Solar wind data assimilation in an operational context: Use of near-real-time data and the forecast value of an L5 monitor.

Harriet Turner[1], Mathew J Owens[1], Matthew Simon Lang[1], Andrew W Smith[2], Pete Riley[3], Mike S Marsh[4], and Siegfried Gonzi[5]

[1]University of Reading
[2]Northumbria University
[3]Predictive Science Inc.
[4]Met Office
[5]UK Met Office

April 20, 2023

Solar wind data assimilation in an operational context: Use of near-real-time data and the forecast value of an L5 monitor.

Harriet Turner [1], Matthew Lang[1], Mathew Owens[1], Andy Smith[2], Pete Riley[3], Mike Marsh[4], Siegfried Gonzi[4]

[1]Department of Meteorology, University of Reading, Reading, UK
[2]Northumbria University, Sutherland Building, Northumberland Road, Newcastle-upon-Tyne, NE1 8ST
[3]Predictive Science Inc., San Diego, CA, USA
[4]Met Office, Exeter, UK

Key Points:

- Solar wind data assimilation needs to perform well with near-real-time data for it to be used operationally for space weather forecasting.
- Despite lower data quality, solar wind speed forecasts based on near-real-time data are comparable to those based on science-level data.
- Assimilation of L1 and L5 data gives forecast error improvement of 15% for lead times up to 5 days over assimilation of only L1 data.

Corresponding author: Harriet Turner, h.turner3@pgr.reading.ac.uk

Abstract

For accurate and timely space weather forecasting, advanced knowledge of the ambient solar wind is required, both for its direct impact on the magnetosphere and for accurately forecasting the propagation of coronal mass ejections to Earth. Data assimilation (DA) combines model output and observations to form an optimum estimation of reality. Initial experiments with assimilation of in situ solar wind speed observations suggest the potential for significant improvement in the forecast skill of near-Earth solar wind conditions. However, these experiments have assimilated science-quality observations, rather than near-real-time (NRT) data that would be available to an operational forecast scheme. Here, we assimilate both NRT and science observations from the Solar Terrestrial Relations Observatory (STEREO) and near-Earth observations from the Advanced Composition Explorer (ACE) and Deep Space Climate Observatory (DSCOVR) spacecraft. We show that solar wind speed forecasts using NRT data are comparable to those based on science-level data. This suggests that an operational solar wind DA scheme would provide significant forecast improvement, with reduction in the mean absolute error (MAE) of solar wind speed around 46% over forecasts without DA. With a proposed space weather monitor planned for the L5 Lagrange point, we also quantify the solar wind forecast gain expected from L5 observations alongside existing observations from L1. This is achieved using configurations of the STEREO and L1 spacecraft. There is a 15% improvement for forecast lead times of less than 5 days when observations from L5 are assimilated alongside those from L1, compared to assimilation of L1 observations alone.

Plain Language Summary

Space weather is the conditions of space in the near-Earth environment, and it has the potential to cause a significant impact on modern day life. For accurate space weather forecasting, knowledge of the background solar wind (a continual stream of charged particles flowing from the Sun) conditions is needed. This can be achieved using data assimilation (DA), which combines existing knowledge of the system with observations to form an optimum estimation of reality. Previous solar wind DA experiments have used cleaned-up ‘science-level’ data, which only become available many days after the observations are made. But for forecasting, where a rapid response is important, DA needs to work with near-real-time (NRT) data. NRT data often contains data gaps, biases and noise when compared to the science-level data. Here, we find that using NRT data does not significantly worsen the forecasts, which is promising for DA forecasting. A future space weather monitoring mission to the L5 Lagrange point (60 degrees behind Earth in longitude) also offers an opportunity for solar wind DA. This is tested using combinations of existing spacecraft observations. Including L5 data alongside observations for Earth improves solar wind forecasting capability for forecasts up to 5 days in the future.

1 Introduction

Space weather poses a threat to modern technologies and human health. It can damage satellites, cause communication failures and destroy electricity transformers causing blackouts. It also puts the health of astronauts in space and passengers on high-altitude flights at risk (Cannon, 2013). Accurate space weather forecasting requires knowledge of the background solar wind, a continual stream of charged particles and magnetic field that fills the heliosphere (Parker, 1958). Stream interaction regions (SIRs) form where fast streams of solar wind catch up with and compress slower streams ahead, leading to regions of higher density and stronger magnetic field (Gosling & Pizzo, 1999; Richardson & Cane, 2012). These can persist for several solar rotations as corotating interaction regions (CIRs) and can be a source of recurrent space weather. The most severe space weather, namely geomagnetic storms, is driven by coronal mass ejections (CMEs), which are huge eruptions of coronal material and magnetic field from the Sun (Webb & Howard, 2012). These propagate through the background solar wind, meaning ambient conditions can impact the CME speed and arrival time at Earth (Cargill, 2004; Case et al., 2008; Riley & Ben-Nun, 2021). Although severe space weather causes the largest impacts, the effect of mild and moderate space weather also causes a considerable economic impact, with estimates of effects on the power grid over the EU and US costing USD1.3 - 2.1 trillion over a century (Schrijver, 2015). With extreme space weather relying less on the background solar wind conditions, the largest improvements in forecasting is expected for mild to moderate space weather events.

Forecasting near-Earth solar wind conditions can be achieved using simple in situ observation-based methods, such as corotation (e.g. M. J. Owens et al., 2013; Thomas et al., 2018; Turner et al., 2022), or data driven methods (Riley et al., 2017). These approaches generally do not capture transient solar wind structures, such as CMEs, and only estimate the solar wind at a single point in space. Global solar wind conditions can be forecast on the basis of remote solar observations. Photospheric magnetic field observations are used to constrain semi-empirical (e.g. WSA, Arge et al., 2003) and more physics-based (e.g. MAS, Linker et al., 1999) models of the corona. The solar wind conditions at the top of the corona can then be propagated to Earth (and beyond) using solar wind models. This is typically achieved with numerical magnetohydrodynamic (MHD) models (e.g. Merkin et al., 2016; Odstrcil, 2003; Riley et al., 2001; Tóth et al., 2005), though reduced-physics approximations can provide a complementary, computationally efficient, approach (HUX, Riley and Lionello (2011); M. J. Owens and Riley (2017) and HUXt, M. Owens (2020)). CME-like disturbances can be introduced at the lower boundary of the solar wind model based on the CME characteristics observed in coronagraph observations (Zhao et al., 2002; Odstrcil et al., 2004). Once ambient and CME inner boundary conditions are supplied to the solar wind models, there are no further observational constraints on the model evolution.

Data assimilation (DA) combines model output and observations to form an optimum estimation of reality. It has led to huge improvements in terrestrial weather forecasting (Migliorini & Candy, 2019), however has not been fully utilised for solar wind forecasting. The Burger Radius Variational Data Assimilation (BRaVDA) scheme (Lang & Owens, 2019) makes use of in situ observations from spacecraft in both near-Earth space and from other locations within the heliosphere. It has been shown to significantly improve the model representation of the ambient solar wind, which is expected to translate to similar forecast gains (Lang et al., 2021). However, all experiments using BRaVDA so far have been carried out using 'science-level' data which has been processed on the ground and is often not made available for weeks or months after the observation date. For solar wind DA to be used operationally to produce timely space weather forecasts, it must be able to perform well with near-real-time (NRT) data. NRT data often includes erroneous results, data gaps, and sometimes systematic biases; a lot of which gets corrected in the subsequent data processing stage. Figure 1 shows one month of NRT and science-level solar wind speed data from 2012/04/01 to 2012/05/01 for Advanced Composition Explorer (ACE, Stone et al., 1998), Solar Terrestrial Relations Observatory (Kaiser et al., 2008) Ahead (STEREO-A) and Behind (STEREO-B) spacecraft. Similarly, Figure 2 shows one month of data from the Deep Space Climate Observatory (DSCOVR, Burt & Smith, 2012) spacecraft from 2017/07/01 to 2017/08/01. There are numerous features that show the differing quality between the NRT and science level data; for example, the step changes in the ACE NRT data, increased noise in the STEREO-B NRT data and large spikes and data gaps in the DSCOVR NRT data (Smith et al., 2022). In this study, we assess the performance of the BRaVDA scheme using archived NRT data for three time periods; 2009/08/01 to 2011/02/01, 2012/04/01 to 2013/10/01 and 2017/07/01 to 2019/01/01. The first interval covers the 18 months up to the effective boundary between solar minimum and solar maximum, whereas the second interval is during solar maximum. These were selected for their solar cycle location, whereas the final interval was an arbitrary 18-month period once the DSCOVR spacecraft was operational.

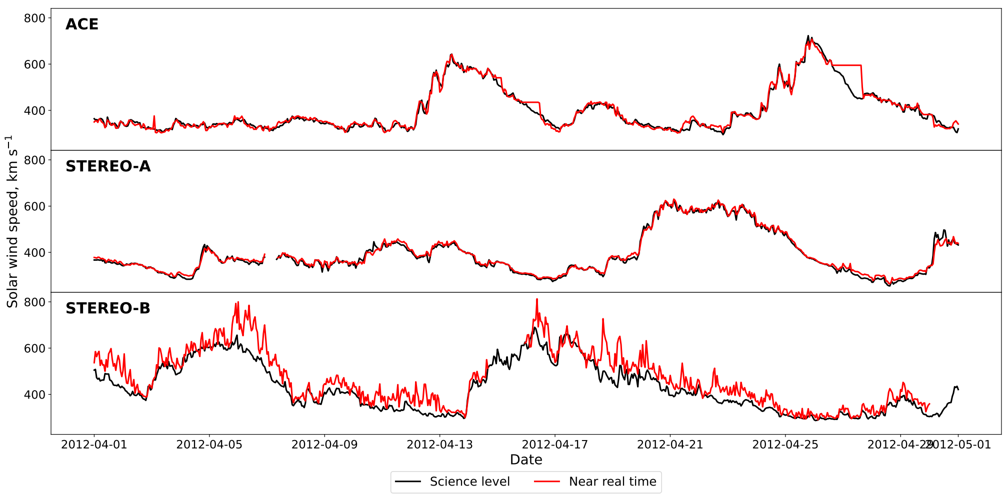

Figure 1. Time series of both science-level (black line) and near-real-time (red line) observations from the ACE, STEREO-A and STEREO-B spacecraft; top, middle and bottom respectively. This covers the interval from 2012/04/01 to 2012/05/01. Data are shown at an hourly resolution.

Future deployment of an operational DA scheme would aim to exploit observations from *Vigil* (Luntama et al., 2020), a planned space weather monitoring mission at the L5 Lagrange point, approximately 60 degrees behind Earth in heliospheric longitude. Alongside data from a monitor at L1, e.g. DSCOVR, this could form a framework for solar wind speed forecasting using data assimilation. Using configurations of observations from STEREO and from near-Earth, we can approximate the future pairing of L5 and L1 monitors. Here, we test the performance of BRaVDA using NRT and science-level observations from spacecraft that are separated by approximately 60 degrees in longitude to simulate an operational L5 solar wind monitor. We can then assess what forecast advantage we can expect from a future mission pairing.

The data used in this work are described in Section 2 and the methods in Section 3. The results and discussion are in Section 4 and the conclusions in Section 5.

2 Data

All data (NRT and science-level) are averaged to an hourly resolution using a boxcar technique with no minimum requirement for the number of data points. This is a good approximation for solar wind speed due to its high autocorrelation (Lockwood et al., 2019).

2.1 STEREO data

The STEREO mission was designed to provide a unique viewpoint of ejecta from the Sun and is comprised of two spacecraft; STEREO ahead (STEREO-A) and STEREO behind (STEREO-B) (Kaiser et al., 2008). These were launched into Earth-like orbits in October 2006, where STEREO-A is ahead in Earth's orbit and STEREO-B behind. The spacecraft separate at approximately 22.5° per year and reached opposition to Earth in 2014. During a planned reset of the spacecraft in October 2014 in preparation for opposition, communication with STEREO-B was lost and has not been regained. The STEREO

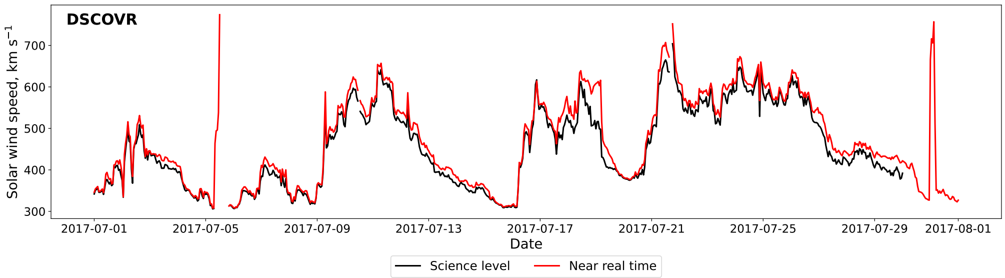

Figure 2. Time series of both science-level (black line) and near-real-time (red line) observations from the DSCOVR spacecraft. This covers the interval from 2017/07/01 to 2017/08/01. Data are shown at an hourly resolution.

near-real-time (beacon) data is available from `https://stereo-ssc.nascom.nasa.gov/data/beacon/` and science-level data from `https://cdaweb.gsfc.nasa.gov/`. Solar wind speed is measured using the Plasma and Suprathermal Ion Composition (PLASTIC) instrument, which provides in situ solar wind and ion observations (Galvin et al., 2008). The science data is level 2 processed data. The beacon data is provided in a continuous broadcast mode, at 1-minute resolution. For use in BRaVDA, this must be lightly processed so that any unphysical values are removed and the data is on the correct time step. As the input data used in BRaVDA is at an hourly cadence, the NRT data is averaged accordingly. This essentially interpolates over any data gaps that are less than an hour long; if there is a single 1-minute value in an hour interval then this will be taken as representative for that hour. Although this technique would not be suitable for other parameters, such as magnetic field direction, it is expected to be an adequate solution for solar wind speed, which has a long auto-correlation time (Lockwood et al., 2019). The NRT data has a typical latency of less than 10-minutes (Biesecker et al., 2008), which would not cause issues for use operationally, as the DA makes use of hourly averages.

The bottom two panels of Figure 1 show an example of one month of data from STEREO-A and STEREO-B. The middle panel shows the STEREO-A data, with NRT in red and science data in black, and in general there is a very good agreement between the two time series. However, the STEREO-B NRT data in the bottom panel shows much greater variability in time compared to the science data. The data plotted is at an averaged hourly resolution, meaning that a large amount of noise must have already been filtered out through this averaging. The greater variability is also demonstrated in Figure 3, with the STEREO data in the bottom two rows. Here we have 2D histograms of NRT against science observations, with the colour representing the density of observations on a log scale. The three time intervals used in this study are shown; 2009/08/01 to 2011/02/01, 2012/04/01 to 2013/10/01 and 2017/07/01 to 2019/01/01, the choice of which is described in Section 4. The STEREO-A NRT data showed periods of low solar wind speed, as shown in the left hand panel of the middle row in Figure 3. This is data from the period of time from October 2009 to January 2010, as shown in more detail in Figure 4. There is a gradual worsening of the relationship between the NRT and science-level observations, before this is resolved and the relationship returns to lie ap-

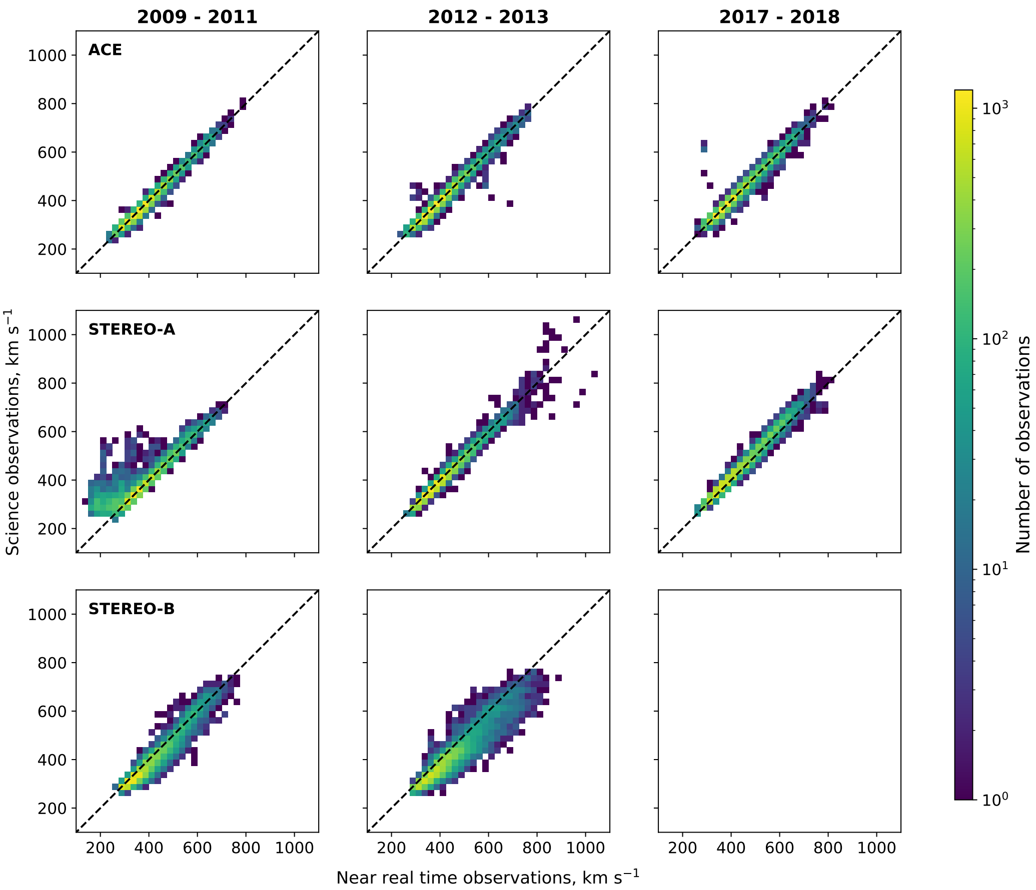

Figure 3. Two-dimensional histograms of science-level observations and near-real-time observations for ACE, STEREO-A and STEREO-B (rows) for the periods of time; 2009/08/01 to 2011/02/01 (left-hand column), 2012/04/01 to 2013/10/01 (middle column) and 2017/07/01 to 2019/01/01 (right-hand column). The black dashed line represents $x = y$. The number of observations are shown as a log scale.

proximately along $y = x$. Although the cause of this is unknown, it provides a useful test for the DA to see how data quality affects the resulting forecasts. The later two time periods show a good relationship between NRT and science data.

The greater variability in the STEREO-B NRT data shown in Figure 1 can also be seen in the greater spread about the $y = x$ line in the bottom row of Figure 3. For the intervals shown, the average standard deviation of the difference between the science and NRT observations is 29.1 kms^{-1}, compared to 13.0 and 23.3 kms^{-1} for ACE and STEREO-A respectively. This is due to a known issue with the detector and is present for the whole operational lifetime of STEREO-B. This issue is resolved in the processing of the data on the ground that produces the science-level data.

2.2 ACE data

The Advanced Composition Explorer (ACE) was launched in August 1997, with the mission aiming to investigate the composition of solar wind plasma at the L1 Lagrange point. The spacecraft carries a suite of instruments, including the Solar Wind Electron, Proton and Alpha Monitor (SWEPAM) and the Real Time Solar Wind monitoring sys-

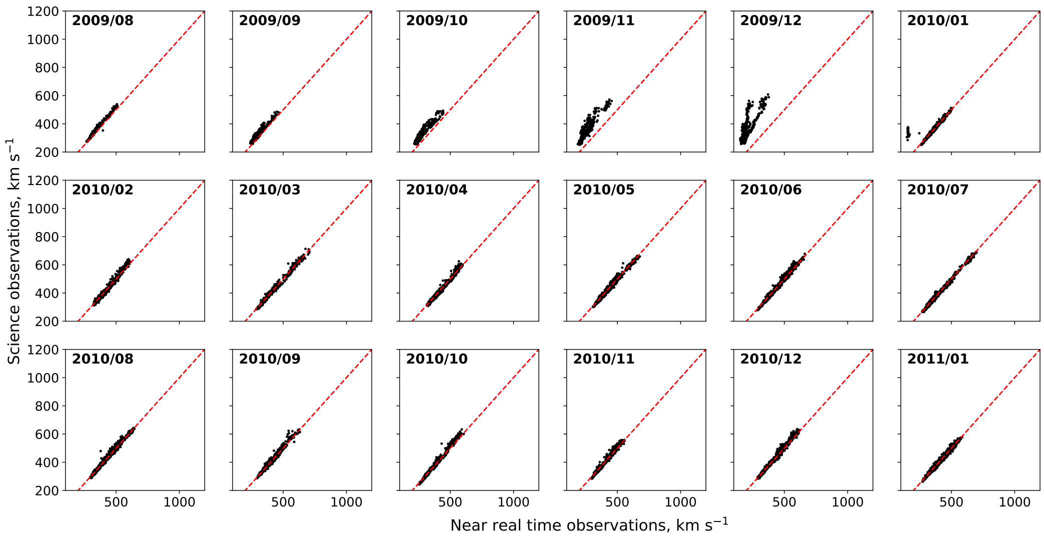

Figure 4. Near-real-time solar wind speeds against science-level solar wind speeds for STEREO-A, during the solar minimum interval (2009-2011) further subdivided into monthly intervals.

tem (RTSW) (Stone et al., 1998). SWEPAM characterises the bulk flow of the solar wind through measurement of electron and ion distribution functions in 3 dimensions (McComas et al., 1998). This is then available as 1-hour science level 2 data through CDAWeb at `https://cdaweb.gsfc.nasa.gov/`. The RTSW experiment also continually transmits a feed of near-real-time data that can provide a warning of solar wind conditions to arrive at Earth up to 1 hour later (Stone et al., 1998). This data is available from NASA's Community Coordinated Modelling Centre at `https://ccmc.gsfc.nasa.gov/requests/GetInput/get_ace_K.php`.

The NRT and science-level data from ACE agree very well. As the top panel in Figure 1 shows, there are some features where the NRT data is constant and then steps back down to the science data. The cause of this is unknown, however, as Figure 3 shows, the observations mostly lie close to the $y = x$ line and so overall there is good agreement.

The NRT data has a typical latency of less than 5 minutes, which is not expected to cause any problems for an operational DA scheme.

2.3 DSCOVR data

The Deep Space Climate Observatory (DSCOVR) was launched in February 2015 to the L1 Lagrange point. The mission was launched to succeed ACE and to aid the National Oceans and Atmosphere Administration (NOAA) in real-time monitoring of space weather. For this study, data from the *PlasMag* instrument was used, which is comprised of a magnetometer, Faraday cup and a top-hat electron electrostatic analyser. Here, we make use of the observations from the Faraday cup, which measures the solar wind velocity, density and temperature. Both the NRT and science-level (level 2) data is available through the DSCOVR Space Weather Data Portal at `https://www.ngdc.noaa.gov/dscovr/portal/index.html#/`. As Figure 2 shows, the NRT data shows erroneous spikes in solar wind speed. This is due to periods of very low solar wind density, meaning that the Faraday cup cannot accurately measure the solar wind speed (Loto'aniu et al., 2022).

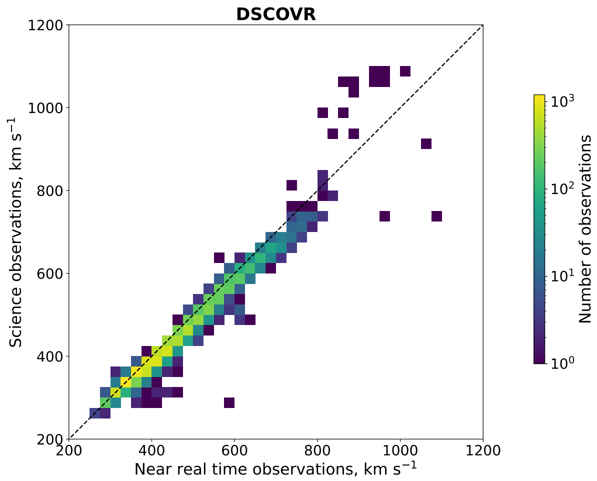

Figure 5. Two-dimensional histogram of science-level observations and near-real-time observations for Earth, using data from the DSCOVR spacecraft. This covers the period of time from 2017/07/01 to 2019/01/01. The black dashed line represents $y = x$. The number of observations is shown on a log scale.

Similarly to ACE, the NRT data latency for DSCOVR is not expected to cause any problems for an operational DA scheme.

3 Methods

3.1 BRaVDA and forecast generation

A complete description of the BRaVDA methodology can be found in Lang and Owens (2019) and the code is available at `https://github.com/University-of-Reading-Space-Science/BRaVDA`. Here, we provide a brief overview of the scheme. BRaVDA combines in situ solar wind speed observations with the steady-state "HUX" model, based on Riley and Lionello (2011). BRaVDA maps information contained within in situ observations, typically at 1 AU, back to the model's inner boundary at 30 solar radii (R_S), where it is combined with the prior inner boundary condition. This prior is defined using output from the HelioMAS model Riley et al. (2001) at 30 R_S. These model data are available at `https://www.predsci.com/portal/home.php`. The information is merged through the minimisation of a cost function, which aims to find the optimum compromise between the prior information and the observations, accounting for the uncertainties in both. Once the inner boundary at 30 R_S is updated, this can then be propagated back out to 1 AU (and beyond) through the use of any solar wind model. For efficiency, HUX is used again for this stage. This produces an estimate of the solar wind over the 2 dimensional domain from $30R_S$ to the outer boundary, which here is set to $245R_S$, to fully include the orbital radii of all spacecraft considered. The 2D plane considered here is the radius/ longitude plane, located at the solar equator.

Note that previous work using BRaVDA (e.g. Lang et al., 2021; Turner et al., 2022) has made the implicit assumption that the observations made from the STEREO spacecraft were taken from 215 R_S (1 AU) and the L1 observations are at 213 R_S. In reality, this is not the case. As shown in Figure 6, Earth varies from 210 to 219 R_S over the year, STEREO-A varies from 206 to 208 R_S and STEREO-B varies from 215 to 234 R_S. These variations are now included into BRaVDA, ensuring that the observations were taken from the correct orbital radius. Due to the highly correlated nature of the solar

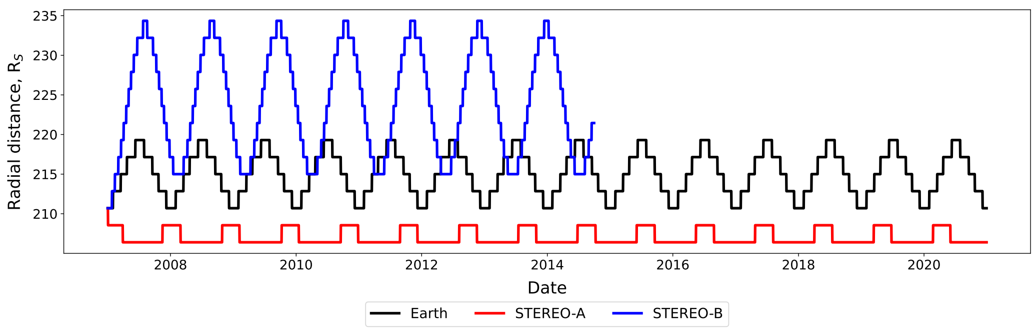

Figure 6. Variation of Earth (black), STEREO-A (red) and STEREO-B (blue) in radial distance from the Sun. The y-axis is shown in solar radii (R_S) and covers the time period from 2007 to 2021. Note that contact with STEREO-B was lost in 2014.

wind, this radial variation did not have a significant impact on the accuracy of the forecasts, however it is important to be as representative of the system as possible.

Forecasts are generated using the output from BRaVDA in the same way as Turner et al. (2022). (As archived data are used for this work, what we state here are forecasts are actually hindcasts. However, as these hindcasts are used to inform the performance we would expect from forecasts, we retain the use of the word 'forecast' for simplicity.) In summary, BRaVDA is run on a daily cadence, which assimilates observations from the previous 27 days to produce a DA solution. Assuming steady state conditions, this can be corotated to produce a forecast for the subsequent 27 days. Here, forecasts are produced from assimilation of NRT and science-level observations, and both are verified against the science-level observations to assess their accuracy.

3.2 L5 experiments

Future deployment of an operational solar wind DA scheme could make use of both observations from near-Earth space (for example, from DSCOVR) and from the planned *Vigil* mission to L5. To test the performance of such a combination, we can use observations from pairs of spacecraft (STEREO-A, STEREO-B and ACE) that are approximately 60 degrees apart in longitude. By using intervals of time where the spacecraft separation is between 50 and 70 degrees, we produce four 'L1-L5' analysis periods. These periods are shown in Table 1 and schematically in Figure 7. The spacecraft lagging with respect to solar rotation acts as the effective L5 monitor and the spacecraft leading with respect to solar rotation is the effective near-Earth, or L1, monitor. We can then assess the forecast performance at the leading spacecraft, as this would represent a forecast at Earth.

4 Results and discussion

Here we conduct a number of experiments to investigate the impact of using near-real-time data on forecasts produced using DA. Here, the science-level observations act as a verification time series for the forecasts to be compared against. The science-level data is also used to produce corotation forecasts, whereby observations are lagged depending on their longitudinal separation from the forecast location. Throughout, we assess the performance of forecasts produced using mean absolute error (MAE) as a function of forecast lead time. As a standard metric, MAE allows for easy comparison of the performance of different forecasts. However, caution must be taken with such "point-

Effective L5	Effective L1	Start	End
STEREO-B	STEREO-A	02/05/2008	30/08/2008
STEREO-B	Earth	30/07/2009	22/01/2010
Earth	STEREO-A	27/05/2009	06/05/2010
STEREO-A	STEREO-B	25/10/2013	09/02/2014

Table 1. Intervals where spacecraft are separated by between 50 and 70 degrees in longitude. These intervals simulate spacecraft at L5 and at L1. In the left and middle panels, the spacecraft are moving away from each other and so the start date indicates the time where they are separated by 50 degrees and the end dates when separated by 70 degrees. In the right panel, the spacecraft are moving towards each other and so the start date is when they are separated by 70 degrees and the end date 50 degrees.

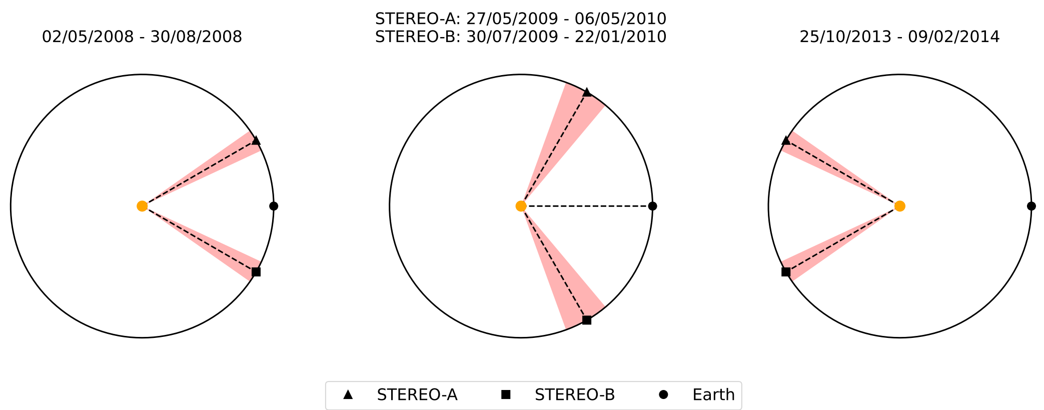

Figure 7. Configurations of the spacecraft used in the experiments assessing the possible combination of an L5 and L1 monitor. The red shaded regions show the time where the spacecraft are separated between 50 and 70 degrees. Earth is indicated by the black circle, STEREO-A by the black triangle and STEREO-B by the black square.

by-point" metrics, as they can be misleading with forecasts of markedly different quality, typically over-penalising forecasts with small timing errors and under-penalising forecasts with very low variance (M. J. Owens et al., 2005). In this study, the difference between near-real-time and science forecasts is generally expected to be a small quantitative change, rather than leading to a qualitatively different time series. For this reason, MAE is found to generally agree with the assessment gained by visual inspection. However, Section 4.1 highlights a case where MAE is inadequate to characterise the forecast performance in isolation.

4.1 Assimilation of single and multiple spacecraft observations

We first assimilate observations from a single spacecraft. We have observations from four sources; ACE, STEREO-A, STEREO-B and DSCOVR. Three time intervals are used for analysis; 2009/08/01 to 2011/02/01, 2012/04/01 to 2013/10/01, and 2017/07/01 to 2019/01/01. The first interval covers the 18 months effectively covers solar minimum, the second interval occurs during solar maximum and the final interval is an arbitrary 18-month period once the DSCOVR spacecraft was operational. Data from all spacecraft are not available for all time periods, as DSCOVR was only launched in 2015 and communication with STEREO-B was lost in 2014.

Figure 8 shows MAE as a function of forecast lead time for experiments assimilating observations from a single spacecraft; ACE observations are assimilated in the top panel, DSCOVR is assimilated in the second panel, STEREO-A in the third panel and STEREO-B in the bottom panel. These are shown for the time intervals where data is available; 2009-to-2011 in the left column, 2012-to-2013 in the middle column and 2017-to-2018 in the right column. Each assimilation experiment is used to produce a forecast at Earth (black lines), a forecast at STEREO-A (red lines) and a forecast at STEREO-B where available (blue lines). Forecasts are verified against the science-level observations at the respective location. Here, and throughout the text, where Earth is used as a forecast verification, this is at the L1 point and so is using data from either ACE or DSCOVR, depending on the respective time period. Forecasts produced using science-level data are shown with a solid line and those using NRT data with a dashed line.

As Figure 8 shows, in general there is little difference between the real time and science forecasts produced using ACE and DSCOVR data. This means that assimilating these data in an operational setting would still produce forecasts of a similar skill to forecasts produced with science-level data.

There is more difference between forecasts based on NRT and science-level data when assimilating only STEREO data. Due to the issues with the STEREO-A beacon data described in Section 2.1 producing a systematic error in the NRT observations, we see a larger difference between the dashed and solid lines for all forecast locations in the 2009-to-2011 panel when assimilating only STEREO-A. This issue is not present in the 2012-to-2013 or 2017-to-2018 data, and we therefore see the NRT and science forecasts producing much more similar results. The forecasts from the STEREO-B real time data in the 2012-to-2013 (approximately solar maximum) interval show greater deviation from the science forecasts than for the 2009-to-2011 (solar minimum) interval.

For the 2009-to-2011 and 2012-to-2013 time intervals, the forecasts assimilating ACE and STEREO data shows the impact from the age of observations, whereby there is a large increase in forecast error when the forecast lead time exceeds the corotation time between the assimilated spacecraft and the forecast location. This is described in more detail in (Turner et al., 2022).

Figure 9 shows the simultaneous assimilation of ACE, STEREO-A and STEREO-B science-level (solid lines) and NRT (dashed lines) data, used for forecasts verified at Earth, STEREO-A and STEREO-B (black, red and blue respectively). Also included

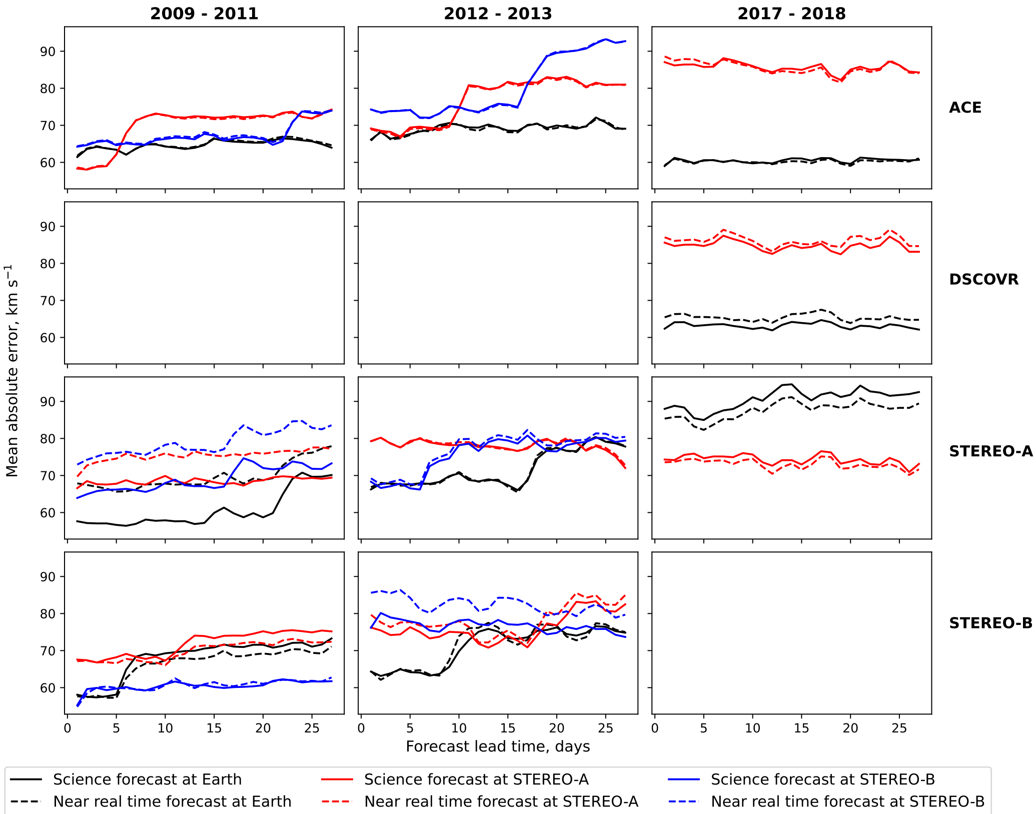

Figure 8. Comparison of solar wind speed forecast MAE for experiments assimilating observations from a single spacecraft, shown in the four rows. For each experiment, forecast MAE is shown at three locations; Earth (black lines), STEREO-A (red lines) and STEREO-B (blue lines). The solid lines show forecasts produced using science-level data and the dashed lines show forecasts using near-real-time data. Three time intervals are shown; 2009/08/01 to 2011/02/01, 2012/04/01 to 2013/10/01 and 2017/07/01 to 2019/01/01. Note that due to loss of communication with STEREO-B, there is no data available for the latest interval.

in this plot is the prior forecast, shown in the dotted line, and the L1 corotation forecast using science-level observations verified at Earth in the light grey shaded region. The prior forecast is the forecast produced from previous available information, before the data assimilation is performed. In this case, the prior forecast is the HelioMAS solution from the photospheric magnetic field that is propagated radially outwards using the HUX solar wind model. The left hand panel is for the 2009-to-2011 interval and the right hand panel for the 2012-to-2013 interval.

Firstly, as Figure 9 shows, especially for Earth, it is clear that assimilating either NRT or science level observations offers a significant improvement in forecast skill from the prior state. Secondly, using L1 corotation (also known as recurrence or persistence) as a baseline forecast, whereby we lag the observations by 27 days and use them as a forecast, we also find an improvement over all lead times using DA. For the 2009-to-2011 and 2012-to-2013 intervals for Earth, L1 corotation gives MAEs of $68.9kms^{-1}$ and $79.8kms^{-1}$ respectively. Using DA also offers improvement over a forecast produced from L1 corotation as it reconstructs the whole domain between the Sun and Earth's orbital radius and provides an updated inner boundary condition that can be used in MHD models. This allows for the propagation of CMEs through the improved background solar wind, something which cannot be achieved through a simple corotation forecast. With CMEs being the main driver of severe space weather, this offers the opportunity to improve their forecasted speed and arrival time.

It can be seen that there is no major difference between the NRT and science forecasts for the earlier interval. Particularly for the 2009 - 2011 interval, it could be expected that the lowest MAE would be seen for forecasts at STEREO-A due to the other observations being closer in longitude behind the spacecraft (with respect to solar rotation). However, it is seen that the lowest MAE are seen for forecasts at Earth. The trends for both Earth and STEREO-A are similar, but there is a systematic offset due to different structures being encountered at the spacecraft over a limited time period. The difference is likely not meaningful due to this reason.

For the 2012-to-2013 interval, from a forecast lead time of approximately 10 days, the forecasts produced using NRT observations appear to perform better than those produced with the science-level observations. As demonstrated below, this improvement comes about due to the NRT-based forecasts producing a 'flatter' solar wind speed time series that doesn't contain the full variability of the observations. Thus, if timing errors are present in both the science-level and NRT-based forecast, the science forecast would suffer greater penalty when assessed by MAE [e.g. Figure 1 of M. J. Owens (2018)]. This is demonstrated in Figure 10, where the number of high-speed events in the forecast time series using the science-level observations (black line) is greater than those using NRT observations (red line) for all lead times. Here, we define a high-speed event as having a solar wind speed greater than 500 km s^{-1}. This encapsulates both CMEs and fast solar wind streams. Both science- and NRT-based forecasts underestimate the number of high-speed events compared with observations, as expected as high-speed CMEs are not captured by the steady state data assimilation.

The forecast characteristics can be displayed using a Taylor diagram, as shown in Figure 11, which summarises the forecast MAE and linear correlation coefficient with the verification data, as well as the standard deviation of the forecasts. As forecasts improve, they move closer to the observation location, shown as a black star. It can be seen that the NRT and science forecasts group into two areas of roughly equal distance from the ideal forecast, but with the science forecasts having a standard deviation more representative of the observations. We can also see that there is an evolution of forecast MAE as the lead time increases, with the longer lead times producing forecasts with a higher MAE.

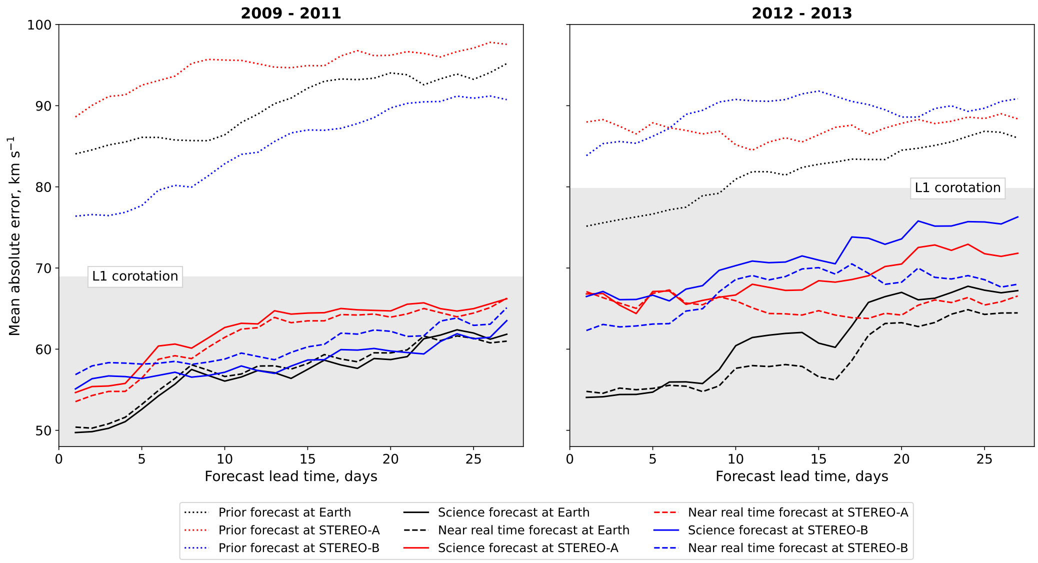

Figure 9. Comparison of solar wind speed forecast MAE for experiments assimilating all available observations; near-Earth, STEREO-A and STEREO-B. Forecast MAE is shown at three locations; Earth (black lines), STEREO-A (red lines) and STEREO-B (blue lines). The solid lines show forecasts produced using science-level data, the dashed lines show forecasts using near-real-time data and the dotted lines show forecasts using the prior estimate (i.e. with no DA). L1 corotation forecast error verified at Earth is shown in the light grey shaded region. Two time intervals are shown; 2009/08/01 to 2011/02/01 and 2012/04/01 to 2013/10/01.

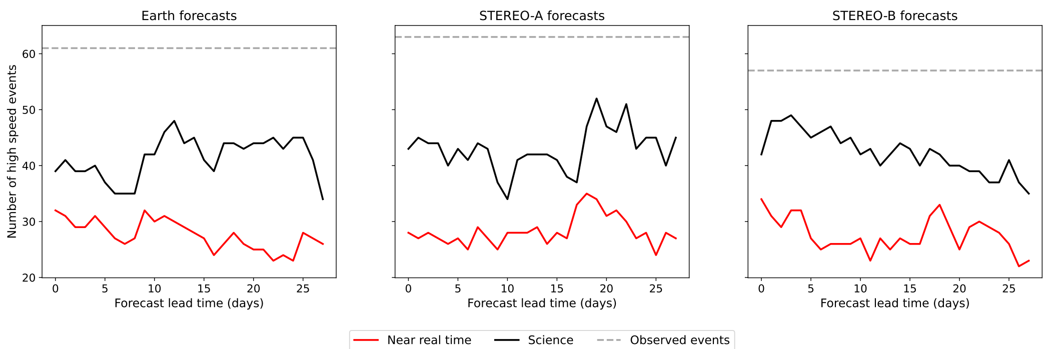

Figure 10. Number of forecast fast events ($> 500kms^{-1}$) for different lead times for forecasts created using near-real-time and science-level. Observed events are seen in the 1-hour resolution science-level observation time series and shown as the grey-dashed lines. Interval shown; 2012/04/01 to 2013/10/01.

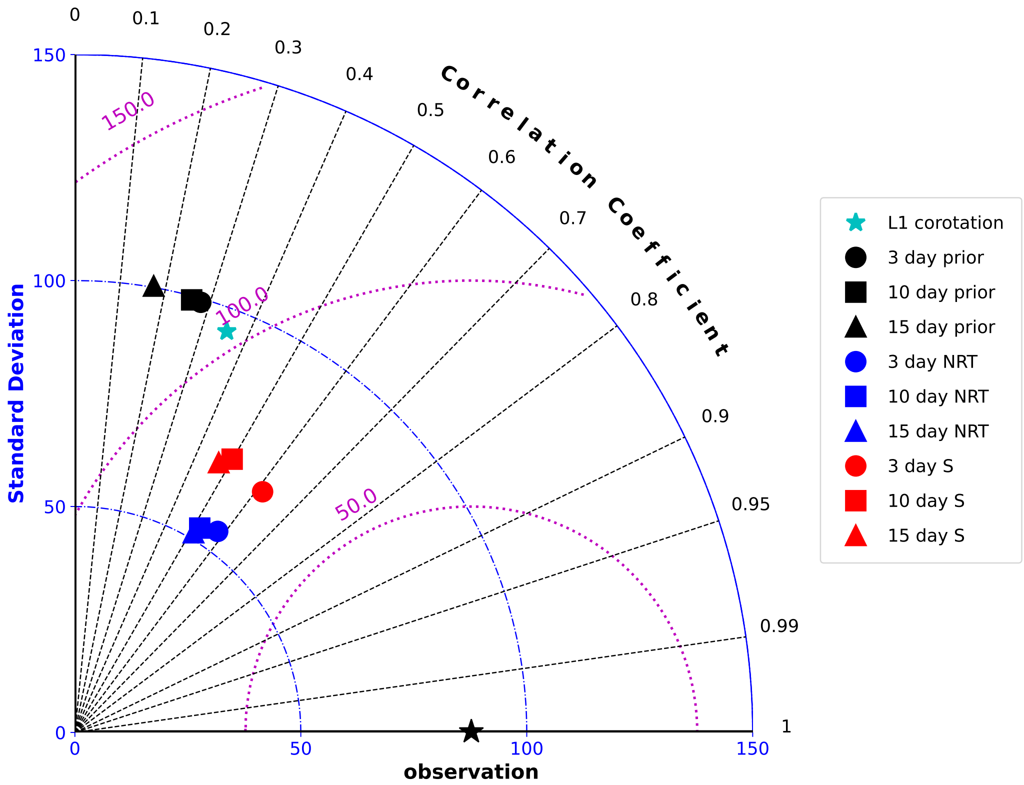

Figure 11. Taylor diagram of 3- (circle), 10- (square) and 15- (triangle) day lead-time forecasts of solar wind speed from 2012/04/01 to 2013/10/01 using the prior (black), near-real-time observations (NRT, blue) and science-level observations (S, red). The L1 corotation forecast for this interval, verified at Earth, is shown as the cyan star. Black radial lines show the correlation coefficient between the forecast and the verification values, the blue circular lines show the standard deviation and the purple circular lines show the forecast MAE. The observation metrics are shown with the black star.

4.2 L5 experiments

The future *Vigil* mission offers a chance for an operational data assimilation scheme to make routine use of simultaneous L5 and L1 data. To test this scenario, we can use combinations of STEREO and ACE data during specific intervals to mimic such a pairing. The forecast at the effective L1 position can then be assessed, as that would be Earth in an operational setting. Four intervals (Table 1) were identified where the spacecraft longitudinal separation was between 50 and 70 degrees, and BRaVDA was run with both NRT and science-level observations. Two sets of experiments were run; assimilating both effective L1 and L5 data and assimilating the effective L1 only. This allows the forecast gains from the L5 mission to be assessed.

Figure 12 shows the forecast MAE variation with forecast lead time. The prior is shown in the solid black line, the L1 only assimilated observations in red and the L1 and L5 assimilated observations in blue. The assimilated science data is shown in the solid coloured lines and the NRT data in dashed. Also shown on this plot are the L1 corotation forecast errors verified at the effective L1 spacecraft in the grey shaded region. These forecasts are made using the science-level observations. The forecasts produced using DA show similar forecast errors to L1 corotation, except in panel d), where the error from corotation is similar to that of the prior. The time interval covered in panel d), 2013/10/25 to 2014/02/09, is at approximately solar maximum, whereas the intervals in panels a-c) are in solar minimum. This means that there are likely more CMEs observed during this time, which cannot be captured in corotation forecasts and would therefore lead to a larger forecast error.

We also compare the DA forecasts to those produced using corotation from L5. Due to the separation of the spacecraft, it takes approximately 5 days for the solar wind to corotate round from the effective L5 spacecraft to the effective L1 spacecraft. As a consequence, the forecast lead time is approximately 5 days, thus giving a L5 corotation forecast. This forecast produces a lower MAE than L1 corotation, due to the shorter amount of time through which the solar wind can evolve whilst the Sun rotates from observation point to forecast point. The darker grey shaded region in Figure 12 shows the MAE from L5 corotation for each associated time interval. For panel a), the DA outperforms L5 corotation in both instances. For panels b) and d), assimilation of L1 and L5 gives the lowest error, whereas L5 corotation outperforms assimilation of L1 only. In panel c), L5 corotation gives the lowest MAE. Although DA offers no significant improvement over L5 corotation purely through MAE, its advantages come from the reconstruction of the whole domain and updating of the inner boundary condition. This means that it can be used to inform and improve MHD models and also allows for CMEs to be propagated through an updated background solar wind. This could lead to improved CME arrival and speed predictions.

The NRT and science-level observations have very similar forecast errors, with no major difference between the solid and the dashed lines. There is one exception; assimilating only STEREO-A NRT as the effective L1. This forecast shows a larger MAE of approximately 10 kms^{-1}, as this interval contains the period of time where there is much lower solar wind speeds in the NRT data when compared with the science-level data, as shown in Figure 4.

In general, it can be seen that the assimilation of both L5 and L1 does not offer a large forecast gain for forecast lead times greater than 4-5 days. However; for less than 5 days, the assimilation of L1 and L5 is $9.0 \pm 1.1 kms^{-1}$ lower in MAE. This is because the corotation time associated with 60 degrees of separation is 4.5 days. Thus the effective age of observations increases significantly after around 4 days, as discussed in Turner et al. (2022).

To further summarise these results, we average the four panels in Figure 12 to give Figure 13, which shows the improvement in the first 5 days of forecast lead time more clearly. Comparing the assimilation of only L1 and of both L1 and L5 against the forecast using the prior information, we can see significant improvements, with a percentage decrease (absolute difference), averaged over all lead times, of $42.7 \pm 3.3\%$ ($44.5 \pm 3.5kms^{-1}$) and $46.3 \pm 3.3\%$ ($48.2 \pm 3.4kms^{-1}$) respectively. Over all lead times, inclusion of L5 in the assimilation provides a $6.2 \pm 1.7\%$ decrease ($3.7 \pm 1.0kms^{-1}$) in MAE from assimilating only L1. However; in the first five days of forecast lead time, there is a $15.1 \pm 1.8\%$ ($9.0 \pm 1.1kms^{-1}$) decrease when including L5 data. This is compared to a $4.1 \pm 1.6\%$ ($2.5 \pm 1.0kms^{-1}$) decrease for lead times greater than 5 days.

As Figure 13 shows, assimilation of both the science and NRT observations for both L1 only and L1 and L5 performs better than corotation from L1. Only assimilation of L1 and L5 together performs better than corotation from L5. However, as discussed above, the DA offers improvements over simple corotation due to it updating the whole domain and for allowing the propagation of CMEs through its output.

Figure 14 summarises the prior and NRT forecast metrics in a Taylor diagram. The forecasts from the prior information are shown in black, assimilation of L1 and L5 NRT data in blue and only L1 NRT in red. Three lead times are shown; 3 days represented with a circle, 10 days with a square and 15 days with a triangle. The observation metrics are shown with a black star. L1 corotation is shown with the cyan star and L5 corotation with the cyan plus. We can see that assimilating L1 and L5 reduces the variability (standard deviation, blue axis) compared to just L1, so there is not much of an improved forecast for lead time greater than 5 days, despite the lower MAE (purple axis). However; for lead times less than 5 days (the blue circle), despite the correlation and standard deviation remaining similar to the other forecasts, there is a genuine improvement in the MAE when including L5 data.

5 Conclusions

In this study we have assessed the performance of the BRaVDA scheme with near-real-time (NRT) observations from the STEREO, ACE and DSCOVR missions. Previous work has been based on the pre-processed, science-level data, but for a solar wind data assimilation scheme to be used operationally it must perform well with NRT data. The forecasts using NRT observations were verified against the science observations, as they are assumed to best represent reality.

Using three test intervals, 2009/08/01 to 2011/02/01 (approximately solar minimum), 2012/04/01 to 2013/10/01 (approximately solar maximum) and 2017/07/01 to 2019/01/01 (interval with DSCOVR availability), BRaVDA was first tested by assimilating individual sources of observations. It was found that for L1 spacecraft (i.e. ACE and DSCOVR), the NRT and science observations produced forecasts with no significant difference, despite there being some quality issues within the input observation time series.

The NRT STEREO observations were found to be more problematic. In the NRT STEREO-A observations, a period of approximately three months at the end of 2009 had anomalously low NRT values compared to the science-level data. This problem gradually worsened over the three months before the NRT values returned close to the science-level observations in 2010/01. The effect of this was seen in the comparison between the DA-forecasts produced using the NRT and science observations, whereby the NRT forecasts have a greater MAE of approximately $10kms^{-1}$. This problem does not occur in the later two periods, showing that the quality of the observations needs to be continually assessed so that issues can be addressed in a timely manner. From a straight comparison between NRT and science data, it is not obvious what will cause a problem in

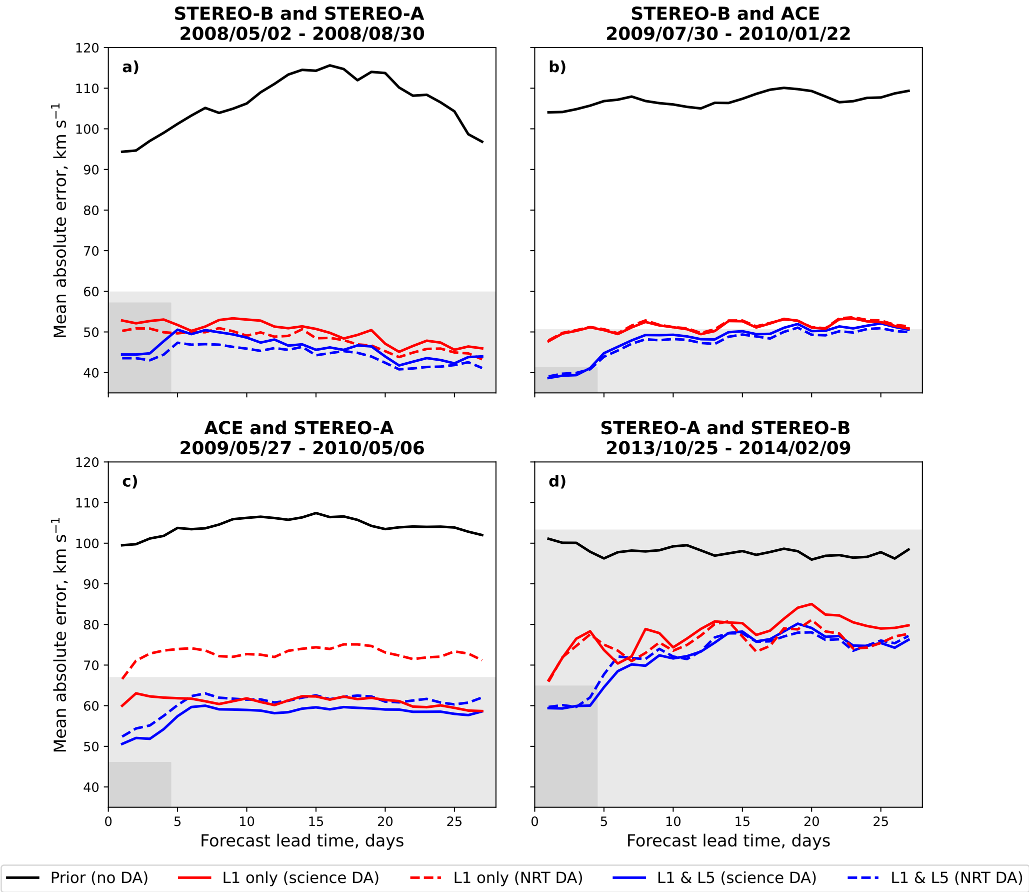

Figure 12. (a - d) Solar wind speed forecast MAE for experiments assimilating only effective L1 (red) and both effective L1 and L5 (blue) observations. The solid lines show forecasts produced using science-level observations and the dashed lines using near-real-time (NRT) observations. The black lines show forecasts produced using the prior (i.e., no DA). L1 corotation forecast error verified at the effective L1 spacecraft is shown in the light grey shaded region and the L5 corotation forecast is shown in the dark grey region for each time period, up to its maximum lead-time of 5 days. The panel a) covers the time period 2008/05/02 to 2008/08/30, panel b) covers 2009/07/30 to 2010/01/22, panel c) covers 2009/05/27 to 2010/05/06 and panel d) covers 2013/10/25 to 2014/02/09.

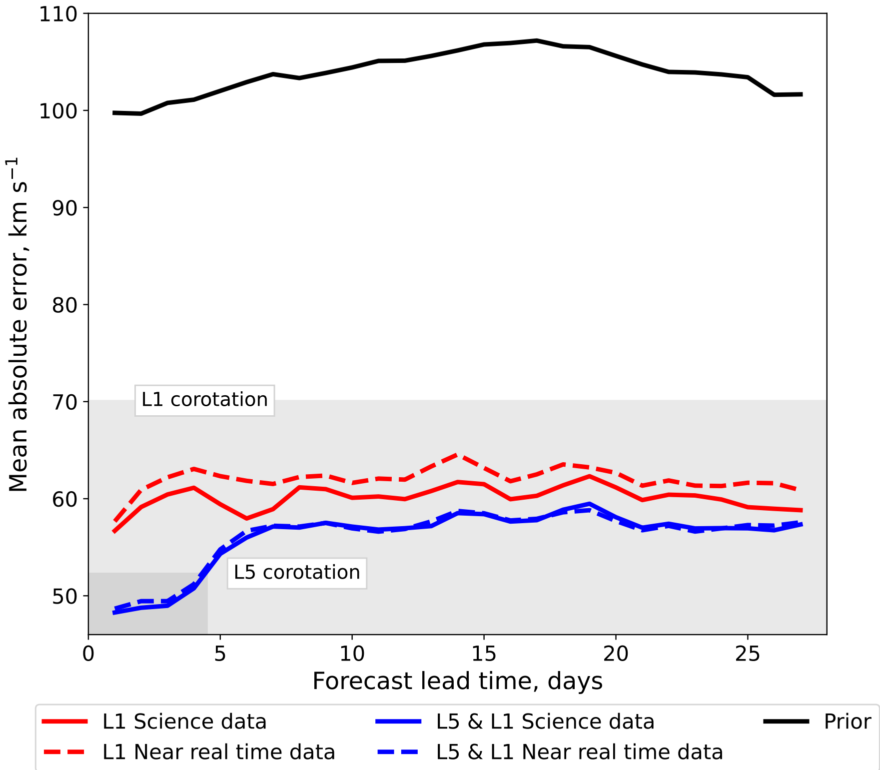

Figure 13. Average of the four L5 experiments shown in Figure 12. Mean absolute error is shown as a function of forecast lead time. The black line shows the forecast produced using the prior, the red line shows the forecasts from assimilation of only the effective L1 observations and the blue line shows assimilation of both L1 and L5. The coloured solid lines use science-level observations and the dashed lines use near-real-time (NRT) observations. The lighter grey shaded region shows the L1 corotation forecast error and the darker shaded region shows the L5 corotation forecast error, averaged from Figure 12.

.

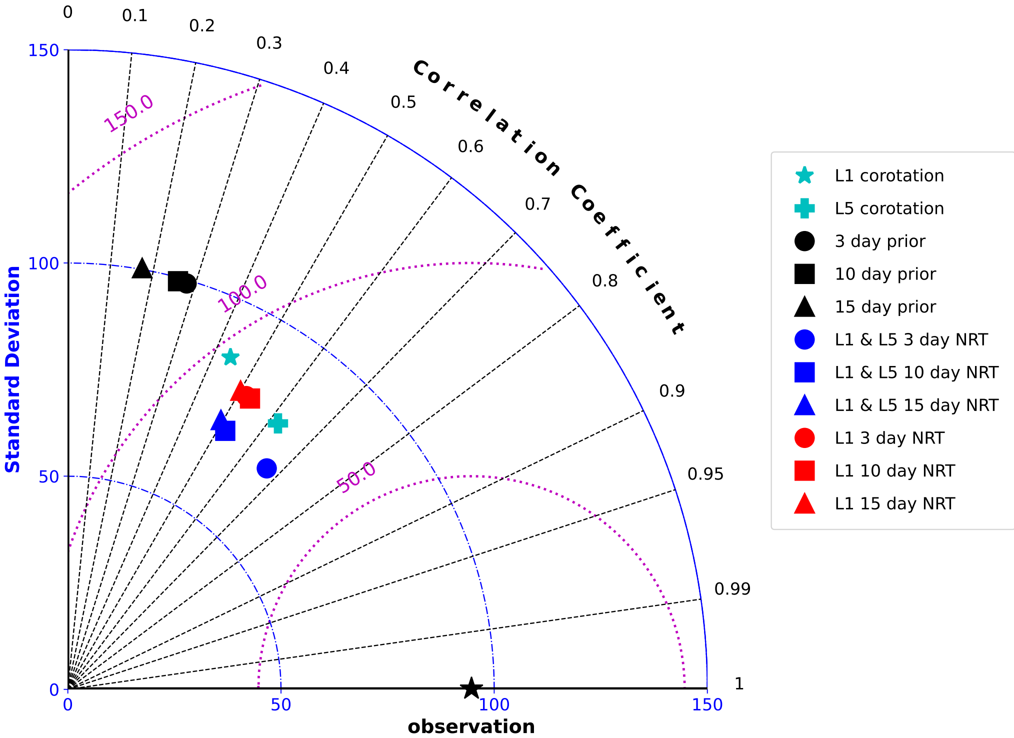

Figure 14. Taylor diagram of selected lead time for the prior forecasts (black), L1 NRT only forecasts (red) and L1 and L5 NRT forecasts (blue). 3-day lead time is shown with a circle, 10-day with a square and 15-day with a triangle. The cyan star shows the L1 corotation forecast and the cyan plus shows the L5 corotation forecast, both averaged over the four intervals. The observation metrics are shown with a black star. Note that the red circle is overlaid by the red square.

the assimilation. So it is important to periodically assess the forecast quality by checking previous NRT forecasts against newly made forecasts using science-level data once it is available.

The STEREO-B NRT observations contain a large amount of noise (i.e. high frequency variations) at roughly the hour timescale compared to the science-level observations. As a result, the STEREO-B NRT data produces an inferior forecast in regards to MAE at the position of STEREO-B itself. At other spacecraft locations, however, there is little difference between NRT and science-level forecasts. The reasons for this difference are not obvious, but may be due to the specific solar wind conditions due to these relatively short intervals. However, as the STEREO-B example shows, despite a slight worsening of the forecast error in one instance, the DA copes well with random errors. In the case of STEREO-B, these were large and of the order of $50kms^{-1}$ on an hour timescale. Comparing this to the systematic error seen in a few months of the STEREO-A data, we see that this produces a systematic error in the forecast. This is due to the assumption of a non-biased prior and observations in the formation of the data assimilation framework in general. This is well-known and accounted for in numerical weather prediction (D. P. Dee, 2006; D. Dee & Uppala, 2008), and bias correction methodologies have been developed. This is an area of active research, which this study falls within, and will be improved in future versions. DA can be used to correct and identify biases in input data, whereas corotation cannot.

BRaVDA was also tested with assimilation of multiple spacecraft observations from ACE, STEREO-A and STEREO-B, for both science and NRT. It was found that assimilation of both science and NRT observations performed better that the prior forecasts (i.e. without DA). Comparing these against a benchmark forecast of L1 corotation, we also see an improvement when using DA. As DA updates the entire domain, rather than a single point forecast that is produced from corotation, its solution can be used to initialise MHD models and allows for the propagation of CMEs through its output. This is not possible using corotation, thus the DA forecast model framework adds significant value to solar wind forecasting.

The future mission to the L5 Lagrange point, *Vigil*, offers the possibility of an operational DA scheme utilising routine NRT data from two vantage points. It is hoped that this will lead to large improvements in solar wind forecasting, but has not been tested from a DA perspective. For this purpose, we used BRaVDA with pairs of the STEREO spacecraft and ACE when they were separated in longitude between 50 and 70 degrees. The forecast was assessed at the effective L1 spacecraft (i.e. 50-70 degrees ahead with respect to solar rotation) to mimic a forecast at Earth. It was found that the NRT observations produce forecasts that are not significantly different to those created with the science-level observation. When these four intervals are averaged together, there is very little difference between the NRT and science forecasts. However, there is a significant improvement when compared to an example of a prior forecast. There is an average improvement of 46.3 ($\pm$3.3)%, showing that DA could offer large improvements to solar wind speed forecasting.

The assimilation of effective L1 and L5 observations was compared against assimilation of effective L1 only. We find improvement from L1 corotation, for both L1 only and L1 and L5, and a similar forecast error to L5 corotation for similar lead times for L1 and L5. As stated above, the DA offers value over corotation as it allows for the whole domain to be updated and for the propagation of CMEs. Although including the L5 observations did not provide a large improvement over L1 only for forecast lead times of more than 5 days, it did offer a 15.1 ($\pm$1.8)% decrease in forecast MAE for lead times less than 5 days. This lead time is of great interest for space weather forecasting, and so the future mission to L5 could be a step forward for solar wind forecasting capability, if solar wind DA is used operationally to exploit these observations.

6 Open Research

STEREO science data were downloaded from the CDAWeb Data Explorer portal at `https://cdaweb.gsfc.nasa.gov/` and STEREO NRT data from `https://stereo-ssc.nascom.nasa.gov/data/beacon/`. ACE science data were also downloaded from CDAWeb and the NRT data from NASA's Community Coordinated Modelling Centre at `https://ccmc.gsfc.nasa.gov/requests/GetInput/get_ace_K.php`. Both DSCOVR science and NRT data were downloaded from the DSCOVR Space Weather Data Portal at `https://www.ngdc.noaa.gov/dscovr/portal/index.html#/`. The code for BRaVDA is available at `https://github.com/University-of-Reading-Space-Science/BRaVDA`. HelioMAS output can be found on the Predictive Science website at `https://www.predsci.com/portal/home.php`.

Acknowledgments

We would like to thank the providers of the data for this project. Harriet Turner is funded through SCENARIO Grant No. NE/S007261/1. Mathew Owens is funded through STFC grants ST/V000497/1 and ST/V00235X/1.

References

Arge, C. N., Odstrcil, D., Pizzo, V. J., & Mayer, L. R. (2003). Improved Method for Specifying Solar Wind Speed Near the Sun. *AIP Conference Proceedings*, *679*, 190–193. doi: 10.1063/1.1618574

Biesecker, D. A., Webb, D. F., & St. Cyr, O. C. (2008). STEREO Space Weather and the Space Weather Beacon. In *The stereo mission* (pp. 45–65). New York, NY: Springer New York. Retrieved from `http://link.springer.com/10.1007/978-0-387-09649-0{\_}4` doi: 10.1007/978-0-387-09649-0_4

Burt, J., & Smith, B. (2012). Deep space climate observatory: The DSCOVR mission. *IEEE Aerospace Conference Proceedings*(November 2001). doi: 10.1109/AERO.2012.6187025

Cannon, P. S. (2013). Extreme space weather - A report published by the UK royal academy of engineering. *Space Weather*, *11*(4), 138–139. doi: 10.1002/swe.20032

Cargill, P. J. (2004). On the aerodynamic drag force acting on interplanetary coronal mass ejections. *Solar Physics*, *221*(1), 135–149. doi: 10.1023/B:SOLA.0000033366.10725.a2

Case, A. W., Spence, H. E., Owens, M. J., Riley, P., & Odstrcil, D. (2008). Ambient solar winds' effect on ICME transit times. *Geophysical Research Letters*, *35*(15), 1–5. doi: 10.1029/2008GL034493

Dee, D., & Uppala, S. (2008). Variational bias correction in ERA-Interim. (October), 26. Retrieved from `http://rda.ucar.edu/datasets/ds627.1/docs/IFS{\_}documentation/ifs.2b.variational{\_}bias{\_}correction.pdf`

Dee, D. P. (2006). Bias and data assimilation. *Quarterly Journal of the Royal Meteorological Society*, *131*(613), 3323–3343. doi: 10.1256/qj.05.137

Galvin, A. B., Kistler, L. M., Popecki, M. A., Farrugia, C. J., Simunac, K. D. C., Ellis, L., … Steinfeld, D. (2008). The plasma and suprathermal ion composition (PLASTIC) investigation on the STEREO observatories. *Space Science Reviews*, *136*(1-4), 437–486. doi: 10.1007/s11214-007-9296-x

Gosling, J. T., & Pizzo, V. J. (1999). Formation and Evolution of Corotating Interaction Regions and Their Three Dimensional Structure. *Space Science Reviews*, *89*, 21–52. doi: 10.1007/978-94-017-1179-1_3

Kaiser, M. L., Kucera, T. A., Davila, J. M., Cyr, O. C., Guhathakurta, M., & Christian, E. (2008). The STEREO mission: An introduction. *Space Science Reviews*, *136*(1-4), 5–16. doi: 10.1007/s11214-007-9277-0

Lang, M., & Owens, M. J. (2019). A Variational Approach to Data Assimilation in the Solar Wind. *Space Weather*, *17*(1), 59–83. doi: 10.1029/2018SW001857

Lang, M., Witherington, J., Owens, M. J., & Turner, H. (2021). Improving solar wind forecasting using Data Assimilation. *Space Weather*, 1–23.

Linker, J. A., Mikić, Z., Biesecker, D. A., Forsyth, R. J., Gibson, S. E., Lazarus, A. J., ... Thompson, B. J. (1999, may). Magnetohydrodynamic modeling of the solar corona during Whole Sun Month. *Journal of Geophysical Research: Space Physics*, *104*(A5), 9809–9830. Retrieved from `https://onlinelibrary.wiley.com/doi/10.1029/1998JA900159` doi: 10.1029/1998JA900159

Lockwood, M., Bentley, S. N., Owens, M. J., Barnard, L. A., Scott, C. J., Watt, C. E., & Allanson, O. (2019). The Development of a Space Climatology: 1. Solar Wind Magnetosphere Coupling as a Function of Timescale and the Effect of Data Gaps. *Space Weather*, *17*(1), 133–156. doi: 10.1029/2018SW001856

Loto'aniu, P. T. M., Romich, K., Rowland, W., Codrescu, S., Biesecker, D., Johnson, J., ... Stevens, M. (2022). Validation of the DSCOVR Spacecraft Mission Space Weather Solar Wind Products. *Space Weather*, *20*(10). doi: 10.1029/2022sw003085

Luntama, J. P., Kraft, S., & Glover, A. (2020). ESA Lagrange mission for enhanced space weather monitoring. In *100th american meteorological society annual meeting.*

McComas, D. J., Bame, S. J., Barker, P., Feldman, W. C., Phillips, J. L., Riley, P., & Griffee, J. W. (1998). Solar wind electron proton alpha monitor (SWEPAM) for the advanced composition explorer. *Space Science Reviews*, *86*(1-4), 563–612. doi: 10.1007/978-94-011-4762-0_20

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 A: Space Physics*, *121*(4), 2866–2890. doi: 10.1002/2015JA022200

Migliorini, S., & Candy, B. (2019). All-sky satellite data assimilation of microwave temperature sounding channels at the Met Office. *Quarterly Journal of the Royal Meteorological Society*, *145*(719), 867–883. doi: 10.1002/qj.3470

Odstrcil, D. (2003). Modeling 3-D solar wind structure. *Advances in Space Research*, *32*(4), 497–506. doi: 10.1016/S0273-1177(03)00332-6

Odstrcil, D., Riley, P., & Zhao, X. P. (2004). Numerical simulation of the 12 May 1997 interplanetary CME event. *Journal of Geophysical Research: Space Physics*, *109*(A2), 1–8. doi: 10.1029/2003JA010135

Owens, M. (2020). *Solar-Wind Structure* (No. July). doi: 10.1093/acrefore/9780190871994.013.19

Owens, M. J. (2018). Time-Window Approaches to Space-Weather Forecast Metrics: A Solar Wind Case Study. *Space Weather*, *16*(11), 1847–1861. doi: 10.1029/2018SW002059

Owens, M. J., Arge, C. N., Spence, H. E., & Pembroke, A. (2005). An event-based approach to validating solar wind speed predictions: High-speed enhancements in the Wang-Sheeley-Arge model. *Journal of Geophysical Research: Space Physics*, *110*(A12), 1–10. doi: 10.1029/2005JA011343

Owens, M. J., Challen, R., Methven, J., Henley, E., & Jackson, D. R. (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. doi: 10.1002/swe.20040

Owens, M. J., & 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. doi: 10.1002/2017SW001679

Parker, E. N. U. o. C. (1958). Dynamics of the Interplanetary Gas and Magnetic Fields. *Astrophysical Journal*, 664 – 676. Retrieved from `https://sci-hub`

.tw/10.1002/iroh.19620470121

Richardson, I. G., & Cane, H. V. (2012). Solar wind drivers of geomagnetic storms during more than four solar cycles. *Journal of Space Weather and Space Climate*, *2*, 1–9. doi: 10.1051/swsc/2012001

Riley, P., & Ben-Nun, M. (2021). On the Sources and Sizes of Uncertainty in Predicting the Arrival Time of Interplanetary Coronal Mass Ejections Using Global MHD Models. *Space Weather*, *19*(6), 1–15. doi: 10.1029/2021SW002775

Riley, P., Ben-Nun, M., Linker, J. A., Owens, M. J., & Horbury, T. S. (2017). Forecasting the properties of the solar wind using simple pattern recognition. *Space Weather*, *15*(3), 526–540. doi: 10.1002/2016SW001589

Riley, P., Linker, J. A., & Mikić, Z. (2001). An empirically-driven global MHD model of the solar corona and inner heliosphere. *Journal of Geophysical Research: Space Physics*, *106*(A8), 15889–15901. doi: 10.1029/2000ja000121

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. doi: 10.1007/s11207-011-9766-x

Schrijver, C. J. (2015). Socio-Economic Hazards and Impacts of Space Weather: The Important Range between Mild and Extreme. *Space Weather*, *13*(9), 524–528. doi: 10.1002/2015SW001252

Smith, A. W., Forsyth, C., Rae, I. J., Garton, T. M., Jackman, C. M., Bakrania, M., ... Johnson, J. M. (2022). On the Considerations of Using Near Real Time Data for Space Weather Hazard Forecasting. *Space Weather*, *20*(7), 1–20. doi: 10.1029/2022SW003098

Stone, E. C., Frandsen, A. M., & Mewaldt, R. A. (1998). The Advanced Composition Explorer. *Space Science Reviews*, *86*, 1–22. Retrieved from 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. doi: 10.1029/2018SW001821

Tóth, G., Sokolov, I. V., Gombosi, T. I., Chesney, D. R., Clauer, C. R., De Zeeuw, D. L., ... Kóta, J. (2005). Space weather modeling framework: A new tool for the space science community. *Journal of Geophysical Research: Space Physics*, *110*(A12), 1–21. doi: 10.1029/2005JA011126

Turner, H., Owens, M., Lang, M., & Gonzi, S. (2022). Quantifying the effect of ICME removal and observation age for in situ solar wind data assimilation. *Space W*, 1–20. doi: 10.1029/2022SW003109

Webb, D. F., & Howard, T. A. (2012). Coronal mass ejections: Observations. *Living Reviews in Solar Physics*, *9*. doi: 10.12942/lrsp-2012-3

Zhao, X. P., Plunkett, S. P., & Liu, W. (2002). Determination of geometrical and kinematical properties of halo coronal mass ejections using the cone model. *Journal of Geophysical Research: Space Physics*, *107*(A8), 1–9. doi: 10.1029/2001JA009143