



## FLIP-MHD-based model sensitivity analysis

C. Skandrani<sup>1</sup>, M. E. Innocenti<sup>2</sup>, L. Bettarini<sup>3</sup>, F. Crespon<sup>1</sup>, J. Lamouroux<sup>1</sup>, and G. Lapenta<sup>2</sup>

<sup>1</sup>NOVELTIS, Space and Remote Sensing Department, Space Weather Unit, rue du Lac 153, 31670 Labège, France

<sup>2</sup>Center for Mathematical Plasma Astrophysics, Department of Mathematics, KULeuven – University of Leuven, Celestijnenlaan 200B, bus2400, 3001 Leuven, Belgium

<sup>3</sup>Fluid and Plasma Dynamics group, Department of Physics, Université Libre de Bruxelles, Campus Plaine – CP231 Boulevard du Triomphe, 1050 Brussels, Belgium

Correspondence to: C. Skandrani (chafih.skandrani@noveltis.fr)

Received: 4 March 2013 – Revised: 4 March 2014 – Accepted: 5 March 2014 – Published: 24 April 2014

**Abstract.** The state of the art in the forecast of the background solar wind speed and of the interplanetary magnetic field at Earth is based on the use as boundary conditions for heliospheric models of the input data provided by solar observations. Magnetogram synoptic maps are used to obtain information on the magnetic field configuration at the solar source surface. Magnetic field inputs at the solar source surface thus constitute one of the main external sources of errors in solar wind models. The assimilation of data into forecasting models used in the terrestrial domain showed the ability to control model state errors. A sensitivity study performed through the analysis of the ensemble variances and the representers technique is used here to assess how process and model state errors propagate in a nonlinear two-dimensional MagnetoHydro Dynamic (MHD) system. The aim is to understand the impact of the source surface input parameters on the evolution of MHD heliospheric models and the potentialities of data assimilation techniques in solar wind forecasting. The representer technique in fact allows one to understand how far from the observation point the improvement granted from the assimilation of a measure propagates.

### 1 Introduction

The interplanetary medium, although once considered a perfect vacuum, is actually a turbulent area dominated by the solar wind, which flows at the average velocities of  $400 \text{ km s}^{-1}$  (slow wind) and  $750 \text{ km s}^{-1}$  (fast wind) (McComas et al., 1998). Among the events which may enhance and provoke a jump in the solar wind speed, coronal mass ejections (CMEs) are the most disrupting. CMEs are one of the most ener-

getic and interesting solar phenomena. They propel magnetic clouds with a mass up to  $10^{17} \text{ g}$  to speeds up to and above  $2600 \text{ km s}^{-1}$  into the heliosphere (Dwivedi and Mohan, 2005), influencing near-Earth plasma conditions, causing fluctuations in the magnetic field near Earth and in the ionospheric density and driving geomagnetic storms and auroras. However, CMEs are not the only source of fluctuations in the solar wind parameters. Coronal holes produce high-speed streams (Levine et al., 1977; Sheeley and Harvey, 1981) and the alternation of slow and fast flows generates perturbations named co-rotating interaction regions (CIRs) (Gosling and Pizzo, 1999). CIRs can themselves impact the near-Earth space. The understanding of these slow and fast flows is of prime importance for the forecast of interplanetary CME (ICME) propagation and associated threats (Vršnak and Zic, 2007). Such threats are relevant for a variety of human assets, both in space, air borne and ground based. Space weather impacts society in diverse ways (Baker, 1998; Siscoe, 2000; Bothmer and Daglis, 2007): apart from posing direct threats to astronauts' lives, space weather events may damage satellite equipment and alter the atmospheric drag on satellites. This causes satellite displacements and losses and disrupts communication and positioning signals. Airline passengers may be exposed to ionising radiations and induced currents at ground level may damage ground-based operations such as pipelines, the power grid and telecommunication cables. Thus there will be a continued and intensified need for space environment services to address health, safety and commercial issues. In recent years, the increased availability of solar wind measurements has allowed the scientific community to develop models and tools aimed both at better understanding the dynamics of solar wind propagation and at

enabling solar wind now-casting and forecasting at 1 astronomical unit (AU) and beyond.

Two main kinds of heliospheric models exist: MagnetoHydro Dynamics (MHD) based models, such as MHD-Around-A-Sphere (MAS, Riley et al., 2001), ENLIL (Odstrcil, 2003) and SWMF/SC/IH (Space Weather Modelling Framework/Solar Corona/ Inner Heliosphere, Toth et al., 2005), and empirically based models, such as the Wang–Sheeley–Arge (WSA) (Wang and Sheeley, 1990) and the model described in Vršnak et al. (2007).

Physics-based kinetic models are still too computationally expensive to be used for large-scale heliospheric simulations. However, recently, methods such as the implicit adaptive Multi Level Multi Domain method (Innocenti et al., 2013; Beck et al., 2013) have been developed with the aim of increasing the portion of space which can be simulated with kinetic codes at a reasonable computational cost.

In the first two kinds of models observations are used as inputs or boundary conditions. See, for example, Wu et al. (2006), where a data-driven MHD model is continuously fed with SOHO/MDI magnetogram data to study the evolution of active regions starting from the observed magnetic field configuration.

A third, less commonly employed direction may be explored in heliospheric modeling: data assimilation (DA) techniques (Bouttier and Courtier, 1999; Evensen, 2009) may be integrated into existing models to enhance their forecasting abilities, as recently explored in Innocenti et al. (2011). There, an empirical model is enhanced with the application of a Kalman filtering DA technique (Kalman, 1960; Welch and Bishop, 2001). It is noticed, among other results, that the assimilation of solar wind temperature observations into the model dramatically improves the forecasts of the solar wind speed and also grants the baseline model some resilience against CME activity.

In Schrijver and Derosa (2003), instead, SOHO/MDI magnetograms are assimilated into a flux dispersal model for the evolution of the photospheric magnetic field, which is then mapped to the solar source surface through a Potential Field Source Surface model. In Barrero Mendoza et al. (2006), an ensemble Kalman filtering (EnKF) technique is applied on top of an MHD model. The variation of the root mean square error for the state estimate is studied as a function of the position and number of the assimilation points. The case study examined is a magnetic storm in the terrestrial magnetosphere. A similar approach is used here for the region of space extending from the source surface to 1 AU. No planetary magnetospheres are considered at this point of the investigation.

The study presented aims at further investigating the topic of DA in the area of space weather forecast. DA combines observational data with the outputs of numerical simulations to produce an optimal estimate of the evolving state of the system. Thus, by assimilating space environment data into global numerical models, a better understanding of the

physics of the interplanetary space and increased forecasting capabilities can be expected. DA has been successfully used by meteorologists and oceanographers in the last twenty years. On the other hand, the space physics community has been more reluctant to implement such techniques.

Applications of DA techniques to the near-Earth system can be found for electron dynamics in the radiation belts (Kondrashov et al., 2007; Rigler et al., 2004) and for global assimilation of ionospheric measurements (GAIM) (Schunk et al., 2003, 2004). Such a rather limited use may be traced to the fact that the space system is much larger and that there is a shortage of measurements when compared to meteorological and oceanographic systems. A massive amount of information for terrestrial weather exists, while the number of satellites collecting space weather-relevant data in the magnetosphere and in the region of space between the Sun and the Earth is rather low. The most prominent of them are SOHO (Solar and Heliospheric Observatory – Fleck et al., 1995), ACE (Advanced Composition Explorer – Stone et al., 1998), the Cluster mission (Laakso et al., 2010) and the PROBA-2 (Project for On-Board Autonomy-2, Gantois et al., 2010) ESA mission. Both SOHO and ACE are located at L1, while the Cluster mission consists of four satellites in the Earth’s magnetosphere. PROBA2, instead, is located in a Sun-synchronous dusk–dawn orbit. The two STEREO spacecrafts, located in the Earth’s orbit but in orbital positions ahead and behind the Earth, can be used for space weather data collection as well, as shown in Turner and Li (2011).

However, more and more observational data will be available soon. The NASA Living With a Star programme promises to provide an incredible amount of data regarding all of the different aspects of the Sun–Earth connection. The first mission, Solar Dynamics Observatory (Pesnell et al., 2012), was launched in 2010 and the community is just starting to use the huge amount of data it delivers.

The application of data assimilation to space sciences will benefit not only from this soon-to-be-available amount of observational data, but also from the recent development of heliospheric models.

One of the main goals of the present paper is to assess the kind of benefits that DA could bring to MHD heliospheric models. These preliminary feasibility studies are essential for understanding the potentialities of DA in the specific field and thus for deciding if more effort should be put into the activity.

As a starting point, the first setting considered is a simplified two-dimensional MHD model of the solar wind between the Sun and the Earth including the Lagrange point L1. The physics-based model used here is the FLIP MHD model (Brackbill, 1991) in a simplified two-dimensional configuration. A description of the model is provided in Sect. 2.

It is well known (see, for example, Wu et al., 2006) that a major source of forecast errors for heliospheric models is the use of non-optimal boundary conditions at the photosphere or at the source surface.

Another purpose of the present paper is to investigate if and how the use of statistical techniques for data assimilation could lead to obtaining a more accurate estimate of magnetogram input parameters and an improved initialisation state, in order to predict the future model variables in the whole area better.

The methodology adopted is reviewed in Sect. 3 and is based on ensemble Monte Carlo techniques (Evensen, 2009). Multiple runs of the model are initialised with statistically guided modifications of the input. Then, the sensitivity study is conducted by diagnosing process and model state errors for the nonlinear two-dimensional MHD system simulated with FLIP-MHD.

The results of the ensemble statistics are presented in Sect. 4.

First, in Sect. 4.1, it is checked if the model errors are distributed according to a Gaussian distribution function. Then, in Sect. 4.2, the model ensemble variances are used to characterise the model sensitivity. The ensemble variance analysis provides an objective tool for evaluating how initial condition errors affect a forecast and whether it would be desirable to gather additional observations to reduce model forecast errors. The variances in time and space of the errors for the solar wind magnetic field and velocity are in fact calculated for model runs which differ only for their initial conditions and can therefore be considered to be the model response to the uncertainties in the source surface inputs.

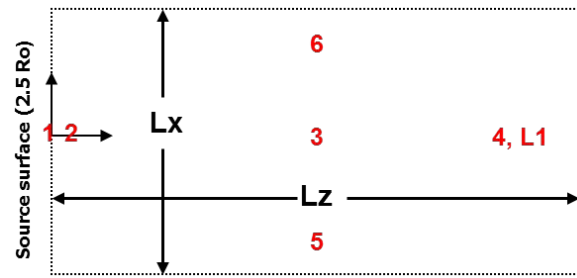
Section 4.3 studies the domain of influence of observations made at a given location on the whole forecasting domain and especially at the source surface. This so-called “representers analysis” allows one to understand at a glance how the system augmented with assimilation propagates the information spatially. Then, it becomes easier to see the potential impact of model improvement using DA.

In the Conclusion a summary of the work presented is provided and further considerations are drawn on the role of data assimilation in space weather forecasting models.

## 2 The FLIP-MHD model

The numerical model used in this study is the FLIP-MHD model, a 2-D resistive MHD code described in Brackbill (1991).

Two directions,  $z$  and  $x$ , are simulated, as depicted in Fig. 1. The direction  $z$  is the longitudinal or streamwise direction, oriented away from the Sun. Open boundary conditions are set at  $z = 0$  and  $z = L_z$ , the domain length in the  $z$  direction. In the transverse or cross-stream direction  $x$  reflecting boundary conditions are set. All three velocity components are retained. The following viscous-resistive compressible MHD equations in dimensionless form, written in the Lagrangian frame, are solved:



**Fig. 1.** Sketch of the domain simulated with FLIP-MHD. The left boundary corresponds to the source surface, conventionally located at 2.5 solar radii. The right boundary is placed at the terrestrial atmosphere.  $L_x$  and  $L_z$  are the box size in the  $x$  and  $z$  directions. The red numbers mark the diagnoses locations, as explained in Sect. 3.

Mass continuity:

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot \rho \mathbf{v} \quad (1)$$

Momentum equation:

$$\rho \frac{\partial \mathbf{v}}{\partial t} = -\rho (\mathbf{v} \cdot \nabla) \mathbf{v} - \nabla p - \left[ \frac{\nabla |\mathbf{B}|^2}{2} - (\mathbf{B} \cdot \nabla) \mathbf{B} \right] + \frac{1}{\Re_\nu} \Delta \mathbf{v} \quad (2)$$

Magnetic flux equation (Faraday’s law):

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{v} \times \mathbf{B}) + \frac{1}{S} \Delta \mathbf{B} \quad (3)$$

Energy equation

$$\rho \frac{\partial I}{\partial t} = -\rho \nabla \cdot \mathbf{v} + \frac{(\nabla \cdot \mathbf{v})^2}{\Re_\lambda} + \frac{\mathbb{I} \cdot \mathbb{I}}{\Re_\nu} + \frac{\mathbf{J} \cdot \mathbf{J}}{S}, \quad (4)$$

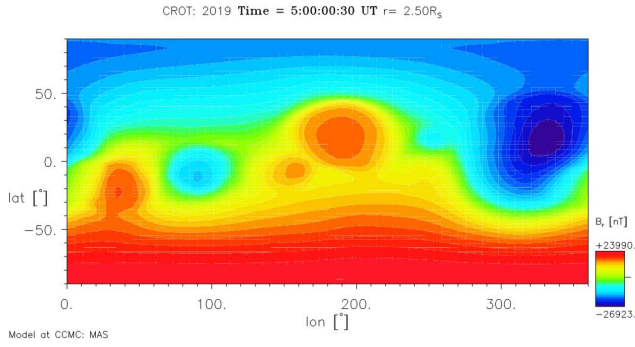
where  $\rho$  is the mass density,  $\mathbf{B}$  is the magnetic field which obeys the Gauss law constraint  $\nabla \cdot \mathbf{B} = 0$ ,  $\mathbf{v}$  is the fluid velocity,  $I$  is the specific internal energy,  $\mathbf{J}$  is the current density,  $p$  is the fluid pressure, and  $\mathbb{I}$  the symmetric rate-of-strain tensor defined as

$$\mathbb{I} = \frac{1}{2} [\nabla \mathbf{v} + \nabla \mathbf{v}^T]. \quad (5)$$

In Eqs. (2)–(4)  $\Re_\lambda$  and  $\Re_\nu$  are the Reynolds numbers measuring respectively the global kinematic and dynamic viscosity in the numerical box; in the simulations shown,  $\Re_\lambda = \Re_\nu = 10^3$  is always assumed. The quantity  $S = 10^3$  is the Lundquist number measuring the global explicit resistivity set within the domain.

### 2.1 Definition of the source function

Currently, operational prediction models are driven by input data at the source surface, which is conventionally located



**Fig. 2.** Example of an extrapolated magnetic field at the source surface. From CCMC, <http://ccmc.gsfc.nasa.gov>.

at a distance of 2.5 solar radii from the solar surface. The boundary conditions at the source surface are in turn obtained from observations at the photosphere, such as GONG magnetogram synoptic maps. Figure 2 shows an example of the extrapolated magnetic field at the source surface obtained using the MAS tool (Riley et al., 2001) at the Community Coordinated Modeling Center, CCMC (<http://ccmc.gsfc.nasa.gov>).

In this paper, the FLIP-MHD model is driven by a synthetic analytical function providing the input conditions for the magnetic and velocity fields at the source surface. For simplicity, the velocity at the source surface is taken as constant. Instead, the boundary condition for the  $z$  component of the magnetic field at the source surface is imposed as follows:

$$b_z = b_0 \tanh(x - x_{\text{centre}}) + \frac{b_0}{f(t)} \sin\left(\frac{2\pi(x - x_{\text{centre}})}{L_B}\right) \quad (6)$$

$x_{\text{centre}}$  is the middle point of the domain in the  $x$  direction.

The equations solved in the code and reported in the results below use normalised units. The magnetic field is normalised to  $b_0$ , the unit length is defined by the thickness  $L_B$ , and the unit time is computed as the time required by an Alfvén wave to travel one unit distance  $L_B$ , the so-called Alfvén time:  $\tau_A = L_B/v_A$ . Note that in this choice of units the reference Alfvén speed is  $v_A = 1$ .

The magnetic field at the source surface varies in time as determined by the function  $f(t)$  defined as

$$f(t) = \begin{cases} 0.5(1 - \cos(\pi(1 - t/t_\alpha))) & t < t_\alpha \\ 1 & \text{Otherwise} \end{cases} \quad (7)$$

$t_\alpha = t_{\text{final}} \cdot \alpha$ . The maximum value of  $t_\alpha$  is reached before the final time,  $t_{\text{final}}/\tau_A = 400$ , when  $\alpha$  is less than 1. In particular, we select a value  $\alpha = 0.4$ , to introduce an initial phase of dynamic perturbation and subsequently observe the relaxation of the system to a constant drive.

The  $y$  and  $x$  components of the magnetic field and the velocity in the longitudinal direction are selected to ensure an initial force-free equilibrium (that is, no force is acting on

the plasma):

$$b_y = -\frac{b_0}{\cosh(x - x_{\text{centre}})}, \quad (8)$$

$$b_x = 0, \quad (9)$$

and

$$u = 1. \quad (10)$$

With these boundary conditions, the evolution of the solar wind magnetic field and velocity is studied between the source surface and 1 AU.

The real heliospheric case is of course spherical. However, a full 3-D description is beyond the scope of the present paper that aims instead at ascertaining the sensitivity of MHD models to boundary drives. For this reason and at the cost of losing the effects of the geometric expansion, a Cartesian grid is used instead. Figure 1 shows the domain used.  $L_x$  and  $L_z$  are the system dimensions in the cross stream wise and stream wise direction respectively. The out of plane  $y$  dimension is assumed ignorable. In code units,  $L_x/L_B = 3.6$  and  $L_z/L_B = 98.4$ . The red numbers mark the diagnoses locations defined later in Sect. 3.

The solar wind speed is assumed to be Alfvénic,  $V_{\text{SW}} = V_A$ , at the injection boundary, and the total simulation time allows roughly 4 transit times in the  $z$  direction. When compared with the real heliospheric case, the distance  $L_z$  represents the Sun to Earth distance and therefore the total simulation time corresponds to approximately 4 times the total time required for the solar wind to reach the Earth. For typical solar wind speeds this corresponds to roughly 12 days. Of course, given the choice of a Cartesian grid this is only an indicative number.

### 3 Methodology

The model sensitivity analysis consists in describing the uncertainties of a reference simulation from a set of simulations over a chosen time period. Information about both the reliability of the model and about which input parameters contribute most to the variability of the model response is provided through such analysis. From the ensemble method (Evensen, 2009), ensemble variances and ensemble covariances can help improve the knowledge of how boundary conditions affect the evolution of the system and which areas are particularly sensitive to their variations. This could be helpful for the identification of the areas where dedicated observations may be useful for the data assimilation process.

Model errors can be simulated in a number of ways. The different options include using a different model, varying the resolution of the discretisation grid or altering some parameters in the model (Evensen, 2009). One of the available alternatives is to characterise the model sensibilities through ensemble variances (Evensen, 2009). This leads to the depiction of the model errors. In the framework of DA, this

ensemble technique is widely used to characterise the inaccuracies in the initialisation of the state of atmospheric and oceanographic forecasting models.

Several initial states are obtained by imposing small variations on top of the initial conditions. Then, starting from each of these initial states, calculations for future states are made. At a given time, all of the calculated states can be represented by a dispersion or probability distribution.

One of the preliminary steps before applying DA to a model is to derive this probability distribution of the state vector. It can be obtained through a Monte Carlo sampling method which describes the uncertainties with respect to a reference simulation of a set of similar simulations over a chosen period of time. The error covariances can thus be specified. The sample characterises the sensitivity of the model forecast to parameter errors around the current state of the system. Based on this ensemble technique, sensitivity analysis and representers study constitute the cornerstone of the approach presented here, leading to the preliminary understanding of how much DA will be effective.

The analysis includes three main stages:

1. an ensemble of 100 members is generated by perturbing the source function as described in Eq. (6);
2. the integration of the FLIP-MHD model over this set of perturbed initial states provides access to a set of model solutions, representing the probability density of the model state (model errors and uncertainties);
3. diagnoses on the simulated sets are conducted around some specific locations such as the Lagrange point L1 (location 4), near the source surface (location 2), in the middle of the domain (location 3) and to both sides of the longitudinal axis near the borders of the domain study (locations 5 and 6) (see Fig. 1).

In Sect. 3.1 more information regarding the generation of the perturbations is provided, while in Sect. 3.2, the ensemble diagnostic statistical tools are introduced in greater detail.

### 3.1 Generation of perturbations

Within the framework of this study, evolved forecast error covariances are constructed using an ensemble of model runs that start from perturbed initial states. Perturbations are built by adding uncertainty to the synthetic source surface magnetic field inputs in the streamwise  $z$  direction. The velocity, instead, is kept constant.

If  $F(x, t)$  is the analytical function at the source surface, a particular perturbed field  $F^{\text{pert}}(x, t)$  is obtained as in

$$F^{\text{pert}}(x, t) = F(x, t) + \gamma(t)F(x, t), \quad (11)$$

where  $\gamma(t)$  is a time-dependent coefficient generated from a Gaussian distribution with mean zero and standard deviation of 0.05. An error of about 5% with respect to the reference field is imposed.

For each member of the ensemble, the simulation is carried out for 100 time steps up to a final time of  $t_{\text{final}}/\tau_A = 400$ .

### 3.2 Ensemble theory and statistical tools

In this section the ensemble statistical tools used in this study are presented in detail. The concepts of representers technique and domain of influence are introduced, based on the ensemble Kalman filter method.

#### 3.2.1 Ensemble Kalman filter and representers

The representers technique has been used widely for oceanic or tidal circulation problems as reported in the literature (Bennet, 1992; Egbert and Bennet, 1996; Lyard, 1999; Echevin et al., 2000; Evensen, 2009). At this stage of the preliminary analysis, it is a simple and robust tool for investigating how the system with DA will spatially propagate the information brought by dedicated observations and how strong the correlation between the observation taken at a particular spatial location and the other grid points is. A brief summary of the ensemble Kalman filter technique is provided here with the aim of facilitating the reading of the rest of the paper.

In the framework of the ensemble Kalman filter, a series of forecasting and analysis steps are executed in a sequence.

First, the forecasting step provides model predictions from the last analysis state and calculates the forecast errors of the model by propagating the analysis errors appropriately.

Then the analysis step at time  $t_i$  corrects the forecasted state  $\mathbf{x}^f(t_i)$  with the information obtained from the observation vector  $\mathbf{y}^o(t_i)$ . The corrected state  $\mathbf{x}^a(t_i)$  is calculated according to the following formula:

$$\mathbf{x}^a(t_i) = \mathbf{x}^f(t_i) + \mathbf{K}(t_i) \left[ \mathbf{y}^o(t_i) - \mathbf{H}(t_i)\mathbf{x}^f(t_i) \right], \quad (12)$$

where  $\mathbf{x}^f(t_i)$  is the forecast of the state at time  $t_i$ ,  $\mathbf{y}^o(t_i)$  is the vector of observations at the same time and the matrix  $\mathbf{H}(t_i)$  is the so-called ‘‘observation operator’’. It ensures the projection of the model space onto the observation space. In this particular case, the state vector  $\mathbf{x}$  includes the solar wind magnetic and velocity fields at all grid points of the simulated domain;  $\mathbf{y}^o(t_i) - \mathbf{H}(t_i)\mathbf{x}^f(t_i)$  is the ‘‘innovation’’ matrix, which carries the information relative to the difference between the actual observations and the corresponding values for the forecasted fields;  $\mathbf{K}(t_i)$  is the Kalman gain matrix at time  $t_i$ , given by:

$$\mathbf{K}(t_i) = \mathbf{P}^f(t_i)\mathbf{H}^T(t_i) \left[ \mathbf{H}(t_i)\mathbf{P}^f(t_i)\mathbf{H}^T(t_i) + \mathbf{r}(t_i) \right]^{-1} \quad (13)$$

where  $\mathbf{P}^f(t_i)$  and  $\mathbf{r}(t_i)$  are the forecast error covariance and the observation error covariance matrices, respectively, at time  $t_i$ . The gain matrix  $\mathbf{K}(t_i)$  is the optimal least square gain at time  $t_i$ , i.e. the solution of the minimisation problem of the model observations error cost function (Ghil and

Malanotte-Rizzoli, 1991; Bouttier and Courtier, 1999). The matrix  $\mathbf{K}(t_i)$  weights the forecast and the observations according to their respective error statistics,  $\mathbf{P}^f(t_i)$  and  $\mathbf{r}(t_i)$ . Notice that Eqs. (12) and (13) are recalled to provide a general background to the Kalman filtering procedure used for data assimilation: indeed, the actual assimilation of heliospheric observations has not been carried out yet. The present work focuses in fact on the concept of representers, which is described below. The rest of this section aims at explaining the role of the representer matrices in this sensitivity study without actually performing the assimilation procedure and without having observations at our disposal.

Representers or influence functions are obtained by reformulating the analysis step of the Kalman filter described in Eq. (12), as in

$$\mathbf{x}^a(t_i) = \mathbf{x}^f(t_i) + \sum_{j=1}^{j=p_i} b_j(t_i) \mathbf{r}_j(t_i), \quad (14)$$

where  $\mathbf{x}^f(t_i)$  is the forecasted state vector at time  $t_i$ ,  $\mathbf{r}_j(t_i)$  is the representer or influence function relative to the  $j$ th observation at time  $t_i$  and  $p_i$  is the number of observations available at time  $t_i$ . The representer  $\mathbf{r}_j(t_i)$  is a vector which has the dimensions of the state vector and  $b_j(t_i)$  is the scalar coefficient associated with the observation  $j$  at time  $t_i$ .

Equation (14) translates in matrix form into

$$\mathbf{x}^a(t_i) = \mathbf{x}^f(t_i) + \mathbf{r}(t_i) \mathbf{b}(t_i), \quad (15)$$

with the matrix  $\mathbf{r}(t_i)$  defined at time  $t_i$  as

$$\mathbf{r}(t_i) = \mathbf{P}^f(t_i) \mathbf{H}^T(t_i), \quad (16)$$

and the vector  $\mathbf{b}(t_i)$  as

$$\mathbf{b}(t_i) = \left[ \mathbf{H}(t_i) \mathbf{P}^f(t_i) \mathbf{H}^T(t_i) + \mathbf{r}(t_i) \right]^{-1} \left[ \mathbf{y}^o(t_i) - \mathbf{H}(t_i) \mathbf{x}^f(t_i) \right]. \quad (17)$$

The matrix  $\mathbf{r}(t_i)$  contains as many columns as there are observations available at time  $t_i$  and each column constitutes the representer  $\mathbf{r}_j(t_i)$  relative to a given observation. The representer formulation of the analysis step of Eq. (14) can be interpreted as the correction of the forecasted state  $\mathbf{x}^f(t_i)$  done by using a combination of representers, depending on the forecast error covariance matrix  $\mathbf{P}^f(t_i)$ . Note that in the case of a multivariate state vector (e.g. solar wind magnetic and velocity fields), the information from the  $j$ th observation can be spread to the other model state variables, thanks to the error cross-covariance of these variables.

Notice that from this point onwards, the time dependence is dropped for ease of reading. It is also assumed that the model grid is composed of  $n$  nodes and that the state vector contains two variables (namely magnetic field and solar wind velocity). The dimension of the state vector is then  $2n$ .

Now, focusing on a single observation, Eq. (16) becomes, at a given time:

$$\mathbf{r}_j = \mathbf{P}^f \mathbf{H}_j^T \quad (18)$$

$\mathbf{H}_j$  is the row of the matrix  $\mathbf{H}$  corresponding to the  $j$ th observation. For the purpose of the demonstration, let us assume that the  $j$ th observation is (1) exactly located at a grid point  $k$  and (2) associated with one of the state vector variables. Thus, this observation is identified in the state vector by its index  $k_j$ . Then,  $\mathbf{H}_j(k_j)$  is the only non-zero element of  $\mathbf{H}_j$ , that is

$$\mathbf{H}_j = (0, \dots, 0, 1, 0, \dots, 0), \quad (19)$$

with only one “1” at index  $k_j$ . Thus, Eq. (18) gives, from the definition of  $\mathbf{P}^f$ ,

$$\mathbf{r}_{k_j} = \mathbf{P}^f \mathbf{H}_{k_j}^T = \begin{pmatrix} \text{cov}(\varepsilon_{k_j}^f, \varepsilon_1^f) \\ \vdots \\ \text{cov}(\varepsilon_{k_j}^f, \varepsilon_n^f) \\ \text{cov}(\varepsilon_{k_j}^f, \varepsilon_{n+1}^f) \\ \vdots \\ \text{cov}(\varepsilon_{k_j}^f, \varepsilon_{2n}^f) \end{pmatrix} \quad (20)$$

$\varepsilon^f$  is the model forecast error, defined as the difference between the forecasted and true states ( $\mathbf{x}^f - \mathbf{x}^t$ ), which is not known. The abbreviation “cov” stands for “covariance”.

Consequently, the representer of observation  $j$  contains the covariances between the model forecast errors at the observation point and all the grid nodes, for all the state variables (e.g. magnetic field, solar wind velocity).

As explained in Sect. 3, the model forecast errors can be estimated through the stochastic modelling approach: if an ensemble of simulations generated by perturbing one or several sources of errors of the model is available, these model forecast errors are approximated by statistics computed from this ensemble. Thus, the covariances of the model forecast errors are estimated by the ensemble covariances of the model forecast errors. Therefore

$$\begin{aligned} \text{cov}(\varepsilon_{k_j}^f, \varepsilon_i^f) &\approx \text{cov}^{\text{ens}}(\varepsilon_{k_j}^f, \varepsilon_i^f) \\ &= \frac{1}{m} \sum_{s=1}^m (\varepsilon_{k_j,s}^f - \frac{1}{m} \sum_{s=1}^m \varepsilon_{k_j,s}^f) (\varepsilon_{i,s}^f - \frac{1}{m} \sum_{s=1}^m \varepsilon_{i,s}^f) \\ &= \overline{(\varepsilon_{k_j}^f - \varepsilon_{k_j}^f)(\varepsilon_i^f - \varepsilon_i^f)}, \end{aligned} \quad (21)$$

where “ens” stands for “ensemble”,  $m$  is the number of members in the ensemble and the horizontal bar represents the ensemble average. Under the assumption of the BLUE (best linear unbiased estimator) hypotheses (Bouttier and Courtier, 1999; Evensen 2009), the model forecast errors are assumed

unbiased. This implies that

$$\begin{aligned} \text{cov}(\varepsilon_{k_j}^f, \varepsilon_i^f) &\approx \overline{(\varepsilon_{k_j}^f)(\varepsilon_i^f)} \\ &\approx \overline{(\mathbf{x}_{k_j}^f - \overline{\mathbf{x}_{k_j}^f})(\mathbf{x}_i^f - \overline{\mathbf{x}_i^f})} \\ &\approx \text{cov}(\mathbf{x}_{k_j}^f, \mathbf{x}_i^f). \end{aligned} \quad (22)$$

Thus, more generally, the variances, covariances and correlations of the model forecast errors are identifiable as the ensemble variances, covariances and correlations of the state variables.

### 3.2.2 Domain of influence of a single observation

Here an explicit expression for the domain of influence of a single observation is presented. Consider an isolated observation  $\mathbf{y}_{k_j}$  relative to the forecasted state variable  $\mathbf{x}^f$  (magnetic field components or solar wind velocity components) and located on a node  $k$  of the mesh. The correction applied on the forecasted state  $\mathbf{x}^f$  calculated by assimilation of the observation  $\mathbf{y}_{k_j}$  is expressed by

$$\delta \mathbf{x} = \mathbf{x}^a - \mathbf{x}^f = b_{k_j} \mathbf{r}_{k_j} \quad (23)$$

with

$$\mathbf{r}_{k_j} = \mathbf{P}^f \mathbf{H}_{k_j}^T = (\text{cov}(\varepsilon_{k_j}^f, \varepsilon_i^f))_{i=1,2n} \quad (24)$$

$$\begin{aligned} b_{k_j} &= \frac{\mathbf{y}_{k_j} - \mathbf{x}_{k_j}}{\text{cov}(\varepsilon_{k_j}^f, \varepsilon_{k_j}^f) + \text{cov}(\varepsilon_{k_j}^o, \varepsilon_{k_j}^o)} \\ &= \frac{\delta \mathbf{y}_{k_j}}{\text{var}(\varepsilon_{k_j}^f) + \text{var}(\varepsilon_{k_j}^o)} \end{aligned} \quad (25)$$

where  $\delta \mathbf{y}_{k_j}$  and  $\varepsilon_{k_j}^o$  are respectively the innovation and the observation error associated with the observation  $\mathbf{y}_{k_j}$ ; the abbreviation ‘‘var’’ stands for the ensemble variance.

Combining Eqs. (23), (24) and (25), the correction at node  $i$  brought by the observation  $\mathbf{y}_{k_j}$  is then expressed as follows:

$$\delta \mathbf{x}_i = \frac{\text{cov}(\varepsilon_{k_j}^f, \varepsilon_i^f)}{\text{var}(\varepsilon_{k_j}^f) + \text{var}(\varepsilon_{k_j}^o)} \delta \mathbf{y}_{k_j} \quad (26)$$

Introducing the formal definition of the correlation, the previous equation becomes

$$\begin{aligned} \delta \mathbf{x}_i &= \frac{\text{cov}(\varepsilon_{k_j}^f, \varepsilon_i^f)}{\sqrt{\text{var}(\varepsilon_{k_j}^f) \text{var}(\varepsilon_i^f)}} \delta \mathbf{y}_{k_j} = \frac{\text{correl}(\varepsilon_{k_j}^f, \varepsilon_i^f)}{1 + \frac{\text{var}(\varepsilon_{k_j}^o)}{\text{var}(\varepsilon_{k_j}^f)}} \sqrt{\frac{\text{var}(\varepsilon_i^f)}{\text{var}(\varepsilon_{k_j}^f)}} \delta \mathbf{y}_{k_j} \\ &\approx \text{correl}(\mathbf{x}_{k_j}^f, \mathbf{x}_i^f) \sqrt{\frac{\text{var}(\varepsilon_i^f)}{\text{var}(\varepsilon_{k_j}^f)}} \delta \mathbf{y}_{k_j} \\ &\approx \text{correl}(\mathbf{x}_{k_j}^f, \mathbf{x}_i^f) \times \text{modulation factor}(\varepsilon_{k_j}^o, \varepsilon_{k_j}^f, \varepsilon_i^f) \delta \mathbf{y}_{k_j}, \end{aligned} \quad (27)$$

where the abbreviation ‘‘correl’’ refers to correlation.

Therefore, the correction at point  $i$ , obtained by the assimilation of the isolated observation  $\mathbf{y}_{k_j}$ , depends on the correlation between the model forecast errors associated with the observed variable  $\mathbf{x}_{k_j}$  and the variable to be corrected  $\mathbf{x}_i$ . According to Eq. (22), this correlation is approximated by the ensemble correlation of the observed variable  $\mathbf{x}_{k_j}$  and the variable to be corrected  $\mathbf{x}_i$ . This correlation term reflects the spatial extent of the correction and thus represents the *domain of influence* (hereafter ‘‘doi’’) of the observation  $\mathbf{y}_{k_j}$ . Note that the magnitude of the correction  $\delta \mathbf{x}_i$  is represented by the *modulation factor* term that depends on the observation error and on the ratio of the model error variances at the point  $i$  and at the observation point  $k$ : if the observation error is strong, the correction has to remain low; if the model error variance is higher/lower at the point  $i$  than at the observed point, the correction must be significant/lower.

As a first step in the implementation of a dedicated DA system, this study will focus on the doi of isolated observations (magnetic or solar wind velocity fields) rather than on the *modulation factor*.

## 4 Results on ensemble statistics

In this section the statistical ensemble variance and domain of influence analysis outlined above are presented. With these experiments, we inspect the potential impact of the DA technique on the improvement of the state at the source surface and, more in general, on the entire simulated domain. All parameters (spatial, time, magnetic and velocity field) in the paper should be considered normalised to the normalisation factors described earlier in Sect. 2.1.

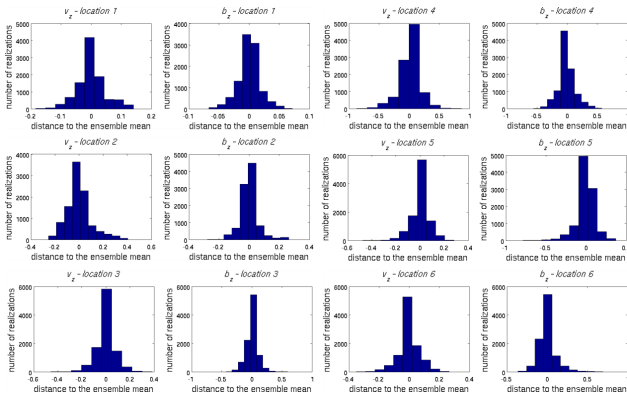
### 4.1 Gaussian character of error statistics

For the model error statistics exploration process of this study, a set of 100 ensemble members is generated. This number is usually chosen for stochastic modelling to ensure a statistical significance at a reasonable computational cost.

When implementing real DA experiment it should be necessary to conduct a convergence analysis to determine the optimal number of members leading to a sufficient stability of the statistics computed from the ensemble members.

The Gaussian nature of the perturbed states is very important for the data assimilation schemes since it ensures that the optimal state calculated in the analysis step is the most probable in the sense of maximum likelihood. In the context of ensemble methods, the forecast error of the model can be estimated from the dispersion of the perturbed model trajectories around the overall average at a given time step.

Figure 3 shows the distribution of the error for the  $z$  components of the solar wind velocity and magnetic field for the different members of the ensemble at the diagnosis locations of Fig. 1. Each time plot at each location depicts



**Fig. 3.** Distribution of the forecast errors for the  $z$  component of the solar wind velocity and magnetic field at the different test locations illustrated in Fig. 1: location 1 (source surface), location 2, location 3, location 4 (Lagrange point, L1), location 5 and location 6. The errors are calculated from the dispersion of the perturbed ensemble member trajectories around the ensemble average.

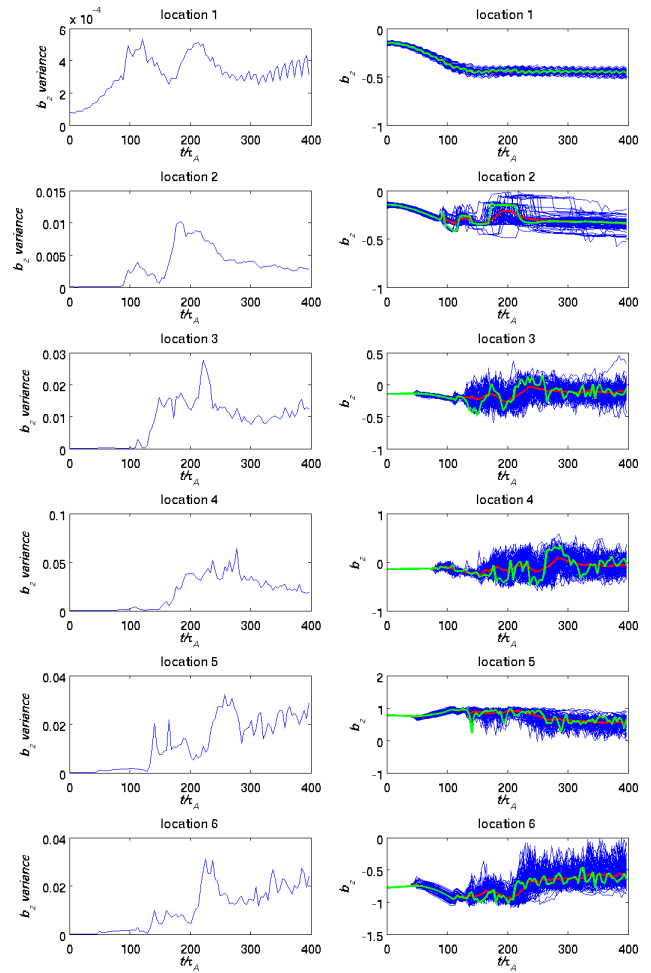
the information relative to all the ensemble members at all the time steps available. These histograms are thus computed on a sample of 10 000 (100 members times 100 outputs) achievements and show the empirical distributions of the model error as a function of the perturbed source surface input.

It can be noted that the shape of the error distribution is slightly broader for locations 4, 5 and 6 compared to locations 1, 2 and 3. This is due to a higher level of non-linear interactions and unsteady state conditions further away from the source surface.

Overall, the sampling distributions can be fitted, in a first approximation, with a quasi-Gaussian. This ensures that the BLUE (cf. Sect. 3.2) hypothesis can be made and that the classical formulation of data assimilation, on which the present study relies, is pertinent. However, it should be mentioned that some reconnection events are triggered in some ensemble members. This could explain, at some locations, the deviation from the Gaussian shape of the distribution of the errors.

## 4.2 Ensemble variances

As part of the sensitivity study, ensemble variance diagnoses for the magnetic field and velocity at the previously designated locations of the domain are discussed. The objective is to estimate the error structures of the model, in order to point out the areas where gathering dedicated observations may be useful. In this framework, the model response to perturbations of the boundary condition inputs at the source surface has been calculated. This will highlight the areas where physical processes are poorly modelled and will also allow estimating of the reliability of the model. In the first subsection, the temporal evolution of the ensemble variances is analysed



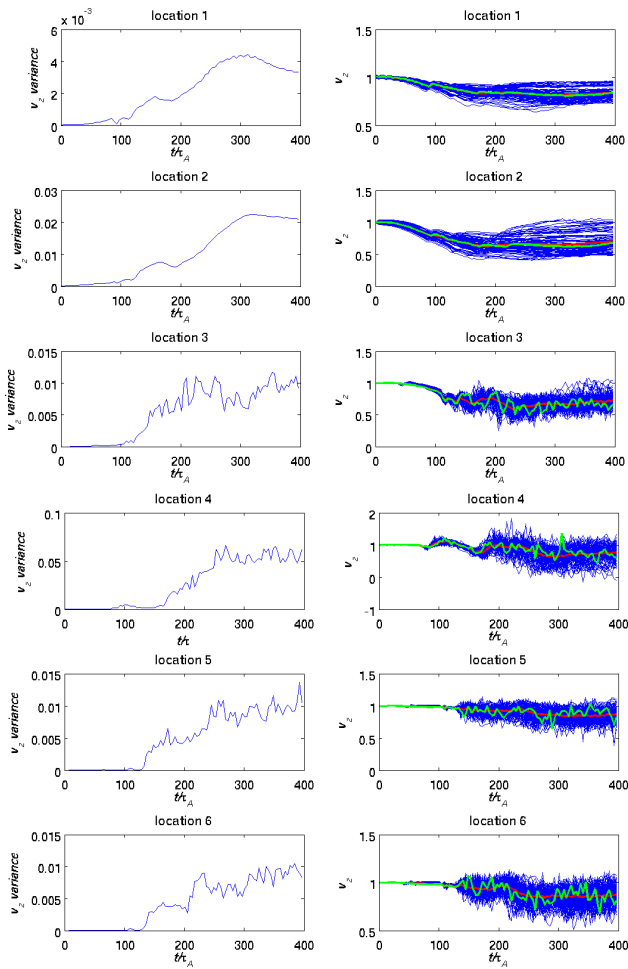
**Fig. 4.** Time evolution of the ensemble variance for the normalised  $b_z$  (left panels) and time evolution of the normalised  $b_z$  for all the ensemble members (right panels) at the selected diagnosis locations. The red curve represents the average of the ensemble and the green curve refers to the reference simulation, where no perturbation is applied to the source surface inputs.  $b_z$  is normalised to  $b_0$  and the time to the Alfvén time  $\tau_A$ . Note that the left and right  $b_z$  scales are not common to the six diagnosis points.

at fixed points in space. In the second subsection, the spatial structures of the ensemble variances are studied in the whole domain simulated.

### 4.2.1 Local temporal evolution of ensemble variances

The temporal evolution of the various perturbed simulations composing the ensemble provides empirical information about the time evolution of the error variances. It has to be recalled here that, while generating the ensemble, only  $b_z$  is perturbed, while  $v_z$  is left unperturbed. Figures 4 and 5 refer to the  $z$  component of the solar wind magnetic field and velocity respectively. They present a comparison of the temporal evolution of the ensemble variances (left panels)





**Fig. 5.** Time evolution of the ensemble variance for the normalised  $v_z$  (left panels) and time evolution of  $v_z$  for all the ensemble members (right panels) at the selected diagnosis locations. The red curve represents the average of the ensemble and the green curve refers to the reference simulation, where no perturbation is applied to the source surface inputs.  $v_z$  is normalised to the Alfvén speed  $v_A$  and the time to the Alfvén time  $\tau_A$ . Note that the left and right  $v_z$  scales are not common to the six diagnosis points.

and their associated maps of evolution for all of the ensemble members (right panels), for six representative locations chosen in the whole domain as depicted in Fig. 1. In the right panels, the red curve represents the temporal evolution of the ensemble average, while the green curve refers to the reference simulation whose source surface input is left unperturbed.

Considering the rate of perturbation that has been applied in this stochastic modelling, one can see that the evolution of the variability in the ensemble remains quite admissible and maintains fairly reasonable error bounds for both  $b_z$  and  $v_z$ .

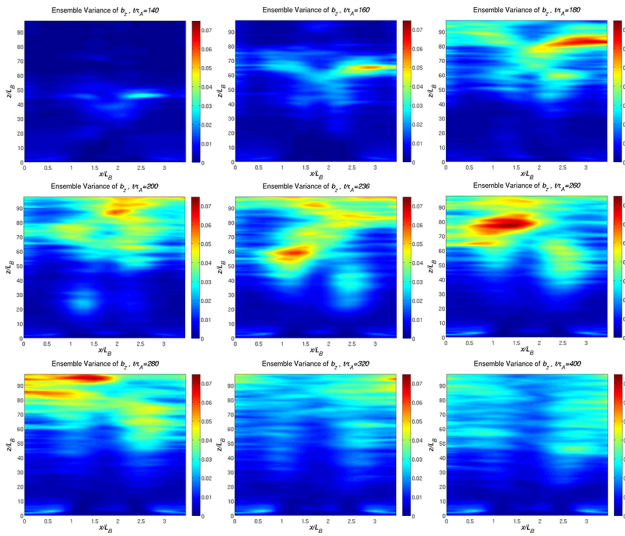
Except for location 1 near the source surface, where the variance remains very low, it can be noted here that the ensemble variance evolution is quite similar for the six loca-

tions: the variance of the ensemble is extremely low at the beginning of the time series. Reasonably, though, the locations far from the source surface show slightly higher differences in the evolution than the closer ones, as depicted in the right panels of Figs. 4 and 5. At later time steps,  $t/\tau_A \approx 88$  for location 2,  $t/\tau_A \approx 120$  for location 3,  $t/\tau_A \approx 140$  for locations 5 and 6 and  $t/\tau_A \approx 160$  for location 4, the variance starts increasing for all of the locations: the ensemble experiences rather different evolutions. These differences in the time evolution of the ensemble variances are the signature of a dynamical process which crosses the domain from the source surface to L1.

This notwithstanding, it can be noticed that ensemble averages and reference simulations behave quite similarly; this confirms that the stochastic modelling protocol is mostly linear, and thus corroborates the quasi-Gaussian feature of the distribution of the errors stated in Sect. 4.1.

#### 4.2.2 Spatial structure of the ensemble variance

The spatial distribution of the ensemble variance structures at a given instant reflects the level of model variability over the ensemble and hence gives an efficient estimation of the errors of the model in response to the perturbations of the  $z$  component of the magnetic field. The time evolution of the velocity variances is quite stable; however, a rather interesting time evolution is noticed for the magnetic field. In Fig. 6 several snapshots taken at different times show the spatial extent of the ensemble variance of the model for the  $z$  component of the magnetic field. Figure 7 shows the  $z$  component of the magnetic field with superimposed magnetic field lines, for the reference simulation, at the same time steps as in Fig. 6. These two figures give an efficient insight into the time evolution of the spatial distribution of the model error structures, in response to the perturbation of the  $z$  component of the magnetic field. In Fig. 6, some “bursts” can be observed in the ensemble dispersion, from time step  $t/\tau_A \approx 140$  to the final time step  $t/\tau_A \approx 400$ . These spatial structures correspond to the ones observed in the temporal evolution of the ensemble variances in Fig. 4. Figure 7, which depicts the magnetic field lines for the reference simulation, shows that the major increases in the ensemble variance seem to be due to some reconnection events in the simulated solar wind. For example, the reference simulation shows signs of reconnection in correspondence with the first model error’s bubble at  $t/\tau_A \approx 140$  and also with the model error’s bubble visible in the ensemble variance at  $t/\tau_A \approx 160$  on the right side of the domain, but not in correspondence with the model error’s bubble visible on the left-hand side. The disturbance, when present, is then advected in time to L1, thus provoking the time-shifted variance signals in the observation points closer to L1. The high variances in correspondence with reconnection events show that not all ensemble instances experience reconnection at the same time and in the same locations. The evolution over time shows that more of these events occur.



**Fig. 6.** Ensemble variance of the normalised  $z$  component of the magnetic field at different time steps and on the entire simulated domain. The spatial dimensions are normalised to  $L_B$ , the length scale of variation of the magnetic field. Time is normalised to the Alfvén time  $\tau_A$ .

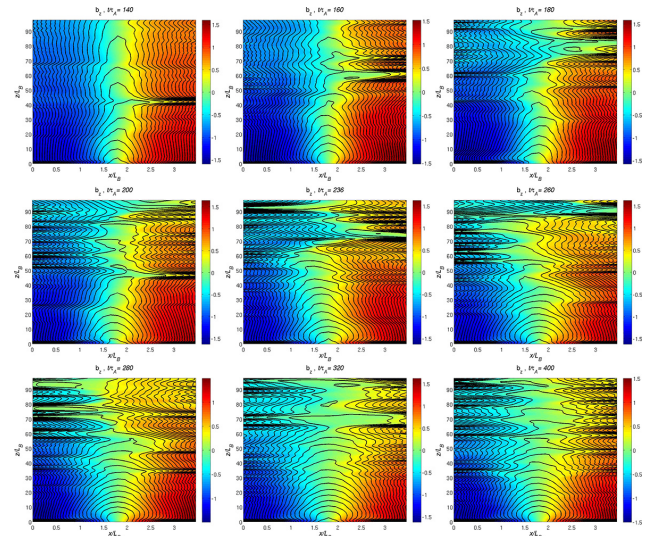
The simulation instances reconnecting “together” (i.e. at similar time and space) are presumably the ones excited with a similar magnetic field at the boundary.

### 4.3 Domains of influence

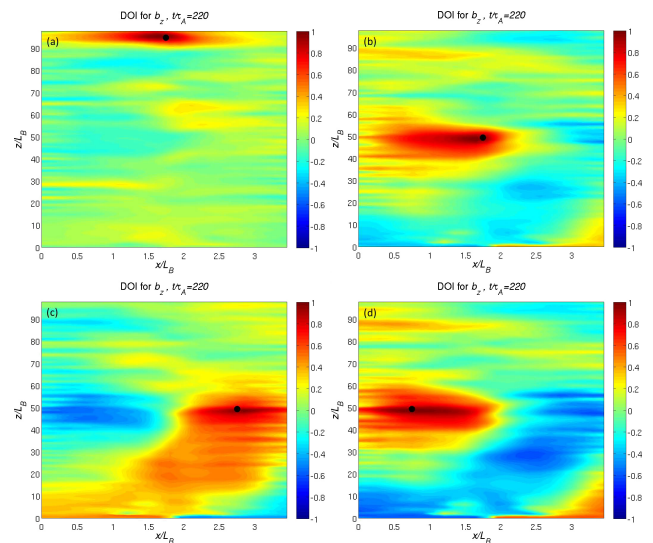
The study of the ensemble variance of the model allowed one to describe how the physical fields respond differently to a perturbation of the model inputs and to characterise the model forecast error. The ensemble modelling protocol that has been set up in the present study provides an efficient way of investigating the model forecast error covariances. These covariances indeed play a leading role in the problem of data assimilation since they ensure the propagation of the information provided by the observations to all model variables. Therefore they determine the spatial extent and magnitude of the state corrections provided by the assimilation of observations. Here the objective is indeed to illustrate how observations taken at a measurement point located in one particular position of the domain would affect the overall model evolution if those observations were assimilated into the model.

For real heliospheric application, the availability of observation points is currently very limited. The main observation point is the ACE spacecraft located at L1 position corresponding to location 4 in this study. As mentioned before, additional important information can be obtained by the STEREO mission.

In Fig. 8, the panels illustrate the domain of influence “doi” as defined in Eq. (27), for  $b_z$ , for observations taken at different locations. The effects of reconnection at the dif-



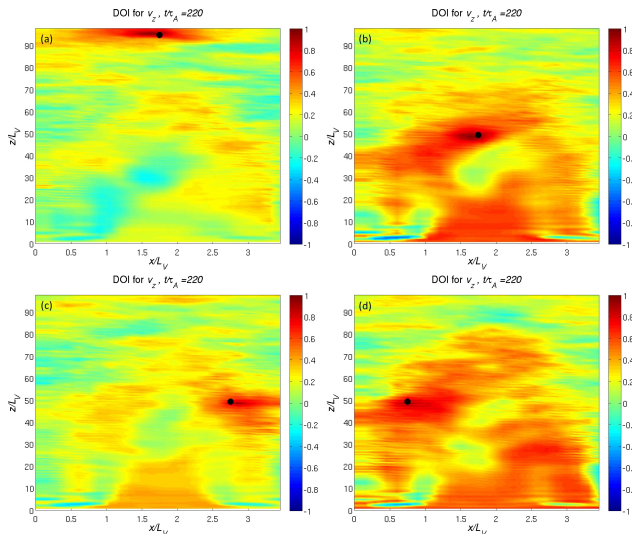
**Fig. 7.**  $z$  component of the magnetic field and, superimposed, the field lines for the reference simulation at different time steps and on the entire simulated domain. The spatial dimensions are normalised to  $L_B$ , the length scale of variation of the magnetic field. Time is normalised to the Alfvén time  $\tau_A$ ,  $b_z$  is normalised to  $b_0$ .



**Fig. 8.** Domain of influence for the  $z$  component of the magnetic field  $b_z$  for the assimilation of a magnetic field observation made at different locations at the same time step,  $t/\tau_A = 220$ . The black circle indicates the data observation location. (a) refers to location 4 (L1, 1 AU), (b) to location 3, (c) to location 5 and (d) to location 6. The spatial dimensions are normalised to  $L_B$ , the length scale of variation of the magnetic field; time is normalised to the Alfvén time  $\tau_A$ .

ferent locations of the simulation domain (see Fig. 1) and at different times are shown.

For example, for panel a, the doi of one virtual observation located at L1 is depicted at  $t/\tau_A \approx 220$ . Both the observation



**Fig. 9.** Domain of influence for the  $z$  component of the solar wind velocity  $v_z$  for the assimilation of a velocity observation taken at different locations at the same time step  $t/\tau_A = 220$ . The black circle indicates the data observation location. (a) refers to location 4 (L1, 1 AU), (b) to location 3, (c) to location 5 and (d) to location 6. The spatial dimensions are normalised to  $L_V = L_B$ , the length scale of variation of the velocity; time is normalised to the Alfvén time  $\tau_A$ .

and the domain of influence are referred to the  $z$  component of the magnetic field.

Notice that the amplitude of the correlation at the observation point location is equal to “1”, which is the maximum positive correlation. Then, the correlation progressively decreases when moving away from the observation point. When the ensemble correlation approaches the maximum negative correlation, value “−1”, degradation may occur. This can actually happen when moving too far away from the observation area. In this case, when implementing DA technique, the analysis step should be restricted to the area with positive correlation. As explained in Sect. 3.2.2, the correlations here reflect the potential spatial extent of the correction around a given location with an isolated observation. Then, the correlation amplitude (Eq. 27) will be modulated, taking into account the observation error (with strong observation errors, the model correction remains low) and the ratio of error variances at the point to be corrected (correction could be significant or lower).

Notice that the domain of influence of the observation is rather localised and decreases to 0 when moving towards the Sun. This result makes sense, since the driver of the simulation is located at the source surface, i.e. far away from the L1 location, and means that observations of the magnetic field taken at L1 can only slightly improve the model evolution closer to the source surface.

It is also important to observe that the occurrence of magnetic reconnection results in the fact that some regions of the

domain remain magnetically isolated. The beneficial feedback from assimilation is thus prevented from spreading into those regions. This effect is particularly severe for the observations taken at L1.

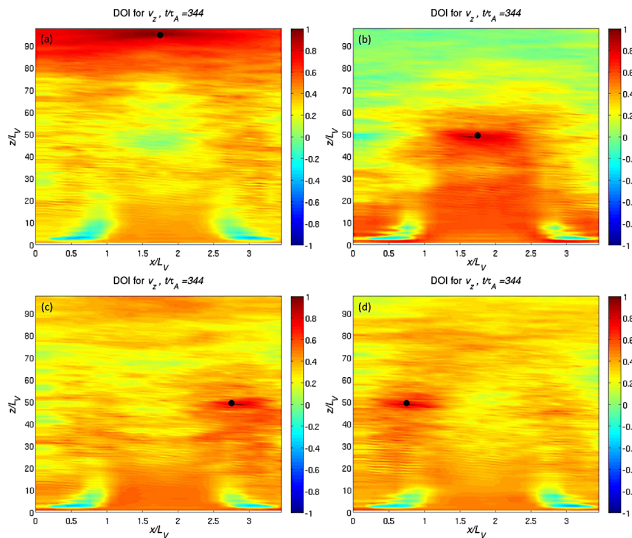
The doi plots from Fig. 8 for location 3 (panel b), location 5 (panel c) and location 6 (panel d) respectively, which again refer to the space and time evolutions of the domain of influence of various magnetic field observations, show more reassuring results: observations taken at those locations would significantly improve the model structure at greater distances, thus making the use of data assimilation a worthy tool for making the model evolution closer to reality. Notice again that the influence of an observation point on the rest of the domain is strictly bound to the occurrence of reconnection, which temporarily magnetically isolates the reconnection area from the rest of the domain. Note that these findings are independent of time and remain similar throughout the whole period.

The same doi analyses are performed for the solar wind velocity  $v_z$ . Figure 9 shows the domain of influence for the solar wind velocity in the case of the assimilation of a velocity observation made at location 4 (panel a), location 3 (panel b), location 5 (panel c) and location 6 (panel d) respectively at time  $t/\tau_A \approx 220$ . Notice that the correlation impact is wider than for magnetic field measures and extends over time to cover the simulation domain entirely, as visible in Fig. 10, which refers to the same spatial locations but to a later time,  $t/\tau_A \approx 344$ . This ensures that data assimilation would significantly minimise the model errors in the velocity field when velocity observations are assimilated. Panel (a) is of particular relevance, since it shows that observations taken at 1 AU, where most of the space weather relevant satellites are located, also have a very strong effect in improving the solar wind velocity representation over almost the entire domain.

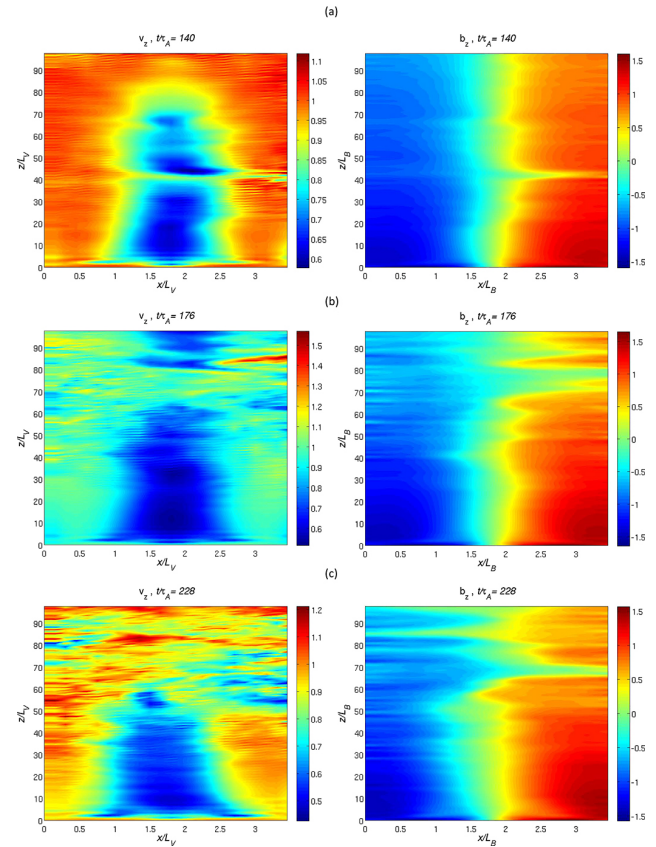
The results thus show that data assimilation should improve the forecast of velocity and magnetic fields not only near and around the observation location, but also for grid points rather far from the measurement location.

It is reminded that, as from Eq. (14), the correction of the forecasted state is a linear combination of representers associated with the observations assimilated. The representers include both contributions from covariances for the same fields (magnetic field or velocity field) and from cross-covariances between different fields (magnetic field with velocity field): each representer propagates information from the observation to the different variables of the state vector. This means, in this case, that assimilation of magnetic field observations might improve the state for the solar wind velocity as well, and vice versa.

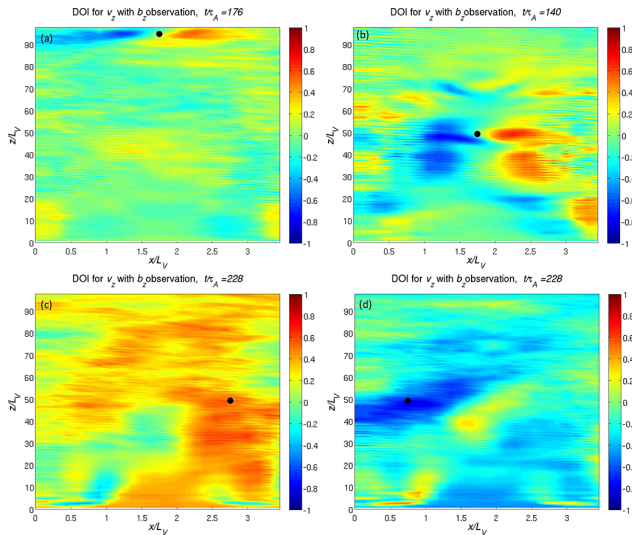
Here, the ensemble cross correlations between the  $z$  components of the magnetic and velocity fields are analysed in the whole domain of study. A virtual  $b_z$  observation is supposed to be assimilated at the same previous location points and the doi on  $v_z$  is computed. Figure 11 shows snapshots of cross correlations for locations 4, 3, 5 and 6, selected



**Fig. 10.** Domain of influence for the  $z$  component of the solar wind velocity  $v_z$  for the assimilation of a velocity observation taken at different locations at the same time step  $t/\tau_A = 334$ . The black circle indicates the data observation location. (a) refers to location 4 (L1, 1 AU), (b) to location 3, (c) to location 5 and (d) to location 6. The spatial dimensions are normalised to  $L_V = L_B$ , the length scale of variation of the velocity; time is normalised to the Alfvén time  $\tau_A$ .



**Fig. 12.**  $z$  components of the solar wind velocity  $v_z$  (left panel) and of the magnetic field  $b_z$  (right panel) of the reference simulation in correspondence with doi plots of Fig. 11, for times (a)  $t/\tau_A = 176$ , (b)  $t/\tau_A = 140$  and (c)  $t/\tau_A = 228$ . The spatial dimensions are normalised to  $L_V = L_B$ , the length scales of variation of the magnetic field and of the velocity; time is normalised to the Alfvén time  $\tau_A$ .  $b_z$  is normalised to  $b_0$  and  $v_z$  is normalised to the Alfvén speed  $v_A$ .



**Fig. 11.** Domain of influence for the  $z$  component of the solar wind velocity  $v_z$  for the assimilation of a magnetic field observation taken at different locations and different times. The black circle indicates the data observation location. (a) refers to location 4 (L1) and time  $t/\tau_A = 176$ , (b) to location 3 and time  $t/\tau_A = 140$ , panel (c) to location 5 and (d) to location 6, both at time  $t/\tau_A = 228$ . The spatial dimensions are normalised to  $L_V = L_B$ , the length scale of variation of the velocity; time is normalised to the Alfvén time  $\tau_A$ .

cases with no correlation are dropped. For both locations 4 and 3 the domain of influence is characterised by the appearance of a dipole around the observation location with a separation of correlated and anti-correlated areas. For location 6 the doi shows an anti-correlated area, while for location 5 the doi is characterised by a positive ensemble correlation. As a reference, Fig. 12 shows the  $z$  components of the solar wind velocity  $v_z$  (left panel) and of the magnetic field  $v_z$  (right panel) of the reference simulation for the same times of the doi plots of Fig. 11. Preliminarily, this figure indicates that the cross correlations are related to the sign of the magnetic and velocity fields, respectively.

It is known in ensemble methods (Evensen, 2009; Boutier and Courtier, 1999) that ensemble correlations are related to the ensemble member structures of model errors. The correlation/anti-correlation space extent seems to be related to the sign of the  $z$  components of the magnetic and velocity fields. While the  $z$  component of velocity is positive for all ensemble members at all locations, as can be noticed from

throughout the simulation at different time steps. It should be noted that only specific cases are shown. The less interesting

Fig. 5 (right panels), the behaviour of  $b_z$  at the different locations is more complicated. From the right panels of Fig. 4 it is possible to notice that at locations 4 and 3, where dipoles are present in Fig. 10, some of the ensemble members exhibit positive values of  $b_z$ , while the rest of the ensemble members experience negative values of the magnetic field. At location 6, where most of the domain shows negative correlation, all the ensemble members have negative magnetic fields. At location 5, which is mostly positively correlated,  $b_z$  is positive for all the ensemble members.

Further studies need to be conducted to better understand this behaviour.

## 5 Conclusions

Investigations were made to assess if and how the use of data assimilation techniques could lead to more accurate estimates of state variables (i.e. solar wind velocity and magnetic field) for solar wind models in different regions of the heliosphere. A location of particular interest is the source surface, since boundary conditions for solar wind models are usually taken there. More in general, the potential domains of influence due to assimilation of virtual measurements taken at different locations of space on the evolution of the FLIP-MHD model for the solar wind were investigated. The general aim of the study was to understand the processes and dynamics in terms of spatial and temporal features of an MHD heliospheric model. A sensitivity study has been conducted through ensemble variances and representers “influence functions” analysis, both performed with multiple runs of the model on statistically guided modifications of the input.

First, the sensitivity study permits us to characterise the sensibility of the model to the variation of the input parameters. This allowed us to better identify the reaction of the model to small variations in the boundary conditions. It is now understood how minimal differences in the boundaries may give rise to a different pattern of reconnection events and that, as a consequence, a faithful representation of the source surface conditions is paramount for the evolution of such models. Hence, the necessity of reliable boundary conditions arises. Another result is the estimate of the error structures of the model, which helps to emphasise the domain areas where dedicated observations may be useful to be collected.

Second, the representer analysis has been used to anticipate and estimate the potential contribution of DA to the model evolution. Indeed, the ensemble modelling led us to investigate the model error covariances with the aim of identifying the potential domains of influence through the assimilation of one virtual observation. Such an analysis was performed by computing the ensemble correlations and cross correlations, both for magnetic field and velocity measurements at different spatial locations. It was shown that, while for magnetic field measurements the model improvement is rather bound to the location of the observation and also to

the occurrence of magnetic reconnection, the assimilation of velocity measures grants model benefits which span further away from the observation location in space and which are less affected by the occurrence of reconnection. Velocity measurements taken at L1 proved to be strongly correlated with the model evolution, also at large distances.

It should be remarked that calculating ensemble correlations also has the advantage of considering the observations one by one, thus isolating the influence of each observation in incrementing the correction. In the intermediate goal of optimising a monitoring network for space weather events, this approach is very interesting and helpful. Indeed, the domain of influence analysis can already provide useful information about the spatial extent of the expected correction obtained by assimilating one isolated observation. This allows one to identify the spatial coverage required for the design of a space weather observation network.

The approach and the results presented here do not assess the success or the failure of the assimilation system. However, the positive results obtained by computing the domains of influence of virtual observation are a further hint of the potential benefits of DA on space weather MHD models. The next step would be to implement and run the complete DA, by applying the correction brought by the observations to the forecasted state  $x^f$  (Eq. 14).

From this study, two main strategies for applying DA to heliospheric models emerge. One consists in updating directly the model fields in the simulation space. If magnetic field values are updated, special attention must be dedicated to the solenoidal condition for the magnetic field. The other one consists in correcting the boundary values at the source surface in order to feed the model continuously with optimal boundary conditions. The study focused on the second strategy, since poor magnetograms are known as a major source of errors in ambient solar wind modelling.

*Acknowledgements.* The authors thank their colleagues for continuing support and discussion. The research leading to these results has received funding from the European Commission’s Seventh Framework Programme (FP7/2007-2013) under grant agreement eHeroes (project no. 284461, [www.eheroes.eu](http://www.eheroes.eu)). Figure 2 has been provided by the Community Coordinated Modeling Center at Goddard Space Flight Center through their public Runs on Request system (<http://ccmc.gsfc.nasa.gov>). The CCMC is a multi-agency partnership between NASA, AFMC, AFOSR, AFRL, AFWA, NOAA, NSF and ONR. The MAS model was developed by J. Linker, Z. Mikic, R. Lionello, P. Riley, N. Arge and D. Odstrcil at PSI, AFRL, U. Colorado.

Edited by: S. Sharma

Reviewed by: V. S. Lukin and two anonymous referees

## References

- Baker, D. N.: What is space weather?, *Adv. Space Res.*, 22, 7–16, 1998.
- Barrero Mendoza, O., De Moor, B., and Bernstein, D. S.: Data assimilation for magnetohydrodynamics systems, *J. Comput. Appl. Mathe.*, 189, 242–259, 2006.
- Beck, A., Innocenti, M. E., Lapenta, G., and Markidis, S.: Multi-level multi-domain algorithm implementation for two-dimensional multiscale particle in cell simulations, doi:10.1016/j.jcp.2013.12.016, 2013.
- Bennet, A. F.: *Inverse Methods in Physical Oceanography, Monographs on Mechanics and Applied Mathematics*, Cambridge University Press, Cambridge, 1992.
- Bothmer, V. and Daglis, I. A.: *Space Weather – Physics and effects*, Springer, Praxis Publishing, Chichester, 2007.
- Bouttier, F. and Courtier, P.: *Data assimilation concepts and methods, Meteorological Training Course Lecture Series*, ECMWF, 1999.
- Brackbill, J. U.: FLIP MHD – A particle-in-cell method for magnetohydrodynamics, *J. Comput. Phys.*, 96, 163–192, 1991.
- Dwivedi, B. N. and Mohan, A.: First tomographic view of coronal mass ejections, *Current Sci.*, 88, 688–690, 2005.
- Echevin, V., De Mey, P., and Evensen, G.: Horizontal and vertical structure of the representer functions for sea surface measurements in a coastal circulation model, *J. Phys. Oceanogr.*, 30, 2627–2635, 2000.
- Egbert, G. D. and Bennett, A. F.: Data assimilation methods for ocean tides, *Modern Approaches to Data Assimilation in Ocean Modeling*, edited by: Malonotte-Rizzoli, P., Elsevier Science, 147–179, 1996.
- Evensen, G.: *Data Assimilation: The Ensemble Kalman Filter*, 2nd Edn., Springer, Berlin, 2009.
- Fleck, B., Domingo, B., and Poland, A. I.: The SOHO Mission, *Solar Phys.*, 162, 1–531, 1995.
- Gantois, K., Santandrea, S., Teston, F., Strauch K., Zender, J., Tilmans, E., and Gerrits, D.: Big year for small satellite – ESA's second in-orbit technology demonstrator mission: PROBA-2, *ESA Bulletin*, 144, 22–33, 2010.
- Ghil, M. and Malanotte-Rizzoli, P.: Data assimilation in meteorology and oceanography, *Adv. Geophys.*, 33, 141–266, 1991.
- Gosling, J. and Pizzo, V.: Formation and evolution of corotating interaction regions and their three dimensional structure, *Space Sci. Rev.*, 89, 21–52, 1999.
- Innocenti, M. E., Lapenta, G., Vrsnak, B., Crespon, F., Skandrani, C., Temmer, M., Veronig, A., Bettarini, L., Markidis, S., and Skender, M.: Improved forecasts of solar wind parameters using the Kalman filter, *Space Weather*, 9, 10005, doi:10.1029/2011SW000659, 2011.
- Innocenti, M. E., Lapenta, G., Markidis, S., Beck, A., and Vapirev, A.: A multi-level multi domain method for particle in cell plasma simulations, *J. Comput. Phys.*, 238, 115–140, doi:10.1016/j.jcp.2012.12.028, 2013.
- Kalman, R.: A new approach to linear filtering and prediction problems, *J. Basic Eng.*, 82, 35–45, 1960.
- Kondrashov, D., Shprits, Y., Ghil, M., and Thorne, R.: A Kalman filter technique to estimate relativistic electron lifetimes in the outer radiation belt, *J. Geophys. Res.*, 112, A10227, doi:10.1029/2007JA012583, 2007.
- Laakso, H., Taylor, M. G. G. T., and Escoubet, C. P.: *The Cluster active archive – Studying the Earth's space plasma environment*, Springer, Astrophysics and Space Science Proceedings, 2010.
- Levine, R. H., Altschuler, M. D., and Harvey, J. W.: Solar Sources of the Interplanetary Magnetic Field and Solar Wind, *J. Geophys. Res.*, 82, 1061–1065, 1977.
- Lyard, F.: Data assimilation in a wave equation: a variational representer approach for the Grenoble tidal model, *J. Comput. Phys.*, 149, 1–31, 1999.
- McComas, D. J., Bame, S. J., Barraclough, B. L., Feldman, W. C., Funsten, H. O., Gosling, J. T., Riley, P., Skoug, R., Balogh, A., Forsyth, R., Goldstein, B. E., and Neugebauer, M.: Ulysses' return to the slow solar wind, *Geophys. Res. Lett.*, 25, 1–4, 1998.
- Odstrcil, D.: Modeling 3-D solar wind structure, *Adv. Space Res.*, 32, 497–506, 2003.
- Pesnell, W. D., Thompson, B. T., and Chamberlin, P. C.: The Solar Dynamics Observatory (SDO), *Solar Phys.*, 275, 3–15, doi:10.1007/978-1-4614-3673-7\_2, 2012.
- Rigler, E., Baker, D., and Weigel, R.: Adaptive linear prediction of radiation belt electrons using the Kalman filter, *Space Weather*, 2, S03003, doi:10.1029/2003SW000036, 2004.
- Riley, P., Linker, J. A., and Mikic, Z.: An empirically-driven global MHD model of the solar corona and inner heliosphere, *J. Geophys. Res.*, 106, 15889–15901, 2001.
- Schunk, R. W., Scherliess, L., and Sojka, J. J.: Recent approaches to modeling ionospheric weather, *Adv. Space Res.*, 31, 819–828, 2003.
- Schunk, R. W., Scherliess, L., Sojka, J. J., Thompson, D. C., Anderson, D. N., Codrescu, M., Minter, C., Fuller-Rowell, T. J., Heelis, R. A., Hairston, M., and Howe, B. M.: Global Assimilation of Ionospheric Measurements (GAIM), *Radio Sci.*, 39, doi:10.1029/2002RS002794, 2004.
- Schrijver, C. J. and Derosa, M. L.: Photospheric and heliospheric magnetic fields, *Solar Phys.*, 212, 165–200, 2003.
- Sheeley, N. R. and Harvey, J. W.: Coronal holes, solar wind streams, and geomagnetic disturbances during 1978 and 1979, *Solar Phys.*, 70, 237–249, 1981.
- Siscoe, G.: The space-weather enterprise: past, present, and future, *J. Atmos. Solar-Terrest. Phys.*, 62, 1223–1232, 2000.
- Stone, E. C., Frandsen, A. M., Mewaldt, R. A., Christian, E. R., Margolies, D., Ormes, J. F., and Snow, F.: The Advanced Composition Explorer, *Space Sci. Rev.*, 86, 1–22, doi:10.1007/978-94-011-4762-0\_1, 1998.
- Toth, G., Sokolov, I. V., Gombosi, T. I., Chesney, D. R., Clauer, C. R., De Zeeuw, D. L., Hansen, K. C., Kane, K. J., Manchester, W. B., Oehmke, R. C., Powell, K. G., Ridley, A. J., Roussev, I. I., Stout, Q. F., Volberg, O., Wolf, R. A., Sazykin, S., Chan, A., Yu, B., and Kta, J.: Space weather modeling framework: A new tool for the space science community, *J. Geophys. Res.*, 110, A12226, doi:10.1029/2005JA011126, 2005.
- Turner, D. L. and Li, X.: Using spacecraft measurements ahead of Earth in the Parker spiral to improve terrestrial space weather forecasts, *Space Weather*, 9, S01002, doi:10.1029/2010SW000627, 2011.
- Vršnak, B. and Žic, T.: Transit times of interplanetary coronal mass ejections and the solar wind speed, *Astro. Astrophys.*, 472, 937–943, 2007.

- Vršnak, B., Temmer, M., and Veronig, A.: Coronal Holes and Solar Wind High-Speed Streams: I. Forecasting the Solar Wind Parameters, *Solar Phys.*, 240, 315–330, 2007.
- Wang, Y. and Sheeley Jr., N. R.: Solar wind speed and coronal flux-tube expansion, *Astrophys. J.*, 355, 726–732, 1990.
- Welch, G. and Bishop, G.: An introduction to the Kalman filter, TR 95-041, University of North Carolina, Department of Computer Science, 2001.
- Wu, S. T., Wang, A. H., Liu, Y., and Hoeksema, J. T.: Data Driven Magnetohydrodynamic model for active region evolution, *Astrophys. J.*, 652, 800–811, 2006.