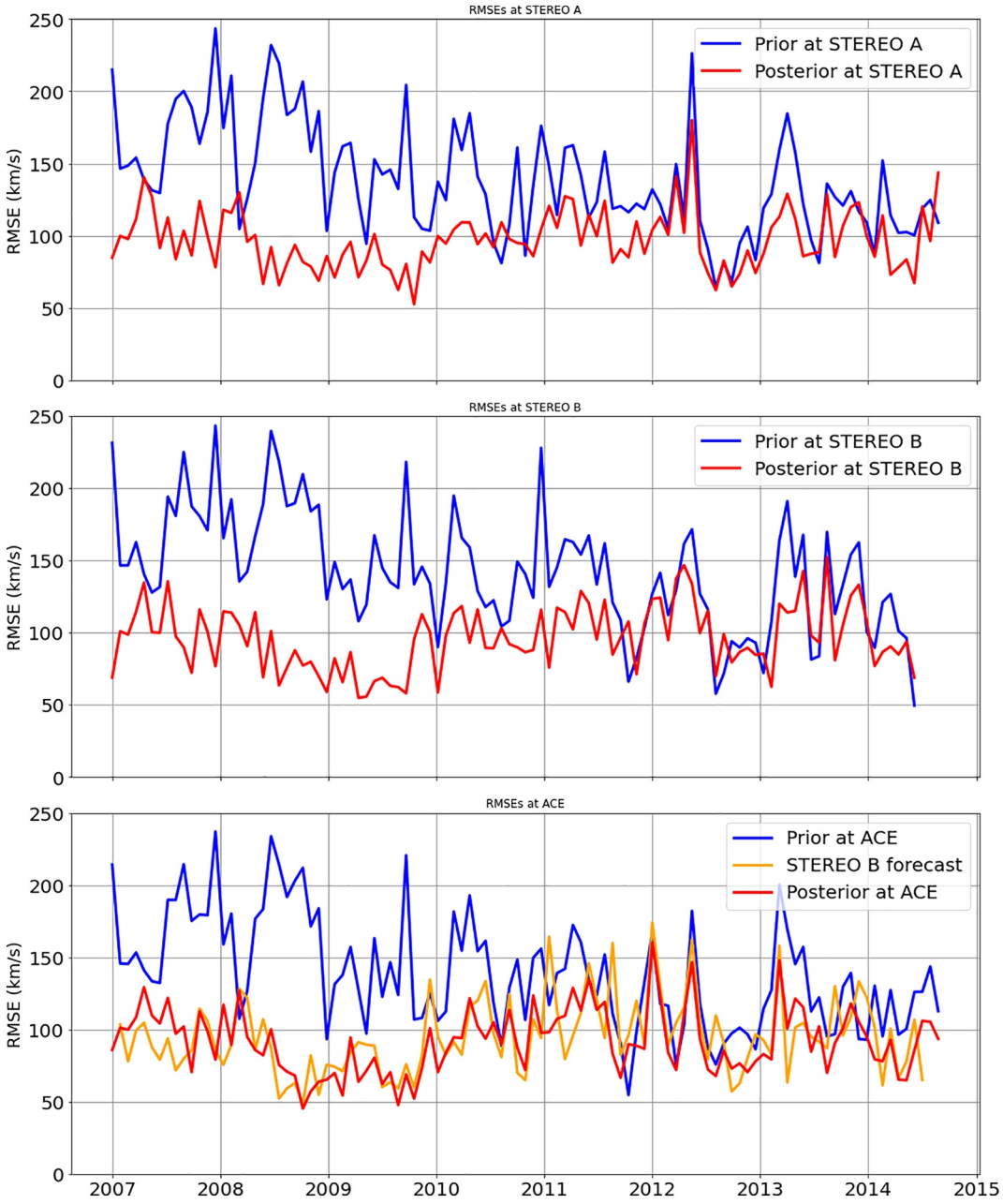


Figure 4. Root mean square error (RMSE) of the solar wind speed averaged over each 27-day window. The top row gives the RMSEs at STEREO-A, the middle row gives the RMSEs at STEREO-B and the bottom row gives the RMSEs at the Advanced Composition Explorer (ACE). The blue lines indicate the prior (no data assimilation [DA]) forecast RMSEs, the red lines indicate the posterior (including DA) forecast RMSEs and the orange line on the bottom plot gives the STEREO-B corotation forecast at ACE.

Figure 5 shows the mean forecast RMSEs as a function of forecast lead-time at STEREO-A, STEREO-B and ACE. The mean RMSE for each forecast lead-time (in days) is calculated as:

$$RMSE_{t,y} = \sqrt{\frac{1}{N_w}\frac{1}{N_t}\sum_{l=1}^{N_w}\sum_{k=s_t}^{e_t}\left[\left(y_{l,k} - x_{l,k}\right)^2\right]}, \tag{7}$$

where N_w is the number of 27-day windows (105 in total), $N_t = e_t - s_t$ is the number of observations during the tth day, $y_{l,k}$ is the kth observation in the lth 27−day window and $x_{l,k}$ is the state corresponding to $y_{l,k}$.

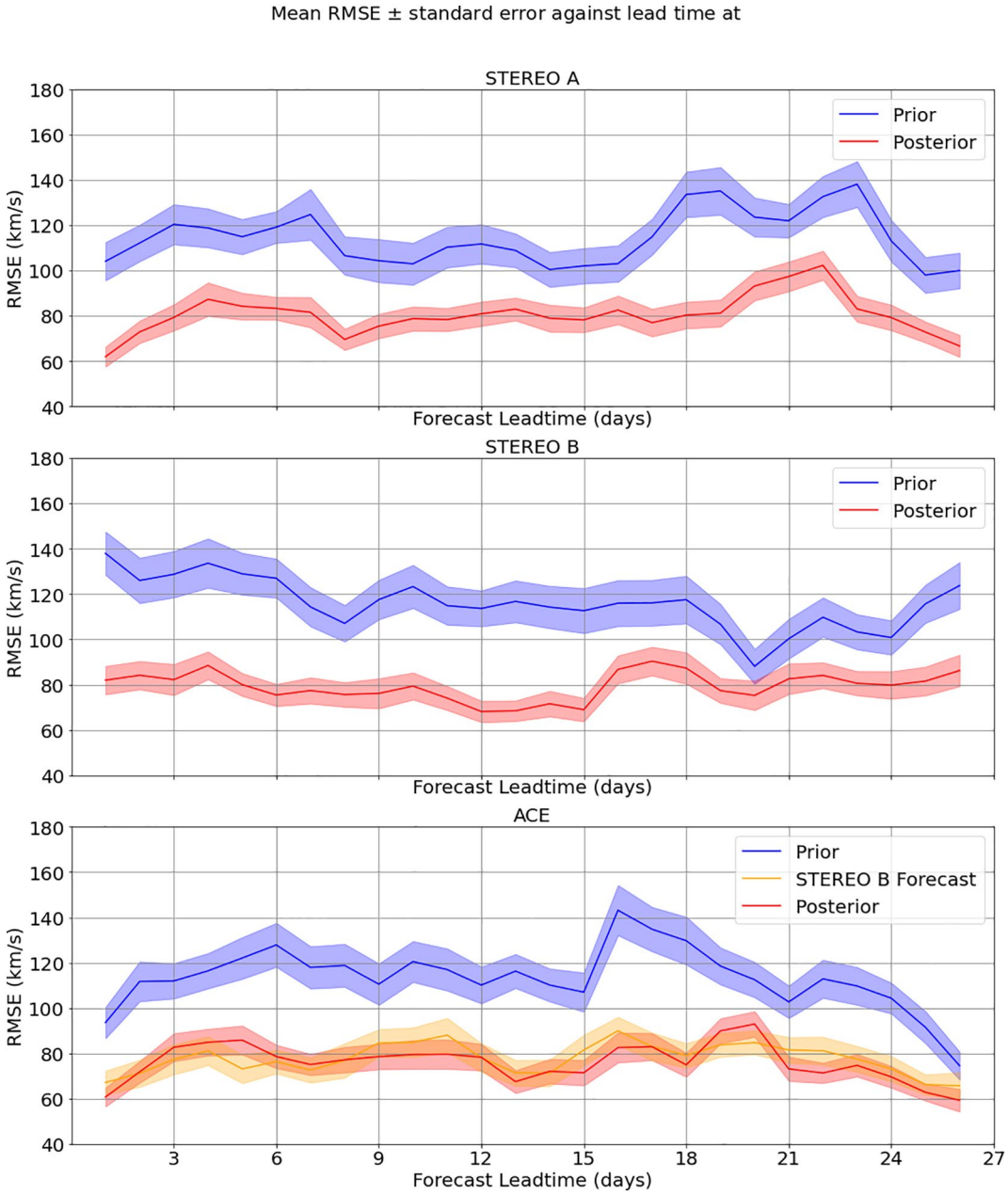


Figure 5. The average forecast root mean squared Errors (RMSEs) of the solar-wind speed plotted against forecast lead-time, for the whole 2007–2014 interval. The top row corresponds to the RMSEs at STEREO-A, the middle row corresponds to the RMSEs at STEREO-B and the bottom row corresponds the RMSEs at the Advanced Composition Explorer (ACE). The blue lines indicate the prior (no data assimilation [DA]) forecast RMSEs, the solid red lines indicate the posterior (DA) forecast RMSEs and the solid orange line on the bottom plot gives the STEREO-B corotation forecast at ACE. The solid lines indicate mean RMSEs while the shaded regions span one standard error.

Looking at all observation locations in Figure 5, the posterior solar wind speed forecast has lower mean RMSE than the prior solar wind speed forecast for all lead-times by about 30–50 km/s. Table 1 summarizes this, with the mean RMSE reduction from prior to posterior being 35.976 km/s over all observation locations (equivalent to a 31.4% reduction in forecast RMSE over all observations). It is also the case that the standard errors for the posterior and STEREO-B corotation forecasts tend to be lower than that of the prior. This is further emphasized by the mean standard errors shown in Table 1 that indicate that the posterior and STEREO-B corotation have standard errors of 5.628 km/s and 5.951 km/s respectively, whereas the prior

Table 1
Table Showing Mean Root Mean Squared Errors (RMSEs) ± Their Standard Errors for the Prior, Posterior and STEREO-B Corotation Forecasts at Each Observation Location Over the Whole-years Window

Observation location	Prior RMSE (km/s)	Posterior RMSE (km/s)	STEREO-B corotation RMSE (km/s)
STEREO-A	114.507 ± 8.618	80.419 ± 5.626	N/A
STEREO-B	115.965 ± 9.111	79.384 ± 5.870	N/A
ACE	113.349 ± 8.416	76.089 ± 5.387	77.752 ± 5.951
All avg.	114.607 ± 8.715	78.631 ± 5.628	N/A

Abbreviation: ACE, Advanced Composition Explorer.

forecast has standard errors averaging 8.715 km/s. This indicates the accuracy of the DA and corotation forecasts is more consistent than the prior. At ACE, the STEREO-B corotation forecast RMSE is very similar to the posterior forecast RMSE over all lead-times. However, the posterior forecast does have a slightly lower mean RMSE (see Table 1) and standard error than the STEREO-B forecast. Furthermore, the mean posterior forecast RMSE is lower for 16 lead-times out of 27 considered (60%).

In Figure 6, the distribution of RMSEs for the forecasts is shown as a function of the absolute heliographic latitude difference between STEREO-B and ACE/Earth. The STEREO-B corotation forecasts are seen to be positively correlated, with a gradient of 1.593 $kms^{-1}deg^{-1}$ and correlation coefficient of 0.301. This is somewhat expected (Owens, Lang, Riley, et al., 2020) and will be explored in more detail later in the study (in particular, we note the possibility for aliasing between latitude and longitude separations, as well as with the solar cycle). Furthermore, the *p*-value of 0.012 indicates that this correlation is significant at the 95% confidence level. As the latitude difference between STEREO-B and ACE becomes greater, the STEREO-B observations are sampling solar wind that emanate from different regions of the Sun (e.g., spacecraft could be sampling solar wind from different coronal holes). On the other hand, the posterior forecast does not display this correlation, the p-value indicating that the correlation found is not significant and the line of best-fit is a flat line (gradient of posterior line of best fit is 0.138 $kms^{-1}deg^{-1}$). This means that when STEREO-B and ACE are at different latitudes, the data assimilation is able to remove this systematic error through combination with the other observation sources (ACE and STEREO-A). However, this does also work against the posterior forecast, as it does not necessarily lead to improvements over the STEREO-B corotation forecast when the latitude differences are small. This could be due to STEREO-A being at a larger latitude offset from ACE, "pulling" the posterior forecast away from the ACE data.

In the next sections, we shall investigate how the data assimilation performs in different solar-cycle regimes and where DA can give the greatest improvements when compared to the STEREO-B forecast.

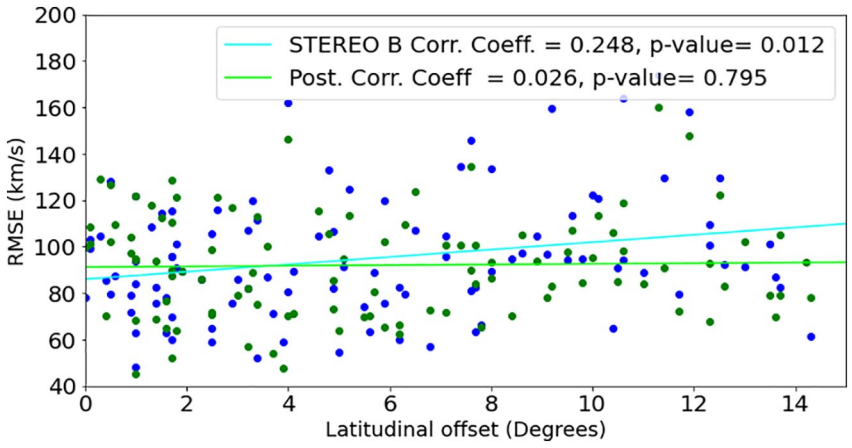


Figure 6. The 27-day forecast root mean squared errors (RMSEs) of the solar-wind speed plotted against the absolute difference between STEREO-B and the Advanced Composition Explorer's (ACE's) latitude, for the entire seven-year period. The blue points correspond to the STEREO-B corotation forecast and the green points relate to the data assimilation (DA) posterior forecast. The blue and green lines represent the line of best-fit for the corotation and posterior forecasts, respectively. The legend gives the Pearson's product-moment correlation coefficient indicating the strength of the correlation between the data and the line of best-fit, along with its associated p-value. The STEREO-B corotation forecast RMSE has a correlation coefficient of 0.248 with STEREO-B latitude, which is significant at the level of $p = 0.012$.

Table 2
Table Showing Mean Root Mean Squared Errors (RMSEs) ± Their Standard Errors for the Prior, Posterior and STEREO-B Corotation Forecasts at Each Observation Location During Solar Minimum

Observation location	Prior RMSE (km/s)	Posterior RMSE (km/s)	STEREO-B corotation RMSE (km/s)
STEREO-A	127.146 ± 12.270	78.056 ± 7.039	N/A
STEREO-B	131.010 ± 12.712	75.096 ± 7.106	N/A
ACE	127.145 ± 12.209	73.386 ± 6.737	73.802 ± 7.122
All avg.	128.434 ± 12.397	75.513 ± 6.961	N/A

Abbreviation: ACE, Advanced Composition Explorer.

5.2. Forecasting the Solar-Wind Speed During Different Solar-Cycle Phases

Solar activity is expected to influence the forecast accuracy of both the DA and corotation forecasts. During solar maximum there is an increased number of transient events, such as coronal mass ejections (CMEs), and a more rapidly evolving solar wind structure. Conversely, solar minimum results in fewer transients and a more slowly evolving solar wind structure. Solar-wind speed is highly structured in latitude near solar minimum. The solar cycle (from one solar minimum to the next) is approximately 11 years. In this section, we denote the phase of the solar cycle as $s_C \in [0, 1]$. 0 denotes the start of the solar cycle (at solar minimum), 1 denotes the end of the solar cycle (at solar minimum) and 0.5 is the midpoint of the solar cycle. Solar maximum typically occurs prior to $s_C = 0.5$.

We split forecasts of the solar-wind speed into two groups: those around solar minimum ($sC \geq 0.7$ or $sC < 0.2$) and those forecast around solar maximum ($0.2 \leq sC < 0.7$). From this definition, solar minimum corresponds to dates between January 01, 2007 (the start of the experiments) and March 20, 2011 and solar maximum from March 20, 2011 to November 02, 2014 (the end of the experiments). In this way, we can diagnose how the solar-wind forecasts behave during these two regimes of the solar cycle. Unfortunately, however, this will not be a like-for-like comparison as the STEREO-B spacecraft is moving away from Earth at a rate of 22.5°/yr, therefore the corotation forecasts produced later in the seven-year period (i.e., during the solar maximum phase) will be of longer lead-time than those produced earlier. This means that we would expect the RMSEs to be larger during the solar maximum phase than those in solar minimum, due to the fact that they occur when STEREO-B is further away and hence forecasts have a longer lead-time. This is in addition to the more rapidly evolving solar wind structure and increased number of solar wind transients near solar maximum.

The seven years that we have STEREO-B observations and have run these solar-wind speed forecasts, from 2007 to 2014, coincide with the end of Solar Cycle 23 (August 1996–December 2008, $s_C > 0.84$) and the first half of Solar Cycle 24 (December 2008–May 2020, $s_C < 0.51$). This means that we have far more forecasts occurring around solar minimum than those around solar maximum.

Tables 2 and 3 show that during the solar minimum regime, the prior solar-wind speed RMSEs and uncertainties are greater than during the solar maximum regime, as seen in the previous time series plots. During solar minimum, the posterior forecast and the STEREO-B corotation forecast show a greater improvement over the prior state and greater reduction in uncertainty.

Figure 7 shows that during solar minimum, the posterior and the STEREO-B corotation forecasts show a very similar RMSE pattern for all lead-times, with a mean RMSE difference of 0.086 km/s over all lead-times. Nonetheless, during solar minimum, the mean posterior forecast RMSEs are lower than the STEREO-B corotation forecast RMSEs for 17 considered lead-times out of 27. During solar maximum (see Figure 8), however, the mean RMSE of the STEREO-B corotation forecast is higher than the posterior forecast RMSE between 20 and 23 days (by an average of 17.393 km/s) and with greater uncertainty (mean standard deviation is 13.95 km/s during this period). Despite this large difference in mean RMSEs, the mean STEREO-B corotation forecasts are better than the posterior forecasts for 11 of the 27 lead-time days. On the whole, Figures 7 and 8 show that the STEREO-B corotation forecast and the posterior forecast RMSEs are broadly similar during solar minimum and maximum, exhibiting similar means and uncertainty patterns.

Table 3
Table Showing Mean Root Mean Squared Errors (RMSEs) ± Their Standard Errors for the Prior, Posterior and STEREO-B Corotation Forecasts at Each Observation Location During Solar Maximum

Observation location	Prior RMSE (km/s)	Posterior RMSE (km/s)	STEREO-B Corotation RMSE (km/s)
STEREO-A	99.057 ± 10.961	83.287 ± 8.732	N/A
STEREO-B	96.992 ± 11.893	84.806 ± 9.521	N/A
ACE	96.967 ± 10.610	79.291 ± 8.294	82.541 ± 9.554
All avg.	97.672 ± 11.155	82.461 ± 8.849	N/A

Abbreviation: ACE, Advanced Composition Explorer.

By contrast, Figure 9 shows a very different picture of forecasting skill as a function of latitude offset during solar minimum and solar maximum. Figure 9a shows the 27-day RMSEs of the STEREO-B corotation forecast and posterior forecasts against latitude offset during solar minimum. It can be seen that there is a significant (at the 95% confidence level) positive

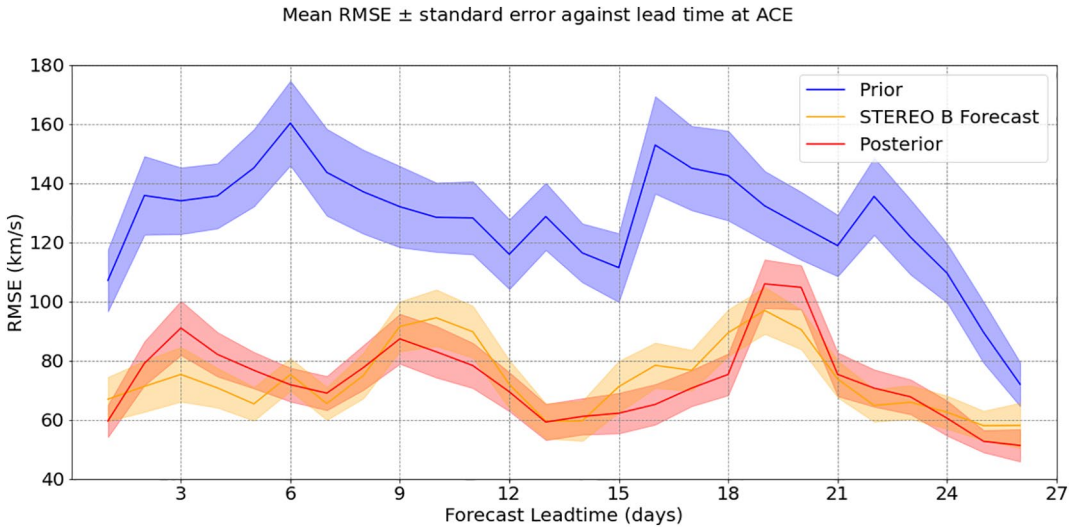


Figure 7. The mean root mean squared Errors (RMSEs) of the solar-wind speed forecasts plotted against lead-time during solar minimum ($sC < 0.2$ or $sC \geq 0.7$) at the Advanced Composition Explorer (ACE). The blue lines indicate the prior (no data assimilation [DA]) forecast RMSEs, the red lines indicate the posterior (including DA) forecast RMSEs and the orange line gives the STEREO-B corotation forecast RMSEs at ACE. The shaded regions indicate one standard error around the mean.

correlation between STEREO-B corotation forecast RMSEs and STEREO-B's latitude difference from Earth during solar minimum (with gradient of 3.078 $kms^{-1}deg^{-1}$). By comparison, during solar maximum, despite STEREO-B reaching a greater latitude offset from ACE, the STEREO-B corotation forecast RMSE is uncorrelated with the latitude offset of STEREO-B (in fact, the line of best-fit's gradient is -0.013 $kms^{-1}deg^{-1}$, and is not significantly anti-correlated with latitude). The reason for this almost counter-intuitive result can be explained by looking at the solar-wind structure during solar minimum and solar maximum Owens, Lang, Riley, et al. (2020). During solar minimum, the slow-wind band (and heliospheric current sheet) is closely aligned with the heliographic equator (±20°), this means that for a relatively small change in latitude, large deviations in solar-wind structure are observed. Therefore, during solar minimum it is more likely that

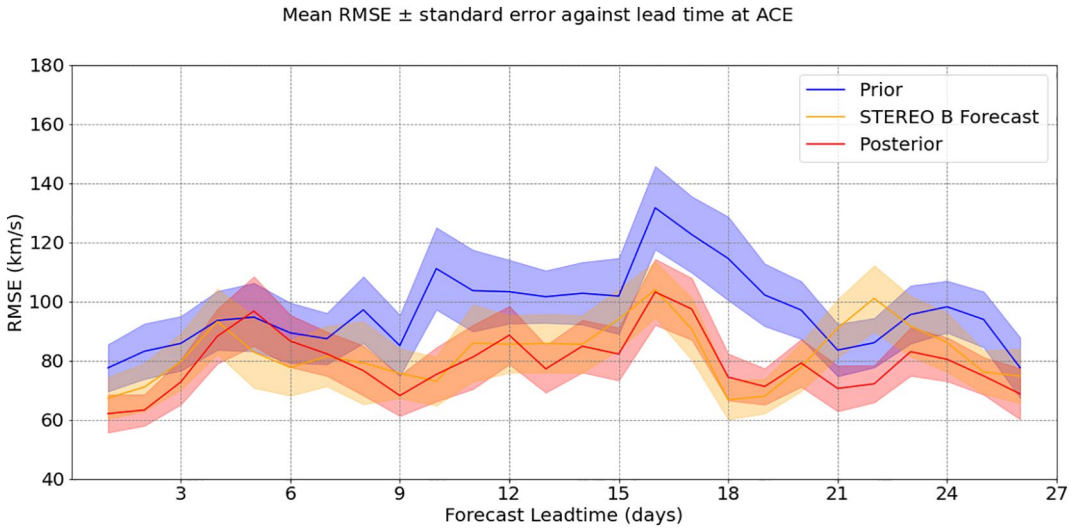
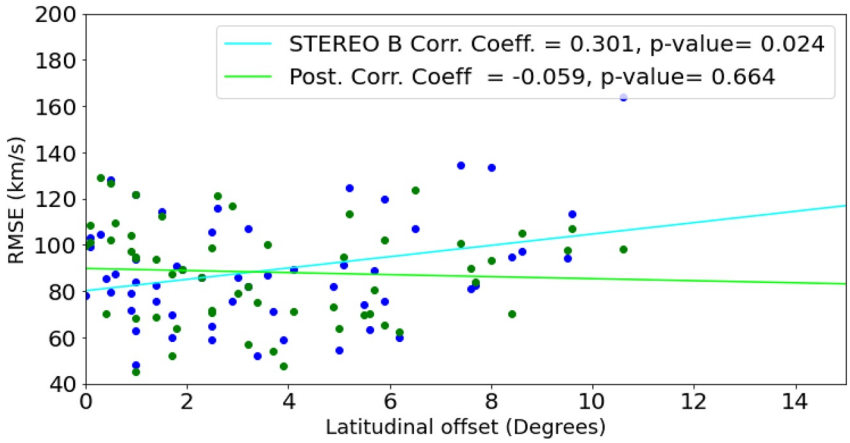
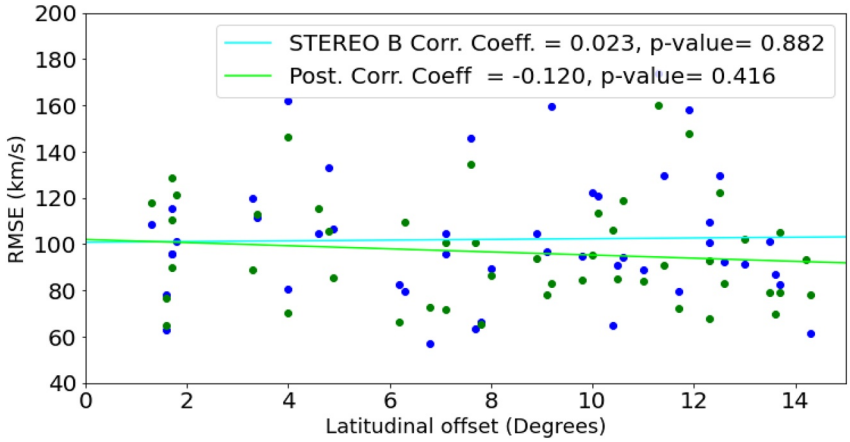


Figure 8. The mean root mean squared errors (RMSEs) of the solar-wind speed forecasts plotted against lead-time during solar maximum ($0.2 \leq sC < 0.7$) at the Advanced Composition Explorer (ACE). The blue lines indicate the prior (no data assimilation [DA]) forecast RMSEs, the solid red lines indicate the posterior (including DA) forecast RMSEs and the solid orange line gives the STEREO-B corotation forecast RMSEs at ACE. The shaded regions span one standard error about the mean.



(a) Solar minimum ($sC < 0.2$ or $sC \geq 0.7$)



(b) Solar maximum ($0.2 \leq sC < 0.7$)

Figure 9. The 27-day root mean squared errors (RMSEs) of the solar-wind speed forecasts plotted against the absolute difference between STEREO-B and the Advanced Composition Explorer's (ACE's) latitude during solar minimum (top panel) and solar maximum (bottom panel). The blue points correspond to the STEREO-B corotation forecast and the green points relate to the posterior forecast. The blue and green lines represent the line of best-fit for the corotation and posterior forecasts, respectively. The legend gives the Pearson's product-moment correlation coefficient indicating the strength of the correlation between the data and the line of best-fit, along with its associated p-value. A significant correlation between latitude offset and RMSE is only observed for the STEREO-B forecast to ACE during solar minimum (at the $p = 0.024$ level).

when STEREO-B is more latitudinally offset from ACE, it will also be more offset with respect to solar-wind structure. In comparison, during solar maximum, slow wind extends to high latitude. This means that there are generally smaller gradients of solar-wind structure with latitude, hence STEREO-B and ACE are more likely to be observing the same solar-wind structures despite their latitudinal offset. For more details on this phenomena, see (Owens, Lang, Riley, et al., 2020).

Reassuringly, the data assimilation is able to compensate for this latitudinal correlation and produce a more consistent forecast for all latitudinal offsets between STEREO-B and ACE. It can be seen that during solar minimum, when this effect is greatest, the posterior forecast's line of best fit is almost flat (its gradient is 0.315 $\text{kms}^{-1}\text{deg}^{-1}$) with very little correlation. As noted before in the plot of latitudinal offset against RMSE for all the data, this does mean that higher RMSE can be observed for the posterior forecasts when STEREO-B's latitude is close to that of Earth. This further illustrates the need to investigate and incorporate the latitudinal representivity into the observation error covariance matrix when performing the data assimilation. During solar maximum, the posterior is barely correlated with the latitudinal offset, nonetheless, it can

be seen that there is a consistent reduction in posterior forecast RMSE when compared to the STEREO-B corotation forecast.

6. Discussion and Conclusions

The experiments shown in this study illustrate how data assimilation can be used to produce superior forecasts of near-Earth solar-wind speed up to one solar rotation (27 days) ahead. As solar wind speed generally orders solar wind structures, it is expected to provide a general improvement in space-weather forecasting capability.

Over the seven-year period of this study, we see that the data assimilation scheme (BRaVDA) is able to improve forecasts of the solar-wind speed when compared to the prior ensemble (with no DA), reducing RMSE by an average of 36 km/s at all spacecraft considered. Additionally, the uncertainties present in the forecast estimates are greatly reduced by assimilating data. The posterior forecast RMSE standard errors decreased by an average of 3.1 km/s compared to the prior forecast RMSE standard deviation. This means that we can have increased confidence in the forecasts produced when using data assimilation. During solar maximum, this improvement in forecast RMSE is reduced to about 14 km/s, but during solar minimum, this is increased to about 48 km/s. This shows that the application of data assimilation, even with a relatively simple solar-wind model, can lead to large improvements in forecasting accuracy. The lack of significant difference in forecast accuracy between solar minimum and solar maximum is likely due to competing effects taking precedence. At solar minimum, the highly structured nature of the solar wind with latitude produces forecast errors for both the corotation and data assimilation approaches. Additionally, the solar-wind speed is generally higher and more variable at solar minimum than solar maximum, which also inflates forecast RMSE. Conversely, at solar maximum, forecast errors are produced by the increased frequency of transient solar-wind structures, which none of the studied forecast approaches incorporate. Also, the STEREO spacecraft are further from Earth during the solar maximum period studied, which increases forecast error.

The data assimilation forecasts at Earth are of similar accuracy and uncertainty to the STEREO-B corotation forecast accuracy and uncertainty. This is somewhat to be expected as the data assimilation is also based upon the STEREO-B observation and is expected to recreate it at the satellite location. But the DA approach is advantageous for two reasons. First, the data assimilation scheme provides far more information than the STEREO-B corotation forecast. Whilst corotation gives us a point forecast at Earth (or can be corotated to any other point at the same latitude and heliocentric distance), the DA scheme maps observational data throughout the model domain. This means that the DA scheme is able to give us improved estimates of the solar wind throughout the inner heliosphere. The advantage of this is that this will give a more accurate representation of the solar wind between the Sun and Earth, which will provide a more suitable medium for CMEs to be propagated through (e.g., see Barnard et al., 2020; Owens, Lang, Barnard, Riley, et al., 2020). This will be elaborated upon more in a later study.

Second, the data assimilation also provides a more consistent and robust forecast for the solar wind compared to the STEREO-B corotation forecast. DA is able to reduce the representivity errors that are caused by the latitudinal offset between STEREO-B and ACE/Earth. The STEREO-B corotation forecast is sensitive to the latitudinal offset between STEREO-B and Earth, with poorer forecasts occurring when STEREO-B is at greater latitudinal separation. This is restricted to solar minimum, when there are greatly differing solar-wind regimes at different latitudes. By relying on more than one observation source and accounting for potential errors in the observations, data assimilation is able to compensate for these errors and remove the correlation between forecast error and latitude offset. The DA scheme will also reduce the effect of transient structures (such as coronal mass ejections) seen only at STEREO-B, thus reducing false positive forecasts of high-speed streams. More study is required to determine how to balance the representivity errors contributed from latitudinal offset and the "age" of observations for the purposes of forecasting at Earth. Both of these effects are expected to have competing solar-cycle variations.

HUXt is intended to act as a surrogate model for more complex MHD models. This study shows that BRaVDA could be used to define an optimum set of boundary conditions using solar wind observations which can then be used to drive more complex models, such as ENLIL and HelioMAS.

Data Availability Statement

HUXt can be downloaded in the Python programming language from https://doi.org/10.5281/zenodo.4889327. BRaVDA can be downloaded in the Python programming language from https://doi.org/10.5281/zenodo.4849682.

Acknowledgments
We are grateful to the Space Physics Data Facility and National Space Science Data Center for OMNI. STEREO-A, STEREO-B and ACE data were downloaded from the OMNIWeb portal at https://omniweb.gsfc.nasa.gov/. MAS model output is available from the Predictive Science Inc. website: (http://www.predsci.com/mhdweb/home.php). Work was part-funded by Science and Technology Facilities Council (STFC) grant number ST/R000921/1, and Natural Environment Research Council (NERC) grant number NE/P016928/1. Matthew Lang was funded by NERC grant number NE/S010033/1. Jake Witherington was funded by a RAS undergraduate bursary. Harriet Turner is funded through SCENARIO grant number NE/S007261/1.

References

Amezcua, J., Goodliff, M., & Leeuwen, Van, P. J. (2017). A weak-constraint 4DEnsembleVar. Part I: Formulation and simple model experiments. *Tellus A: Dynamic Meteorology and Oceanography*, *69*(1), 1271. https://doi.org/10.1080/16000870.2016.1271564

Arge, C. N., Henney, C. J., Koller, J., Compeau, C. R., Young, S., MacKenzie, D., et al. (2010). Air force data assimilative photospheric flux transport (ADAPT) model. In: *AIP Conference Proceedings* (Vol. 1216, pp. 343–346). American Institute of Physics. https://doi.org/10.1063/1.3395870

Arge, C. N., Odstrcil, D., Pizzo, V. J., & Mayer, L. R. (2003). Improved method for specifying solar wind speed near the Sun. In: *AIP Conference Proceedings* (Vol. 679, pp. 190–193). American Institute of Physics Inc. https://doi.org/10.1063/1.1618574

Bannister, R. (2007). *Elementary 4D-var (DARC technical report)*.

Barnard, L., Owens, M. J., Scott, C. J., & Koning, C. A. (2020). Ensemble CME modeling constrained by heliospheric imager observations. *AGU Advances*, *1*(3). https://doi.org/10.1029/2020av000214

Bayes, T., & Price, R. (1763). An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFRS. *Philosophical Transactions*, *53*, 370–418. https://doi.org/10.1098/rstl.1763.0053

Bocquet, M., & Sakov, P. (2014). An iterative ensemble Kalman smoother. *Quarterly Journal of the Royal Meteorological Society*, *140*(682), 1521–1535. https://doi.org/10.1002/qj.2236

Breen, A. R., Fallows, R. A., Bisi, M. M., Thomasson, P., Jordan, C. A., Wannberg, G., & Jones, R. A. (2006). Extremely long baseline interplanetary scintillation measurements of solar wind velocity. *Journal of Geophysical Research*, *111*(A8), 104. https://doi.org/10.1029/2005ja011485

Browne, P. A., & Wilson, S. (2015). A simple method for integrating a complex model into an ensemble data assimilation system using MPI. *Environmental Modelling & Software*, *68*, 122–128. https://doi.org/10.1016/j.envsoft.2015.02.003

Bust, G. S., & Mitchell, C. N. (2008). History, current state, and future directions of ionospheric imaging. *Reviews of Geophysics*, *46*(1), RG1003. https://doi.org/10.1029/2006RG000212

Durazo, J. A., Kostelich, E. J., & Mahalov, A. (2017). Local ensemble transform Kalman filter for ionospheric data assimilation: Observation influence analysis during a geomagnetic storm event. *Journal of Geophysical Research: Space Physics*, *122*(9), 9652–9669. https://doi.org/10.1002/2017ja024274

Errico, R. M. (1997). What is an adjoint model? *Bulletin of the American Meteorological Society*, *78*(11), 2577–2591. https://doi.org/10.1175/1520-0477(1997)078<2577:wiaam>2.0.co;2

Evensen, G., Dee, D. P., & Schröter, J. (1998). Parameter estimation in dynamical models. In: Ocean modeling and parameterization (pp. 373–398). Springer. https://doi.org/10.1007/978-94-011-5096-5_16

Eyles, C. J., Harrison, R. A., Davis, C. J., Waltham, N. R., Shaughnessy, B. M., Mapson-Menard, H. C. A., et al. (2009). The heliospheric imagers onboard the STEREO mission. *Solar Physics*, *254*(2), 387–445. https://doi.org/10.1007/s11207-008-9299-0

Goodliff, M., Amezcua, J., & Leeuwen, Van, P. J. (2017). A weak-constraint 4DEnsembleVar. Part II: Experiments with larger models. *Tellus A: Dynamic Meteorology and Oceanography*, *69*(1), 1271. https://doi.org/10.1080/16000870.2016.1271565

Hapgood, M. A. (2011). Towards a scientific understanding of the risk from extreme space weather. *Advances in Space Research*, *47*(12), 2059–2072. https://doi.org/10.1016/j.asr.2010.02.007

Howes, K. E., Fowler, A. M., & Lawless, A. S. (2017). Accounting for model error in strong-constraint 4D-Var data assimilation. *Quarterly Journal of the Royal Meteorological Society*, *143*(704), 1227–1240. https://doi.org/10.1002/qj.2996

Kaiser, M. L., Kucera, T. A., Davila, J. M., Cyr, O. C. S., Guhathakurta, M., & Christian, E. (2008). The STEREO mission: An introduction. In: The STEREO mission (pp. 5–16). Springer. https://doi.org/10.1007/978-0-387-09649-0_2

Kalnay, E. (2003). *Atmospheric modeling, data assimilation and predictability*. Cambridge university press.

Kohutova, P., Bocquet, F.-X., Henley, E. M., & Owens, M. J. (2016). Improving solar wind persistence forecasts: Removing transient space weather events, and using observations away from the Sun-Earth line. *Space Weather*, *14*(10), 802–818. https://doi.org/10.1002/2016SW001447

Lang, M., Browne, P., Leeuwen, Van, P. J., & Owens, M. (2017). Data assimilation in the solar wind: Challenges and first results. *Space Weather*, *15*(11), 1490–1510. https://doi.org/10.1002/2017SW001681

Lang, M., Leeuwen, van, P. J., & Browne, P. A. (2016). A systematic method of parameterisation estimation using data assimilation. *Tellus*, *68*(1). https://doi.org/10.3402/tellusa.v68.29012

Lang, M., & Owens, M. J. (2019). A variational approach to data assimilation in the solar wind. *Space Weather*, *17*(1), 59–83. https://doi.org/10.1029/2018SW001857

Linker, J. A., Mikić, Z., Biesecker, D. A., Forsyth, R. J., Gibson, S. E., Lazarus, A. J., et al. (1999). Magnetohydrodynamic modeling of the solar corona during whole Sun month. *Journal of Geophysical Research*, *104*(A5), 9809–9830. https://doi.org/10.1029/1998JA900159

Mackay, D., & Yeates, A. (2012). The Sun's global photospheric and coronal magnetic fields: Observations and models. *Living Reviews in Solar Physics*, *9*(1), 6. https://doi.org/10.12942/lrsp-2012-6

McGregor, S. L., Hughes, W. J., Arge, C. N., Owens, M. J., & Odstrcil, D. (2011). The distribution of solar wind speeds during solar minimum: Calibration for numerical solar wind modeling constraints on the source of the slow solar wind. *Journal of Geophysical Research*, *116*(A3). https://doi.org/10.1029/2010JA015881

Merkin, V. G., Lyon, J. G., Lario, D., Arge, C. N., & Henney, C. J. (2016). Time-dependent magnetohydrodynamic simulations of the inner heliosphere. *Journal of Geophysical Research: Space Physics*, *121*(4), 2866–2890. https://doi.org/10.1002/2015JA022200

Odstrcil, D. (2003). Modeling 3-D solar wind structure. *Advances in Space Research*, *32*(4), 497–506. https://doi.org/10.1016/s0273-1177(03)00332-6

Owens, M., Lang, M., Barnard, L., Riley, P., Ben-Nun, M., Scott, C. J., et al. (2020a). A Computationally efficient, time-dependent model of the solar wind for use as a surrogate to three-dimensional numerical magnetohydrodynamic simulations. *Solar Physics*, *295*(3), 43. https://doi.org/10.1007/s11207-020-01605-310.1007/s11207-020-01605-3

Owens, M. J. (2018). Time-window approaches to space-weather forecast metrics: A solar wind case study. *Space Weather*, *16*(11), 1847–1861. https://doi.org/10.1029/2018SW002059

Owens, M. J. (2020). Solar-wind structure. In: Oxford research encyclopedia of physics. Oxford University Press. https://doi.org/10.1093/acrefore/9780190871994.013.19

Owens, M. J., & Forsyth, R. J. (2013). The heliospheric magnetic field. *Living Reviews in Solar Physics*, *10*(5), 5. https://doi.org/10.12942/lrsp-2013-5

Owens, M. J., Lang, M., Riley, P., Lockwood, M., & Lawless, A. S. (2020b). Quantifying the latitudinal representivity of in situ solar wind observations. *Journal of Space Weather and Space Climate*, *10*, 8. https://doi.org/10.1051/swsc/2020009

Owens, M. J., & Riley, P. (2017). Probabilistic solar wind forecasting using large ensembles of ear-Sun conditions with a simple one-dimensional "upwind" scheme. *Space Weather*, *15*(11), 1461–1474. https://doi.org/10.1002/2017SW001679

Poedts, S., & Pomoell, J. (2017). EUHFORIA: A solar wind and CME evolution model. In: *19th EGU General Assembly, EGU2017, Proceedings from the Conference Held 23-28 April* (Vol. 19, p. 7396). Vienna: Elsevier Procedia.

Riley, P., Linker, J. A., & Arge, C. N. (2015). On the role played by magnetic expansion factor in the prediction of solar wind speed. *Space Weather*, *13*(3), 154–169. https://doi.org/10.1002/2014SW001144

Riley, P., Linker, J. A., Lionello, R., & Mikic, Z. (2012). Corotating interaction regions during the recent solar minimum: The power and limitations of global MHD modeling. *Journal of Atmospheric and Solar-Terrestrial Physics*, *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. *Solar Physics*, *270*(2), 575–592. https://doi.org/10.1007/s11207-011-9766-x

Stone, E. C., Frandsen, A. M., Mewaldt, R. A., Christian, E. R., Margolies, D., Ormes, J. F., & Snow, F. (1998). The advanced composition explorer. In: *The Advanced Composition Explorer Mission* (Vol. 86, pp. 1–22). Springer. https://doi.org/10.1023/A:1005082526237

Thomas, S. R., Fazakerley, A., Wicks, R. T., & Green, L. (2018). Evaluating the skill of forecasts of the near-Earth solar wind using a space weather monitor at L5. *Space Weather*, *16*(7), 814–828. https://doi.org/10.1029/2018SW001821