

Domain of Influence analysis: implications for Data Assimilation in space weather forecasting

Dimitrios Millas^{1,4,*}, Maria Elena Innocenti^{1,*}, Brecht Laperre¹, Joachim Raeder², Stefaan Poedts^{1,3} and Giovanni Lapenta¹

¹ Centre for mathematical Plasma Astrophysics, Department of Mathematics, KU Leuven, Leuven, Belgium

² Institute for the Study of Earth, Oceans and Space, University of New Hampshire, Durham, NH, USA

³ Institute of Physics, University of Maria Curie-Skłodowska, Lublin, Poland

⁴ Department of Physics and Astronomy, University College London, London, UK

Correspondence*:

Dimitrios Millas

dimitrios.millas@kuleuven.be

Maria Elena Innocenti

mariaelena.innocenti@kuleuven.be

ABSTRACT

Solar activity, ranging from the background solar wind to energetic coronal mass ejections (CMEs), is the main driver of the conditions in the interplanetary space and in the terrestrial space environment, known as space weather. A better understanding of the Sun-Earth connection carries enormous potential to mitigate negative space weather effects with economic and social benefits. Effective space weather forecasting relies on data and models. In this paper, we discuss some of the most used space weather models, and propose suitable locations for data gathering with space weather purposes. We report on the application of *Representer analysis (RA)* and *Domain of Influence (DOI) analysis* to three models simulating different stages of the Sun-Earth connection: the OpenGGCM and Tsyganenko models, focusing on solar wind - magnetosphere interaction, and the PLUTO model, used to simulate CME propagation in interplanetary space. Our analysis is promising for space weather purposes for several reasons. First, we obtain quantitative information about the most useful locations of observation points, such as solar wind monitors. For example, we find that the absolute values of the DOI are extremely low in the magnetospheric plasma sheet. Since knowledge of that particular sub-system is crucial for space weather, enhanced monitoring of the region would be most beneficial. Second, we are able to better characterize the models. Although the current analysis focuses on spatial rather than temporal correlations, we find that time-independent models are less useful for Data Assimilation activities than time-dependent models. Third, we take the first steps towards the ambitious goal of identifying the most relevant heliospheric parameters for modelling CME propagation in the heliosphere, their arrival time, and their geoeffectiveness at Earth.

Keywords: solar wind, coronal mass ejections (CMEs), magnetohydrodynamics (MHD), numerical simulations, statistical tools, domain of influence, observations

1 INTRODUCTION

Solar activity affects the terrestrial environment with a constantly present but highly variable solar wind and with higher energy, transient events, such as flares and Coronal Mass Ejections (CMEs).

“Space weather” (Bothmer and Daglis, 2007) is the discipline that focuses on the impact of these solar drives on the solar system and in particular on the Earth and its near space environment.

Space weather events can have serious effects on the health of astronauts and on technology, with potentially large economic costs (Eastwood et al., 2017). The importance of space weather forecasting has grown with societal dependence on advanced space technology, on communication and on the electrical grid. For example, the Halloween 2003 solar storms that impacted Earth between 19th of October 2003 and 7th of November 2003 caused an hour long power outage in Sweden (Pulkkinen et al., 2005), forced airline flight reroutes, and affected communication and satellite systems (Plunkett, 2005). The “great geomagnetic storm” of March 13-14, 1989 caused, among other disruptions, a blackout of up to nine hours in most of Quebec Province, due to a massive failure experienced by the power grid Hydro-Quebec Power Company (Allen et al., 1989).

In order to improve space weather forecasting, accurate models of the Sun-Earth connection are needed. Such forecasts are challenging because of the complexity of the processes involved and the large range of spatial and temporal scales. Commonly the heliosphere is divided into sub-systems, where each one is simulated with a different model, such that the models feed into each other (Luhmann et al., 2004; Tóth et al., 2005). These models can be either physics-based or empirical. Empirical models (such as, in the solar domain, Altschuler and Newkirk (1969); Schatten et al. (1969); Schatten (1971); Wang and Sheeley Jr (1992); Nikolić (2019)) usually require less computational resources, enabling faster forecasts. They can also serve as a baseline for physics-based models (Siscoe et al., 2004). However, empirical models lack the sophistication of more expensive first-principles based numerical models. Recently, machine learning methods have emerged, that can provide a new approach to space weather forecasting (Camporeale, 2019; Laperre et al., 2020). Most of these methods, while promising, must still undergo extensive validation.

The technique of Data Assimilation (DA) was developed to improve model predictions by properly initializing models from data and by keeping a model on track during its time evolution. (Kalnay, 2003; Bouttier and Courtier, 1999; Evensen, 2009). DA methods were originally applied to atmosphere and ocean models, which exhibit a large degree of inertia. The latter is also true for the solar wind, but not for the magnetosphere-ionosphere system, which is strongly driven and dissipative. Therefore, in space weather forecasting, DA aims not only at initializing the models, but also at using information from various observations to bring the evolution of a system as predicted from a model closer to the real system evolution (Kalman, 1960; Bouttier and Courtier, 1999; Bishop et al., 2001; Evensen, 2009; Le Dimet and Talagrand, 1986; Innocenti et al., 2011), making up for model deficiencies in the terms of resolution and incomplete physical description.

The quantity and quality of available data is a critical factor in the effectiveness of Data Assimilation. This is the reason why fields where data can be obtained more easily and continuously have shown early successes in DA implementations. Examples of these fields are meteorology and oceanography, and, in space sciences, ionospheric and radiation belt physics (Bennett, 1992; Ghil and Malanotte-Rizzoli, 1991; Egbert and Bennett, 1996; Kalnay, 2003; Rigler et al., 2004; Schunk et al., 2004; Kondrashov et al., 2007). Examples of DA applications targeting specifically the interplanetary space environments are Schrijver and DeRosa (2003); Mendoza et al. (2006); Arge et al. (2010); Innocenti et al. (2011); Skandran et al. (2014); Lang and Owens (2019); Lang and Owens (2019); Lang et al. (2020).

Representer analysis (RA) and Domain of Influence analysis (DOI) (Bennett, 1992; Egbert and Bennett, 1996; Echevin et al., 2000; Evensen, 2009; Skandran et al., 2014), briefly summarized in Section 2, are powerful statistical tools used to estimate the effectiveness of DA techniques when applied to a specific model, without assimilating actual data. Such analysis can be used in several ways. It allows us to optimize assimilation strategies, it may uncover model biases that can then be addressed by further model development, and it may be used to optimize the observation systems that provide operational data for DA. For example, RA/DOI can be used to optimize locations for solar wind monitors, such as locations proposed near the L5 Lagrangian point (Vourlidas, 2015; Lavraud et al., 2016; Pevtsov et al., 2020).

In the present paper, the RA and DOI analysis is applied to three models: the OpenGGCM magnetosphere - ionosphere model (Section 3.1), two of the empirical Tsyganenko magnetosphere magnetic field models (Section 3.2), and a solar wind simulation based on the PLUTO code (Section 3.3). These models simulate critical sub-systems in the Sun-Earth connection with a focus on the terrestrial magnetosphere and Coronal Mass Ejection propagation.

The present paper provides insights into the locations of the terrestrial magnetosphere that should be prioritized (ideally, in absence of orbital constraints) for space weather forecasting and monitoring activity. We compare a time-dependent, physics based model (e.g., OpenGGCM) and time-independent, empirical (e.g., Tsyganenko) models in terms of the expected benefits that DA can provide. We conclude that time-dependent models should be preferentially chosen for DA. We take the first steps towards the goal if understanding the main physical parameters, close to the Sun and in interplanetary space, that control CME propagation and hence their arrival time at Earth.

This manuscript is organized as follows: in Section 2 we introduce the theoretical background on RA and the DOI; Section 3 discusses the application of the method to the different models; in Section 4 we summarize the results and discuss potential improvements and new applications.

2 REPRESENTER ANALYSIS AND DOMAIN OF INFLUENCE ANALYSIS

This Section introduces the mathematical basis of RA and DOI analysis. The reader is referred to Skandran et al. (2014) and references therein for an in depth derivation.

Let us start from a system described by the state variable vector $\mathbf{x}_t \in \mathbb{R}^n$. \mathbf{x}_t is a vector containing the n state variables that describe the system at a time t . Let us assume that the evolution of the system can be described as a discrete-time process controlled by an evolution law \mathbf{A} . The state of the system then evolves as follows: $\mathbf{x}_t = \mathbf{A}(\mathbf{x}_{t-1}) + \mathbf{w}_{t-1}$, where $\mathbf{w} \in \mathbb{R}^n$ is process noise. The process noise is assumed to be Gaussian and with covariance matrix \mathbf{Q} .

If a model \mathbf{M} (for example, a simulation model) of the evolution law \mathbf{A} is available, we can obtain, following Kalman (1960); Evensen (2009), a prior estimate $\hat{\mathbf{x}}_t^-$ of the state variable \mathbf{x}_t through the simulation model as

$$\hat{\mathbf{x}}_t^- = \mathbf{M}(\hat{\mathbf{x}}_{t-1}^-) + \mathbf{w}_{t-1}. \quad (1)$$

Assume now that we have m observational values or measurements $\mathbf{z}_t \in \mathbb{R}^m$. These measurements can be mapped to the current state \mathbf{x}_t through the so-called observation operator $\mathbf{H} \in \mathbb{R}^{m \times n}$, such that $\mathbf{z}_t = \mathbf{H}\mathbf{x}_t + \nu_t$. Here, ν is the (assumed Gaussian) measurement noise, with a covariance matrix \mathbf{R} .

It can be then shown (Bishop et al., 2001) that a posterior estimate of the state ($\hat{\mathbf{x}}_t$) can be obtained from the prior estimate of the state ($\hat{\mathbf{x}}_t^-$), obtained from Eq. 1, as follows:

$$\hat{\mathbf{x}}_t = \hat{\mathbf{x}}_t^- + \mathbf{K}_t (\mathbf{z}_t - \mathbf{H}\hat{\mathbf{x}}_t^-). \quad (2)$$

Here, the term $(\mathbf{z}_t - \mathbf{H}\hat{\mathbf{x}}_t^-)$ is called the “innovation”, and represents the difference between the measurements and their expected values, calculated by applying the observation operator to the prior state estimate. The Kalman gain \mathbf{K}_t is obtained by minimizing the posterior error covariance matrix. This is the “correction” step of the Kalman filter, where the Kalman gain is calculated and the estimate and error covariance matrix of the posterior state are updated. The “prediction” (forecast) phase of the filter results in the calculation of the prior state estimate and prior error covariance matrix (used to compute the Kalman gain).

The prior and posterior error covariance matrices are respectively defined as

$$\mathbf{P}_t^- = \mathbb{E} \left[(\mathbf{x}_t - \hat{\mathbf{x}}_t^-) (\mathbf{x}_t - \hat{\mathbf{x}}_t^-)^T \right], \quad \mathbf{P}_t = \mathbb{E} \left[(\mathbf{x}_t - \hat{\mathbf{x}}_t) (\mathbf{x}_t - \hat{\mathbf{x}}_t)^T \right], \quad (3)$$

where \mathbb{E} is the expected value, \mathbf{x}_t is the “real”, unknown system state and $\epsilon^- = (\hat{\mathbf{x}}_t^- - \mathbf{x}_t)$ and $\epsilon = (\hat{\mathbf{x}}_t - \mathbf{x}_t)$ are the prior and posterior errors, calculated as the difference between the prior ($\hat{\mathbf{x}}_t^-$)/posterior ($\hat{\mathbf{x}}_t$) state and the real state, \mathbf{x}_t . Notice that, although these are the definitions of the error covariances, this is not how they are computed in the filter, since the real state is not known.

The formula for the calculation of the posterior state, Eq. 2, can be written as

$$\hat{\mathbf{x}}_t = \hat{\mathbf{x}}_t^- + \mathbf{r}\mathbf{b} \quad (4)$$

where $\mathbf{r} \in \mathbb{R}^{n \times m}$ and $\mathbf{b} \in \mathbb{R}^m$ are defined as

$$\mathbf{r} = \mathbf{P}^- \mathbf{H}^T, \quad \mathbf{b} = \left(\mathbf{H} \mathbf{P}^- \mathbf{H}^T + \mathbf{R} \right)^{-1} (\mathbf{z} - \mathbf{H}\hat{\mathbf{x}}^-), \quad (5)$$

with \mathbf{R} the measurement noise covariance matrix. The time index t has been dropped for ease of reading.

We will from now on assume that the system (and in particular, the observation operator \mathbf{H}) is linear. Then, each column of the matrix \mathbf{r} , denoted as \mathbf{r}_j with $j = 1, \dots, m$, is the representer associated to a given observation z_j (remember that \mathbf{z} is the vector with m observations), and gives a measure of the impact of that observation in “correcting” the prior state estimate. If we further assume that each observation j is located at grid point k_j , and is associated to a state variable, then each column \mathbf{r}_j (now \mathbf{r}_{k_j} , given the assumption mentioned above) contains the covariances (“cov”) between the prior errors at the observation point k_j and at every other grid node, for all the state variables.

Since the real state is not available for error covariance calculations, an ensemble of simulations can be used to estimate the prior errors instead. An ensemble (Evensen, 2009) can be generated by perturbing one or several of the sources of model errors. In this work, ensembles are generated for each model by perturbing the respective initial / boundary conditions. Once the ensemble is available, the covariances of the prior errors at a certain simulated time can be approximated as the ensemble covariances (“cov^{ens}”) of the prior errors. These in turn become the ensemble covariances of the simulated state variables, if one assumes that the prior errors are unbiased. The ensemble covariance between the state variable x and y is

defined as

$$\text{cov}^{ens}(x, y) = \frac{1}{N} \sum_{s=1}^N \left[\left(x_s^- - \frac{1}{N} \sum_{s=1}^N x_s^- \right) \left(y_s^- - \frac{1}{N} \sum_{s=1}^N y_s^- \right) \right] \quad (6)$$

where N is the number of members in the ensemble, and x and y represent two state variables (notice that, for ease of reading, we indicate here two state variables as x and y , while earlier we indicated as vector \mathbf{x} all the state variables).

In a set of simulations, the prior state variables are the simulation results at a specific time. Being able to calculate the representers associated to observations from the ensemble rather than from assimilating observations simplifies the RA significantly.

Following the discussion of the representer term \mathbf{r} in the calculation of the posterior state, Eq. 4 and Eq. 5, we now examine the term \mathbf{b} . Assuming that the only observation point for observation j is at grid node k_j (i.e., the row j of the matrix \mathbf{H} has only the term k_j different from zero), the element of \mathbf{b} associated to the observation at grid point k_j , denoted as b_{k_j} , becomes, from Eq. 5:

$$b_{k_j} = \frac{z_{k_j} - \hat{x}_{k_j}^-}{\text{cov}(\epsilon_{k_j}^-, \epsilon_{k_j}^-) + \text{cov}(\epsilon_{k_j}^z, \epsilon_{k_j}^z)}, \quad (7)$$

where we have made use of the simplified form of the matrix \mathbf{H} and where $\epsilon_{k_j}^z$ is the observation error associated with the observation z_{k_j} .

If x_i is one of the state variables at grid node i , the correction to x_i brought by the assimilation of the measure z_{k_j} , following Eq. 4, Eq. 5, Eq. 7 and some straightforward manipulation based on the definitions of covariance, variance, correlation, can be written as (Skandiani et al., 2014)

$$\hat{x}_i - \hat{x}_i^- = \text{corr}^{ens}(\hat{x}_{k_j}^-, \hat{x}_i^-) F(z_{k_j}). \quad (8)$$

Here, $F(z_{k_j})$ is the modulation factor and $\text{corr}^{ens}(\hat{x}_{k_j}^-, \hat{x}_i^-)$ the correlation. The correlation is computed from the ensemble, and is calculated between the state variable at node k_j and at node i . This correlation reflects how a change at node k_j , caused e.g. by the assimilation of the measurement z_{k_j} , will influence the node i , and is what we call the DOI. The correlation over the ensemble is defined, using the dummy variables x and y for brevity, as:

$$\text{corr}^{ens}(x, y) = \frac{\text{cov}^{ens}(x, y)}{\sqrt{\text{var}(x) \text{var}(y)}}. \quad (9)$$

The modulation factor $F(z_{k_j})$ in Eq. 8 depends, among other things, on the measurement z_{k_j} and on the error associated to the measure z_{k_j} . Hence Data Assimilation has to be performed to calculate this term. The $\text{corr}^{ens}(\hat{x}_{k_j}^-, \hat{x}_i^-)$ term reflects how large we can expect the area that will be affected by the assimilation of z_{k_j} to be. But to know how large the difference between the posterior and prior state, $\hat{x}_i - \hat{x}_i^-$, will be, we need to know the modulation factor as well.

So now the DOI of the measurement z_{k_j} on the state variable at grid point i , x_i , can be defined as

$$DOI(z_{k_j}, x_i) = \text{corr}^{ens} \left(\hat{x}_{k_j}^-, \hat{x}_i^- \right). \quad (10)$$

One can see from its definition that the DOI can be calculated *before assimilation* by computing the ensemble correlation of the state variable value at the grid point k_j with that at grid point i . Dropping the i index, i.e. examining the expected impact of measurement z_{k_j} on all the state variables \mathbf{x} at all grid points, we obtain the more general definition of the DOI as

$$DOI(z_{k_j}) = \text{corr}^{ens} \left(\hat{x}_{k_j}^-, \hat{\mathbf{x}}^- \right). \quad (11)$$

We can then draw “DOI maps” that show the correlation between a field at grid point k_j , the “observation point”, and the other grid points.

Notice that in this derivation we have assumed, for simplicity, that the measurement z and the state variable x refer to the same field, for example, the \mathbf{x} component of the magnetic field, or of the velocity. This simplifies the formulation of the numerator of Eq. 7 and improves the readability of the derivation. Skandran et al. (2014) shows examples where the DOI is calculated between different fields, e.g. magnetic field and velocity.

DOI analysis has the advantage that it can be calculated for all state variables and at any grid point without actual assimilation, i.e. without the need for measurements \mathbf{z} . In order to compute the DOI at a time step t , we only require evolving the ensemble up to said time step t , and then performing the correlation over the ensemble between the state variable value at the observation point k_j and at all other grid points.

Because DOI values are derived from a correlation they are bounded between -1 and 1. $|DOI| \sim 1$ indicates that the field at that specific point significantly changes when the same field (or a different field, in the case of cross-correlation) varies at the observation point. $|DOI| \sim 0$ indicates the opposite, i.e., variation at the observation point have little or no effect. Thus, DOI analysis also provides information on how information propagates within the model, and therefore sheds light on the physical processes within the model. We will exploit this property in Section 3.2.

We note that in this study we only focus on spatial correlations, neglecting temporal correlations. In other words, the following analysis (i.e. the calculation of variances, DOI, etc.) is restricted to specific instances in time, rather than examining correlations between fields at difference times as well. The dependence on time will be addressed in a future project.

The RA is applied to “artificial” data, obtained from ensembles of simulations focused on different processes of interest in the Sun-Earth connection: the interaction of the solar wind with the terrestrial magnetosphere (via OpenGGCM and Tsyganenko simulations) and the propagation of a CME-like event in the steady solar wind (PLUTO).

OpenGGCM and the Tsyganenko Geomagnetic Field Models both simulate the interaction of the solar wind with the magnetospheric system. OpenGGCM is a physics-based magnetohydrodynamic (MHD) model, while the Tsyganenko models are semi-empirical best-fit representations for the magnetic field, based on a large number of satellite observations (Tsyganenko, 1995, 2002a; Tsyganenko and Sitnov, 2005).

PLUTO is an MHD-modelling software used to simulate the propagation of a CME in the background solar wind. This software can be used to numerically solve the partial differential equations encountered in plasma physics problems, in conservative form, in different regimes (from hydrodynamics to relativistic MHD). The structure of the software is explained in Mignone et al. (2007, 2012). Full documentation and references can be found in the relevant web page: <http://plutocode.ph.unito.it/>.

Because the DOI analysis is an ensemble based technique the size of the ensemble and its properties matter. In order to test for sufficient size, we performed the DOI calculation using a limited, random subset of the ensemble, which we gradually expanded. We found that using at 25 runs were sufficient to obtain a consistent ensemble mean, variance, and DOI. We note, however, that this may change for different choices of simulation resolution and parameters used for the generation of the ensemble.

3 APPLICATIONS

3.1 Magnetospheric applications I: OpenGGCM

The OpenGGCM (Open Geospace General Circulation Model) is a MHD based model that simulates the interaction of the solar wind with the magnetosphere-ionosphere-thermosphere system. OpenGGCM is available at the Community Coordinated Modeling Center at NASA/GSFC for model runs on demand (see: <http://ccmc.gsfc.nasa.gov>). This model has been developed and continually improved over more than two decades. Besides numerically solving the MHD equations with high spatial resolution in a large volume containing the magnetosphere, the model also includes ionospheric processes and their electrodynamic coupling with the magnetosphere.

The mathematical formulation of the software is described in Raeder (2003). The latest version of OpenGGCM, used here, is coupled with the Rice Convection Model (RCM), (Toffoletto et al., 2001), which treats the inner magnetosphere drift physics better than MHD and allows for more realistic simulations that involve the ring current (Cramer et al., 2017). The model is both modular and efficiently parallelized using the message passing interface (MPI). It is written in Fortran and C, and uses extensive Perl scripting for pre-processing. The software runs on virtually any massively parallel supercomputer available today.

OpenGGCM uses a stretched Cartesian grid (Raeder, 2003) that is quite flexible. There is a minimal useful resolution, about 150x100x100 cells, that yields the main magnetosphere features but does not resolve mesoscale structures such as FTEs or small plasmoids in the tail plasma sheet. At the other end, we have run simulations with grids as large as $\sim 1000^3$ (on some 20,000 cores). In terms of computational cost that is almost a 10^4 ratio. Here, we used a grid of 325x150x150 cells which is sufficient for the purposes of this study and runs faster than real time on a modest number of cores.

OpenGGCM has been used for numerous studies of magnetospheric phenomena such as storms (Raeder et al., 2001b; Raeder and Lu, 2005; Connor et al., 2016), substorms (Raeder et al., 2001a; Ge et al., 2011; Raeder et al., 2010), magnetic reconnection (Dorelli, 2004; Raeder, 2006; Berchem et al., 1995), field-aligned currents (Moretto et al., 2006; Vernerstrom et al., 2005; Raeder et al., 2017; Anderson et al., 2017), and magnetotail processes (Zhu et al., 2009; Zhou et al., 2012; Shi et al., 2014), to name a few.

The boundary conditions require the specification of the three components of the solar wind velocity and magnetic field, the plasma pressure and the plasma number density at 1 AU, which are obtained from ACE observations (Stone et al., 1998) and applied for the entire duration of the simulation at the sunward boundary.

We generate an ensemble of 50 OpenGGCM simulations by perturbing the v_x component of the input solar wind velocity. Changing this particular parameter guarantees a direct and easy way to interpret magnetospheric response.

First, we run a reference simulation using the observed solar wind values at 1 AU starting from May 8th, 2004, 09:00 UTC (denoted as t_0) until 13:00 UTC on May 8th, 2004. We choose this period because it is relatively quiet: no iCME were registered in the Richardson/ Cane list of near-Earth interplanetary CMEs (Richardson and Cane, 2010) and, as it is common during the declining phase of the solar cycle, geomagnetic activity is driven by Corotating Interaction Regions and High Speed Streams (Tsurutani et al., 2006). The study of outlier events, such as CME arrival at Earth, is left as future work.

To generate the ensemble, the solar wind compression is changed in each of the “perturbed” simulations. The perturbed velocity in the x (Earth-to-Sun) direction is set, for the entire duration of each simulation, to a *constant* value obtained by multiplying the average observed v_x^{avg} by a random number S sampled from a normal distribution with mean $\mu = 1$ and standard deviation $\sigma = 0.1$:

$$v_x = Sv_x^{avg}, \text{ with } S \in \mathcal{N}(1, (0.1)^2). \quad (12)$$

The time period we use to calculate the average is the duration of the reference simulation, between 9:00 UTC and 13:00 UTC on May 8th, 2004.

Our choices for the generation of the ensemble are determined by the necessity to preserve both the Gaussian characteristic of the model error and the physical significance of the simulations: the solar wind compression in all ensemble members is not too far from the reference value. With such low standard deviation, the average of the obtained perturbed value plus/minus several sigmas are still within the typical range for the solar wind: the minimum and maximum values of the constant, perturbed input velocities are $|v_x| \sim 363$ km/s and $|v_x| \sim 583$ km/s respectively. The solar wind velocity v_x is negative in the Geocentric Solar Ecliptic (GSE) coordinates used here. The v_x values that we obtain in this way are not supposed to be representative of the full range of values that v_x can assume; they are used to generate an ensemble of simulations “slightly perturbed” with respect to our reference simulation. We note that the real distribution of solar wind velocity is far from a normal distribution, with two distinct peaks and extreme outliers, and would not be appropriate to produce the required ensemble. We refer the reader to Fortin et al. (2014) for optimal procedures on how to choose the variance of the ensemble.

The ensemble analysis requires running 50 simulations to produce the ensemble members, plus the unperturbed reference simulation. Each run takes ~ 12 hours on 52 cores on the supercomputer Marconi-Broadwell (Cineca, Italy), for a total cost of ~ 32000 core hours.

We verify that the prior errors are unbiased (as assumed in the derivation of the method summarized in Section 2) by comparing the reference simulation and the average of the ensemble. We note that the ensemble mean is an appropriate metric to use in this case because the perturbed simulations have not been generated in order to represent all possible solar wind velocity values, but small perturbations around the reference case.

Figure 1 shows this comparison in the xz plane for the x component of the velocity and of the magnetic field at a fixed time, (172 minutes after the beginning of the simulation), for both the reference simulation (panel (a) and (b)) and the average of the ensemble (panel (c) and (d)). The magnetic field lines, depicted in black, are calculated from the reference simulation in panel (a) and (b) and from the ensemble average in panel (c) and (d). The distances are normalized by the Earth radius R_E .

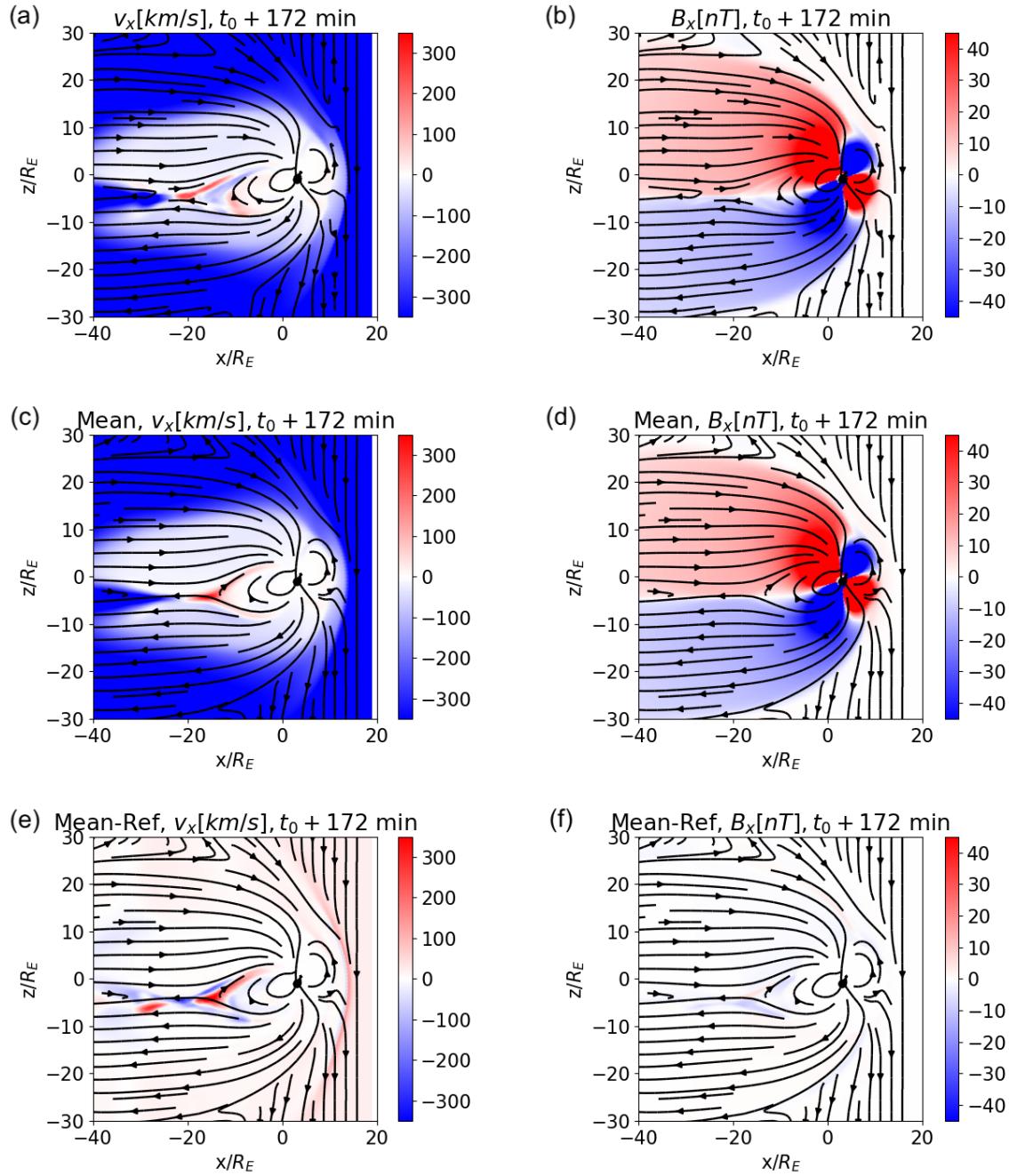


Figure 1. Reference simulation (panel (a) and (b)), ensemble mean (panel (c) and (d)) and difference between the ensemble mean and the reference simulation results (panel (e) and (f)) for the ensemble of OpenGGCM magnetosospheric simulations. The ensemble is generated by perturbing the v_x solar wind boundary condition. v_x is depicted in panel (a), (c), (e), b_x in (b), (d), (f). The coordinate system is GSE. The depicted time is 172 minutes after the beginning of the simulation, May 8th 2004, 9:00 UTC. The boundary conditions at the sunward boundary of the “reference” simulation are observed solar wind values.

Visual inspection of panel (a), (b), (c), (d) and of the difference between the ensemble mean and the reference simulation results, depicted in panels (e) and (f) for v_x and B_x respectively, highlight the areas where the behavior differs most within the ensemble: the bow shock, the plasma sheet, and the neutral line. The former is a plasma discontinuity that moves back and forth in response to the changing solar wind

Mach number, and thus gets smeared out in the ensemble. The latter is a region of marginal stability in the magnetosphere that reacts in a non-linear way to solar wind changes.

In order to determine if 50 ensemble members are sufficient for our analysis, we have compared corresponding plots of the difference between the ensemble mean and the reference simulation for v_x with decreasing number (40, 30, 20) of ensemble members. We find that, with decreasingly smaller ensembles, the plasma sheet structure seen in Fig. 1, panel (e), is only minimally affected. However, the differences around the bow shock become more pronounced. The velocity difference increases in the solar wind and magnetosheath as well, and the magnetic field structure at the magnetosphere/ solar wind interface (as shown by the magnetic field lines, which are drawn for the average field in panel (e) and (f) and similar analysis) begin to change significantly with respect to the reference simulation. By comparing the plots with 50, 40, 30 and 20 ensemble members, we conclude that 30 is the minimum number of ensemble members that gives average fields compatible with the reference simulation, with our choice of perturbation to generate the ensemble.

Panels (a) and (b) of Figure 1 show characteristic signatures of magnetic reconnection in the magnetotail, i.e, the X pattern and the formation of dipolarization fronts in the magnetic field lines, and the presence of earthward and tailward jets in v_x departing from the X point. We provide a movie showing the dynamic evolution in the supplementary material (ReferenceSimVx.avi). The movie shows the solar wind b_z time variation and the subsequent occurrence of several magnetopause/ magnetotail reconnection events. The “formation” of the magnetosphere occurs during the first ~ 30 minutes of the simulation and should be disregarded.

The movie MeanVx.avi, also in the supplementary material, shows how the global evolution changes in the ensemble mean: the magnetopause and magnetotail reconnection patterns are still overall visible, but smoothed out by the averaging procedure with respect to the reference simulation, since the different ensemble instances reconnect at different times and the smaller scale features of each single run are averaged away.

In Figure 2 and movies LowerVx.avi and HigherVx.avi we compare the evolution of the members of the ensemble generated with the lower ($|v_x| \sim 363$ km/s, panel (a)) and higher ($|v_x| \sim 583$ km/s, panel (b)) absolute value of the v_x velocity component. The movies show that the velocity values and magnetic field line patterns are significantly different from the reference simulation and from the ensemble average, demonstrating that the perturbations are not trivial.

We now discuss the RA and DOI analysis for a set of different observation points, depicted as white stars in the following figures, in the inflowing solar wind (a), in the magnetosheath (b), in the northern lobe (c), and in the plasma sheet (d), for the same plane and time as Figure 1. The coordinates of each of these points in the xz -plane are given in Table 1, the y coordinate being $y/R_E = 0$. Figure 3 and Figure 4 show the DOI maps for v_x and b_x respectively. Note that the correlations which are displayed are not cross-correlations: the correlation is done between the value of a field at the observation point and the values of the same field at the other grid points.

Figures 3 and 4 show that the correlations are mostly ordered by the principal regions of the magnetosphere such as the lobes, the magnetosheath, and the plasma sheet. For example, Figure 3 shows the results for the v_x correlations. The DOI values for the plasma sheet are different from those in the magnetosheath and the lobes in all panels. As expected, the stronger correlations are somewhat localized around the observation point, for example, the strongest correlations in panel (d), where the observation point is in the plasma sheet, are in the plasma sheet itself and its immediate surroundings. However, some other observation

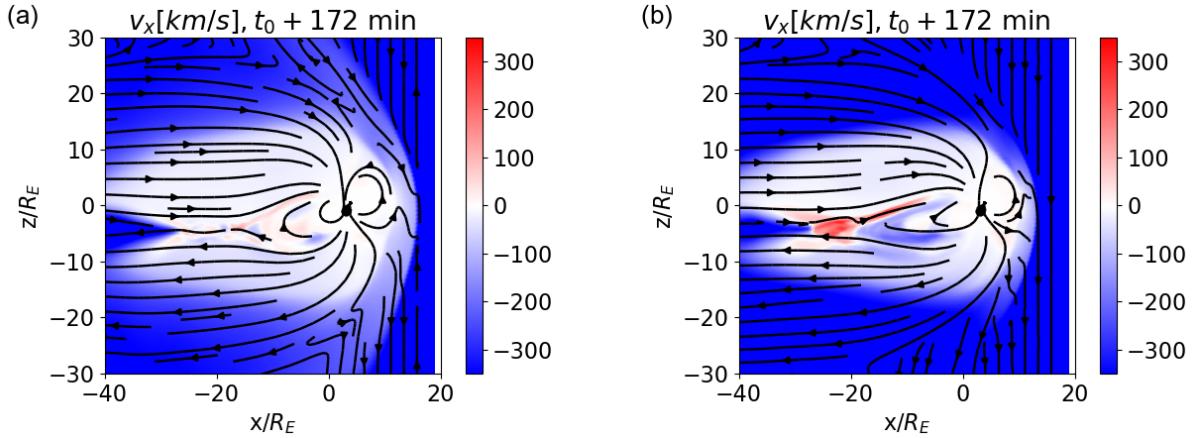


Figure 2. v_x field component, at the same time as Figure 1, for the ensemble member with the lowest (panel (a)) and highest (panel (b)) absolute value of the perturbed, inflowing v_x velocity component in the OpenGGCM ensemble.

	x/R_E	y/R_E	z/R_E
solar wind	15	0	20
magnetosheath	-10	0	20
northern lobe	-10	0	5
plasma sheet	-20	0	-3

Table 1. Coordinates of the observation points used in the DOI analysis.

points have a much larger DOI, such as the ones in the solar wind and the magnetosheath. This makes physical sense, because v_x variations in those regions will propagate through much of the magnetosphere. Figure 4 shows the B_x correlations. The northern and southern lobes clearly stand out, with opposite DOI values, and the magnetosheath stands out as well. Because the B_x values have opposite signs in the two lobes, the DOI values also have opposite signs. The correlation values depend on the variability of the field value at the observation point, thus the panels that exhibit lower correlations are those with observation point in the plasma sheet, where the intrinsic variance of both v_x and B_x is higher (see Figure 1) due to the different reconnection patterns in the different members of the ensemble. For example, in Figure 3, panel (d) the velocity value at the observation point in the plasma sheet exhibits little correlation with the v_x values outside of the plasma sheet and the neighboring areas. This is a consequence of the jet structure which is caused by internal magnetospheric dynamics rather than the solar wind driver.

The temporal dynamic DOI behavior is similar: the DOI maps of v_x and B_x with observation point in the plasma sheet exhibit higher temporal variability than those with a observation point in the magnetosheath, as can be seen in the DOI movies DOI_bx_bx_MSsheath.avi, DOI_bx_bx_pSheet.avi, DOI_vx_vx_MSsheath.avi, DOI_vx_vx_pSheet.avi and in Figure 5. The Figure shows the DOI map for v_x (panel (a) and (b)) and B_x (panel (c) and (d)) with observation point in the plasma sheet (panel (a) and (c)) and in the magnetosheath (panel (b) and (d)) at $t_0 + 192$ minutes. All the previous figures, Fig. 1, 2, 3 and 4, were at $t_0 + 172$ minutes.

We note that the plots with observation points in the magnetosheath are not significantly different to earlier plots (see Figures 3, 4), except for the plasma sheet plots, which differ profoundly.

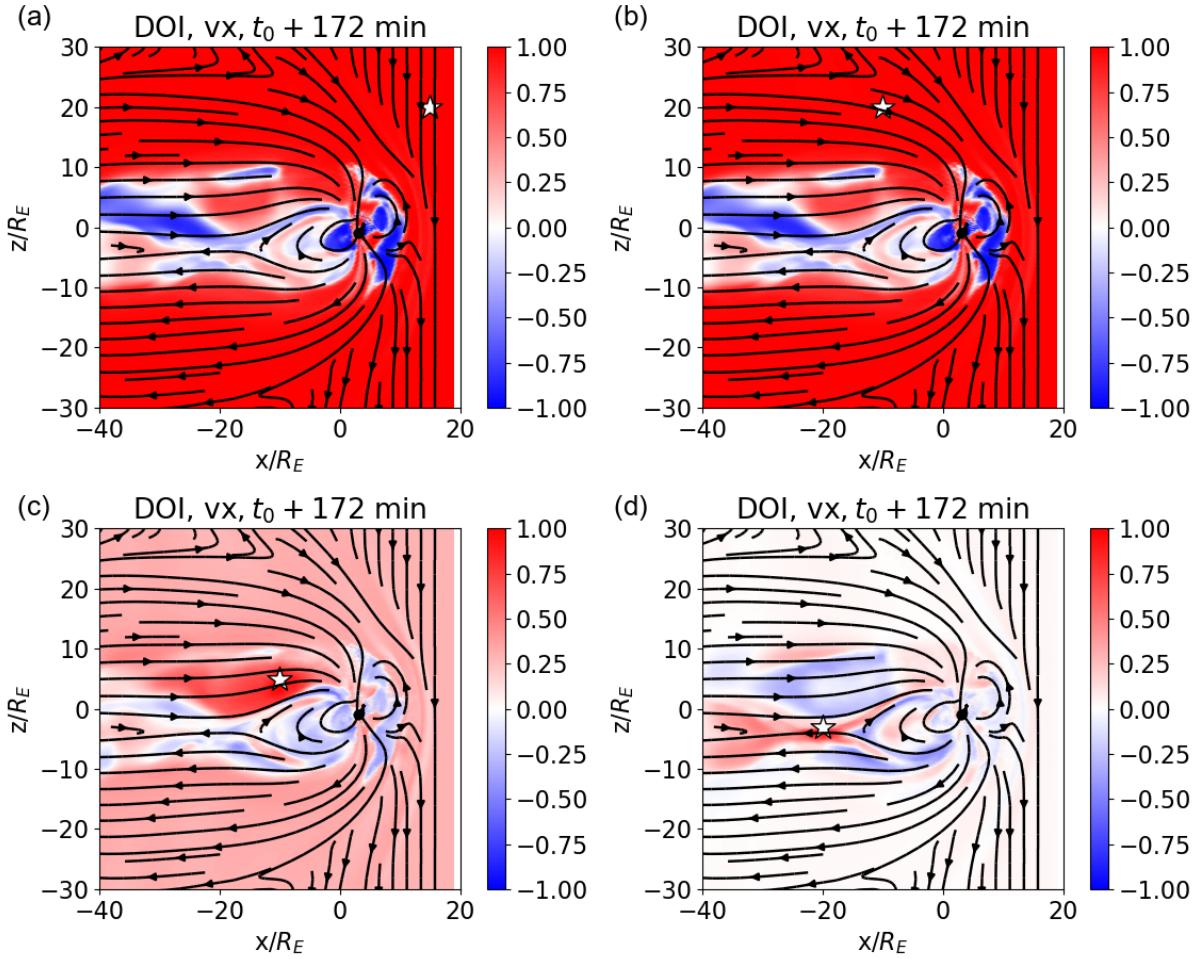


Figure 3. DOI maps for v_x , computed from the correlation of an ensemble of OpenGGCM magnetosospheric simulation, with observation points in the solar wind (panel (a)), magnetosheath (panel (b)), northern lobe (panel (c)), plasma sheet (panel (d)).

To summarize and interpret the OpenGGCM results, the DOI analysis is well in line with our understanding of the terrestrial magnetosphere. In the v_x case, when the observation point is in the solar wind or in the magnetosheath, the $|DOI|$ values are very high in both the solar wind and the magnetosheath region. This is expected, because v_x in the solar wind is a correlation with itself (and thus a sanity test for the calculation), whereas the magnetosheath is largely driven by the interaction between solar wind and the bow shock, where the Rankine-Hugoniot conditions predict a positive correlation of the downstream velocity with the upstream velocity. When the observation points are in the solar wind and magnetosheath regions, the $|DOI|$ values in the plasma sheet are expected to be lower due to internal transient dynamics (e.g., reconnection events, bursty bulk flows) in the sheet which may be triggered by local plasma sheet dynamics, rather than solar wind compression. Local dynamics in the sheet is also the reason why, in Figure 3, panel (d), when the observation point is in the plasma sheet, the correlation with the solar wind and magnetosheath regions is close to zero. Even if, in global terms, magnetic reconnection in the plasma sheet were triggered by magnetopause dynamics, in any region of the plasma sheet v_x may flow sunward or anti-sunward, depending on the location of the reconnection site, and thus would be uncorrelated with the velocity in the solar wind or in the magnetosheath. The lobe magnetic field is expected to be directly driven by the solar wind dynamic pressure, and thus by v_x . As the dynamic pressure

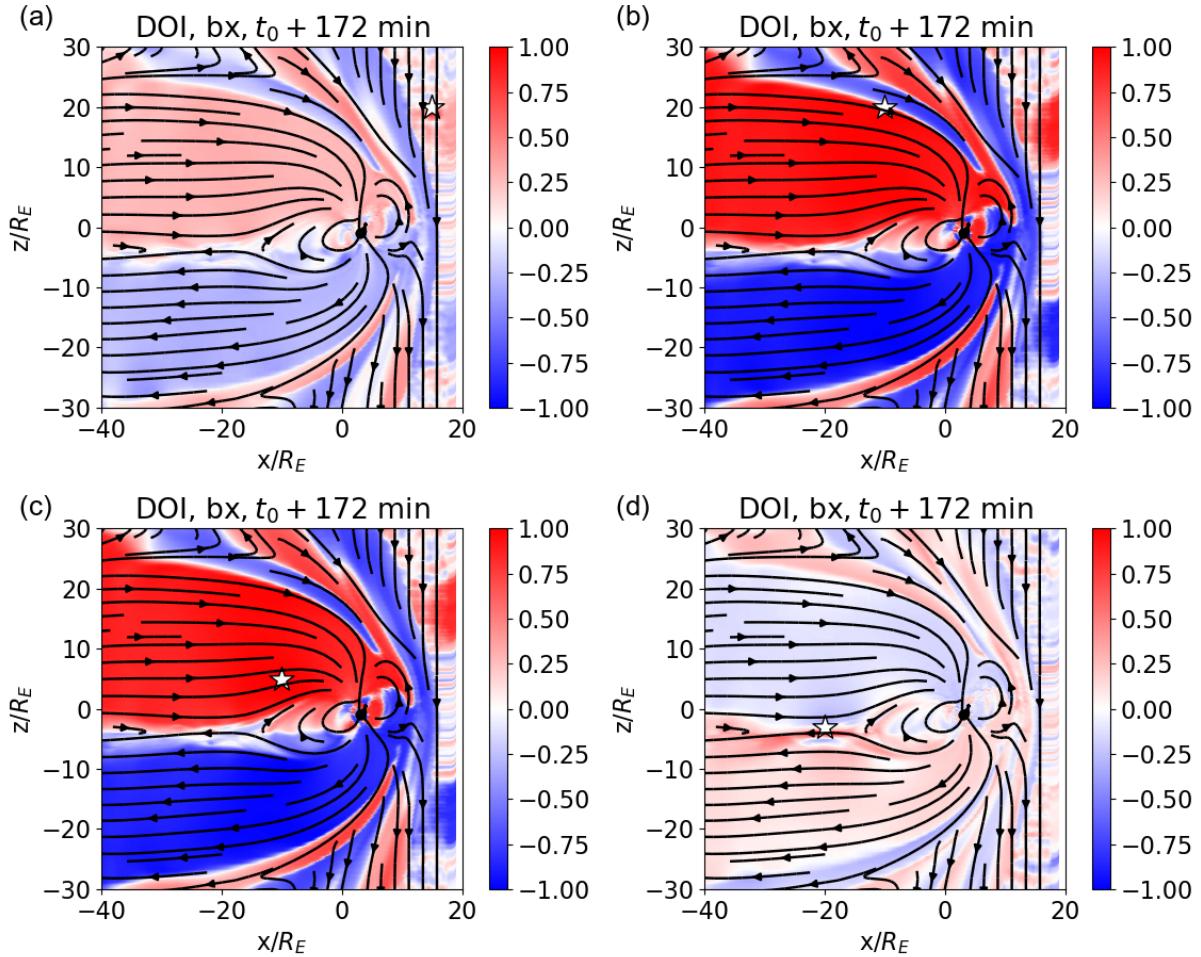


Figure 4. DOI maps for B_x , computed from the correlation of an ensemble of OpenGGCM magnetospheric simulations, with observation points in the solar wind (panel (a)), magnetosheath (panel (b)), northern lobe (panel (c)), plasma sheet (panel (d)).

increases, the lobe flare angle decreases, and vice versa. As the flare angle decreases, the lobe field gets compressed. Figure 4, panels (b) and (c) show that effect, as expected.

Similar consideration broadly apply to the B_x DOI results shown in Figure 4. There is, however, a significant difference between panels (a) in 3 and 4. When the observation point is in the solar wind, high correlations are obtained in large parts of the magnetosphere for v_x , while B_x correlations are much lower. This can be attributed to the fact that B_x in the solar wind is not a major driver of magnetospheric dynamics, unlike the solar wind speed and solar wind dynamic pressure. The geo-effective component of the interplanetary magnetic field (IMF) is the B_z component, which controls reconnection at the magnetopause and thus the dominant energy input into the magnetosphere.

3.2 Magnetospheric applications II: Tsyganenko Model

The Tsyganenko models are a family of empirical, static terrestrial magnetic field models (Tsyganenko, 1987, 1989, 1995, 2002a,b; Tsyganenko et al., 2003; Tsyganenko and Sitnov, 2005). The successive model versions (available at <http://geo.phys.spbu.ru/~tsyganenko/modeling.html>) reflect increasing knowledge of the magnetospheric systems and are based on an increasing amount of data from all regions in the magnetosphere.

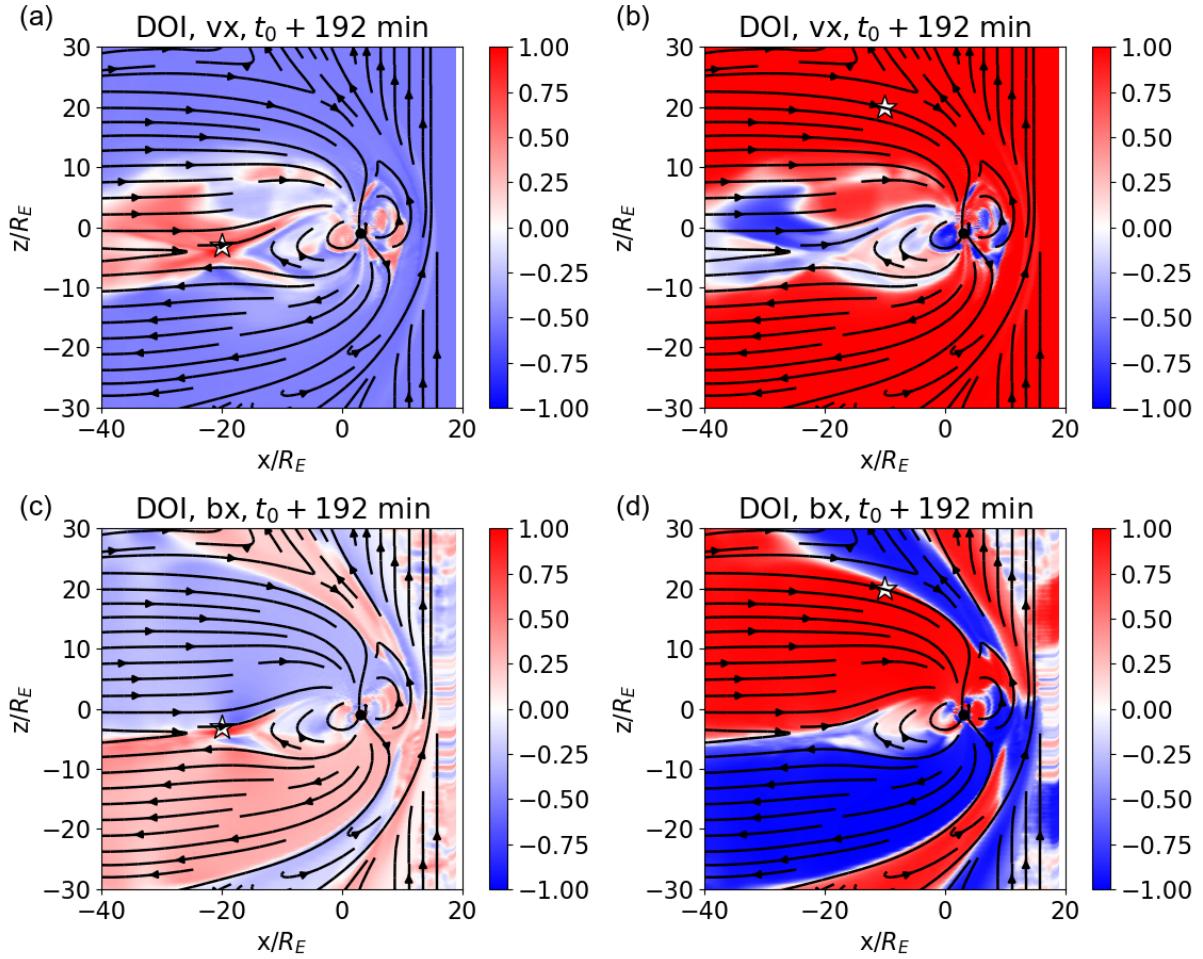


Figure 5. DOI maps for v_x (panel (a) and (b)) and B_x (panel (c) and (d)) from an ensemble of OpenGGCM magnetosospheric simulations with observation point in the plasma sheet (panel (a) and (c)) and in the magnetosheath (panel (b) and (d)), at $t_0 + 192$ minutes. All previous Figures were at $t_0 + 172$ minutes.

The models are based on a mathematical description of the magnetosphere, which includes contributions from major magnetospheric current sources such as the Chapman-Ferraro current, the ring current, the cross-tail current sheet and large-scale field-aligned currents. Terms are added to account for the magnetopause and for partial penetration of the IMF into the magnetosphere. The most recent versions can also take into account the dipole tilt, the dawn-dusk asymmetry, and allow for open magnetospheric configurations. The parameters of the models are derived from a regression to magnetic field observations, and keyed to magnetic indices and/or solar wind parameters. The model requires the user to specify a date and time for the dipole orientation. The other model parameters, either an index such as the Kp, or solar wind variables, are to be given by the user. In more recent models, Tsyganenko also provides yearly input data files for his models. From these inputs, an approximation of the magnetosphere is created for the specified date and time. Notice that the Tsyganenko models are static, and only provides a snapshot of the magnetosphere. However, since the parameters are time dependent the model can be used in a quasi-dynamic mode.

Several versions of the Tsyganenko model have been tested over the years against observations and physics-based, MHD models (Thomsen et al., 1996; Huang et al., 2006; Woodfield et al., 2007). While the Tsyganenko models do not account for the Earth's internal magnetic field, methods are provided to add the internal field model as described in the above cited literature.

In order to simulate the evolution of the magnetosphere with the chosen Tsyganenko model, we create snapshots of the magnetosphere at different times. The time May 8th, 2004, 09:00 UTC is taken as t_0 , the same time as the OpenGGCM simulations presented in Section 3.1. The model is “evolved” by using a time series of the required input parameters, which are obtained from the OMNIWeb database (King and Papitashvili, 2005).

We use two versions of the Tsyganenko model, the T96 model (Tsyganenko, 1996) and the TA15 model (Tsyganenko and Andreeva, 2015). We generate the Tsyganenko ensembles in the same way as the OpenGGCM examples, by using a distribution of v_x values as described in section 3.1.

Before we analyse the results of the T96 ensemble, we show the magnetospheric configuration computed by the model using the original solar wind data. In Figure 6, the first row of figures shows the results of the “reference” simulation, e.g. the simulation without any perturbed inputs, at time $t_0 + 85\text{min}$, for B_x (panel (a)) and B_z (panel (b)).

Unlike the OpenGGCM, the T96 model cannot model reconnection, although some approximation of reconnection is included in later Tsyganenko models (Tsyganenko, 2002a,b). Also, the day-side magnetospheric structure is only approximated with respect to physics-based models, and bow shock and magnetosheath are not clearly distinguishable. In Figure 6, the second row shows the average of the ensemble at the same time of the reference simulation, for B_x (panel (c)) and B_z (panel (d)). The results are similar to the reference simulation, as shown by the logarithm of the absolute difference between the reference and ensemble mean, e.g., Figure 6, panel (e) and (f). The only significant difference is located at the magnetopause, which is expected since varying the solar wind velocity changes the standoff distance.

Next we analyse the DOI maps of the T96 model. Figure 7 shows the DOI maps for the B_x and B_z field components at time $t_0 + 85\text{ min}$. Although the T96 model is parameterized by the solar wind velocity, it only models the magnetic field in the magnetosphere. Because of this, we are only able to analyze the DOI maps of the magnetic field components. The observation points are placed in the northern lobe, in proximity to the current sheet, at the dayside magnetosphere, and in the southern lobe. Like in the OpenGGCM case, the DOI maps reflect the general regions of the magnetosphere as reproduced by the T96 model. However, the correlation only takes values of ± 1 in the magnetosphere, and zero in the solar wind. The latter is simply a consequence of the fact that the model does not predict the IMF, which is therefore independent of the v_x variations of the ensemble. The former is due to the fact that the model has no intrinsic time dependence. Any variations of the solar wind affect the entire magnetosphere instantly and in proportion to the variation. Thus, after normalization, only the sign matters, i.e., whether a given change at the observation point leads to a positive or negative change at a different point.

Now we focus on the results of the TA15 model. Figure 8 shows the reference simulation and ensemble mean for the B_x and B_z fields, with superimposed field lines, at time $t_0 + 85\text{ min}$, together with the difference between reference and ensemble mean. We observe that the reproduced dayside magnetosphere structure is improved compared to the T96 model, at the expense of unrealistically high magnetic field values in the inflowing solar wind, and correspondingly distorted magnetic field lines. These artificial boundary conditions in the Sunwards boundary are used to obtain an “open” magnetosphere which blends with the inflowing solar wind, without seeming to form a nightside magnetosheath. Notice also the high values at these artificial boundary conditions in the difference plot, indicating that there is a high variability in their values.

From the DOI maps in Figure 9 (with observation points at the same positions as Figure 7), we can confirm that the modelled IMF is used to construct the internal magnetospheric solution. While in the

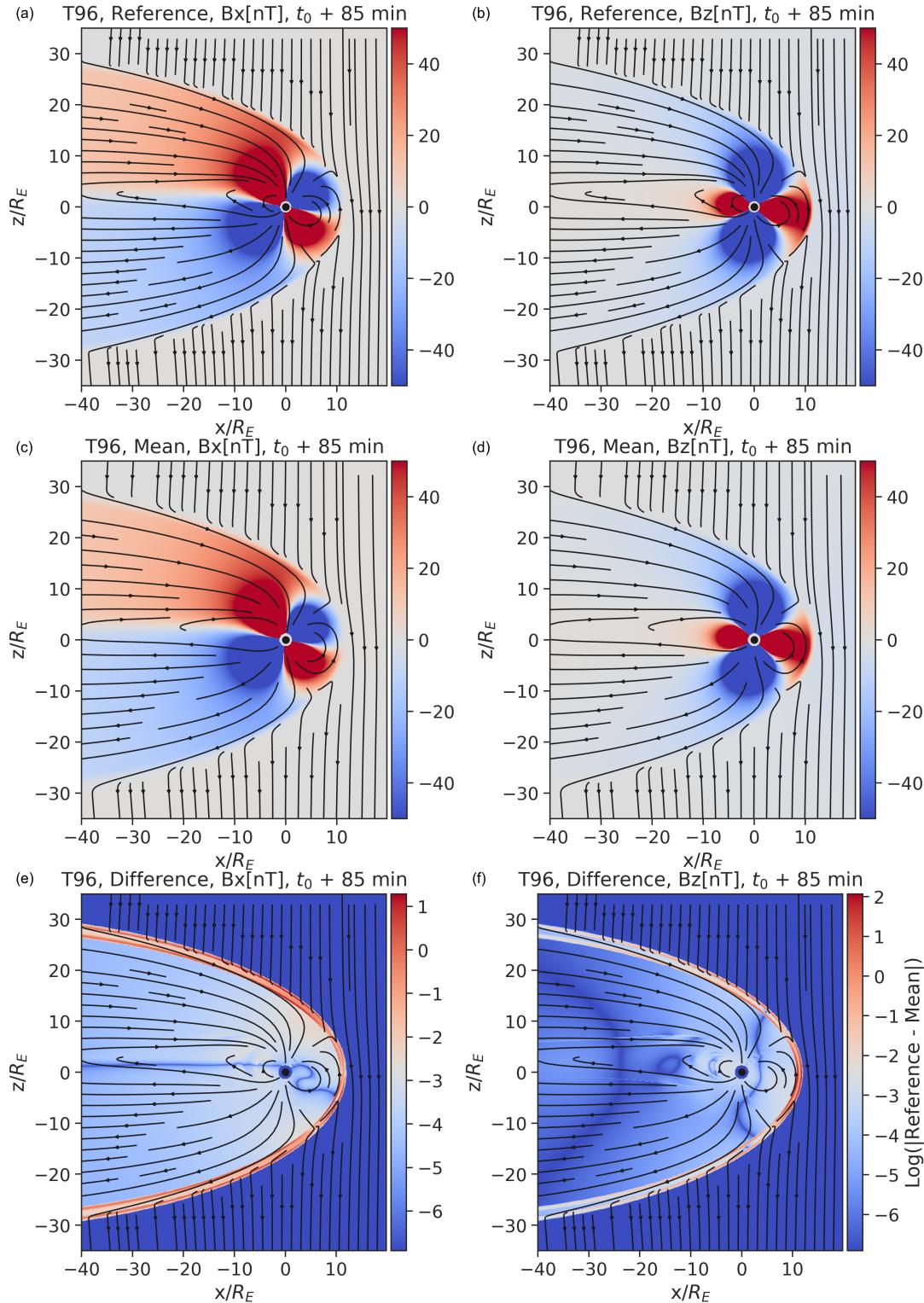


Figure 6. Results from the T96 Tsyganenko model. The top row shows the reference magnetic field (B_x and B_z component) at time $t_0 + 85$ min, with a negative IMF. The middle row shows the ensemble mean of the same magnetic field components at the same time $t_0 + 85$ min. We clipped the magnetic field values to $|50nT|$, in order to make variations in the tail better visible. The last row shows the logarithm of the absolute difference between the reference simulation and the mean of the ensemble.

T96 model the solar wind B_x and B_z values were uncorrelated with the magnetospheric values, here the absolute value (i.e. ignoring the sign) of the correlation is very high: the solar wind input strictly determines the inner magnetospheric solution, making the correlation practically unitary. This could be because of the deterministic analytical formula used to construct the magnetic field, where everything is exactly determined on a global scale.

Note that the correlations reported are spatial and not temporal, therefore no causality is implied. High correlation between the IMF and magnetospheric fields point to the fact that, in an ensemble generated by perturbing the solar wind input, the model is built in such a way that variations in the magnetic field are highly correlated through the system, apparently without highlighting the boundary regions that we were able to spot in the DOI maps for the OpenGGCM and Tsyganenko T96 simulations.

A last remark on the DOI analysis applied to the Tsyganenko models is the following. The analysis helps us understand and visualize how the different models are built, with regards to the relationship between the solar input and the magnetospheric solution. DA analysis then proves useful here as a model investigation tool.

It also highlights that caution should be used when deciding to apply DA techniques to a particular model, depending on the objectives of the investigation. The Tsyganenko models were built to provide time-independent, empirical-based insights into the structure of the magnetosphere at a particular instant in time. They do not aim at representing the state of the pristine solar wind, which is used only to better the magnetospheric solution (hence the somehow unrealistic solar wind patterns identified in Figure 8 and 9). Also, they do not intend to reproduce temporal dynamics in the magnetosphere. These factors result in DOI maps where the absolute value of the correlation is always either 1 or -1. When using DOI techniques with the purposes of identifying useful locations for satellite placement, these are not useful results: we are interested in the *value* of the DOI, not in the *sign*. Hence, caution should be used before using empirical, time-independent models for this particular purpose: more significant information will possibly be acquired from their physics-based, time-dependent counterparts. This consideration does not intend to diminish the importance of empirical, time-independent models for other scientific objectives such as, most importantly, quick forecasting.

3.3 Heliospheric application: PLUTO

In this section, we study the propagation of a Coronal Mass Ejection in a solar meridional plane, which is defined by the rotation axis of the Sun and a radial vector in the equatorial plane. In all the runs of the ensemble, the computational domain is $1R_{\odot} \leq r \leq 216R_{\odot}$ and $0 \leq \theta \leq \pi$ in spherical coordinates, where R_{\odot} is the solar radius and θ is the polar angle (or colatitude), corotating with the Sun. Assuming axisymmetry around the solar rotation axis, we may limit our analysis to 2.5D (pseudo 3D) simulations. The grid resolution is uniform in both directions, 384×384 cells, which is sufficient to capture the structure of the background solar wind while keeping the computational cost and output size manageable.

We simulate the background solar wind using a simple adiabatic model with effective polytropic index $\Gamma = 1.13$ (Keppens and Goedbloed, 2000). We also assume a time-independent dipole background magnetic field:

$$B_r = -2B_o \cos \theta / r^3, \quad (13)$$

$$B_\theta = -B_o \sin \theta / r^3, \quad (14)$$

where B_r, B_θ are the r, θ components of the magnetic field in spherical coordinates and B_o a constant used to scale the field to $B = 1.1\text{G}$ on the solar equator. We impose the density distribution ρ as a function of the latitude θ at the inner boundary to achieve a “dead zone” of low velocity near the equator and a fast solar wind near the poles simultaneously (see Keppens and Goedbloed (2000) and Chané et al. (2008)). The differential rotation of the Sun is also taken into account, following Chané et al. (2005); this is achieved by imposing a varying azimuthal velocity $v_\phi = v_\phi(\theta)$ at the inflow boundary.

Once the simulation reaches a steady state, roughly after ~ 2.5 days or $t = 10$ in normalized units, the radial velocity at 1 AU is ~ 300 km/s near the equator and ~ 850 km/s near the poles. This is consistent with the large-scale bimodal solar wind structure that is typically observed during solar minimum (McComas et al., 1998)).

We create two ensembles of 100 simulations each. In the first ensemble, the velocity of the CME in each case is randomly selected from a Gaussian distribution with mean $\mu = 900$ km/s and standard deviation $\sigma = 25$ km/s. The resulting values are typical of strong CME events. In the second ensemble, the spatial extent of the boundary conditions that launch the CME varies as well, along with the velocity of the CME as described above. The half-width of this region is also randomly selected from a Gaussian distribution with $\mu = 10^\circ$ and $\sigma = 0.5^\circ$. All other parameters remain the same in every run.

The values of the CME widths that are used here are comparable to observed events. The choice of parameters in the second ensemble is less constrained by observations and leads to the appearance of very small values of variance. We thus find large areas where the DOI ~ 1 , since the simulations in the ensemble do not differ significantly. This was confirmed by creating and analyzing a third ensemble, where the width of the CME is chosen from a Gaussian with $\mu = 20^\circ$ and $\sigma = 2.0^\circ$.

Figure 10 shows the evolution of the radial velocity average over the whole ensemble. Up to $t=11$, i.e., before the CME is launched, all runs are identical. The CME is initialized similar to the simplified approach of Keppens and Goedbloed (2000), such that the boundary conditions on the solar surface are modified to represent a change of mass flux. In our case, we modify the boundary conditions at $R=1R_\odot$, in a given region around $\theta = 80^\circ$. A tracer (a passive scalar only present as an advected quantity within the flow, without effect on the plasma) is also injected with the CME, to facilitate monitoring its propagation. In the middle and right panel of Fig. 10 we show the ejection of the CME and its propagation. The CME front can be clearly distinguished at $t=12$.

To apply the RA technique, as explained in Section 2, we select a point of interest and perform the analysis based on (a) the plasma density, or (b) the radial velocity. We present results at $t=14$, when the CME has reached a distance of $\sim 150R_\odot$, for two detection points (at $R = 90$ and $R = 150R_\odot$, $\theta = 80^\circ$). At times earlier than $t = 10$ (when the solar wind reaches a steady state and the CME is injected), the DOI is zero, since the observation point is disconnected from the rest of the domain before the CME reaches it.

The propagation of the CME can be monitored in the MHD simulations easily via e.g. a tracer (or the radial velocity). The DOI map, when the tracer is used as a criterion, follows closely the CME propagation

pattern observed in the MHD runs. However, this is of limited use, besides testing, as the tracer (in our case) does not represent a real physical quantity.

The DOI map for the first ensemble, where we perturb only the radial velocity of the CME, is shown in Fig. 11. The regions where information from the CME front has not yet arrived have $DOI=0$, as shown in the radial velocity DOI map (Fig. 11). When only one parameter is modified (first ensemble), the density DOI map shows a very large area of the domain saturated with correlation $\simeq 1$. This is probably due to variations in density of the background solar wind induced by the propagation of the CME. The regions with absolute value of the $DOI \simeq 1$ that are located far from the CME propagation front (at small or large angles θ) are the areas of high radial velocity in Fig. 10, where the information on the perturbation introduced when triggering the CME has already propagated. The density and radial velocity of all ensemble members are modified in a similar way, hence the large $|DOI|$ values.

In the second ensemble, where the width of the CME is also modified, the DOI map of the density shows smaller correlation values (compare especially panel (b) and (d) in the two Figures) compared to the previous ensemble, because the differences between the runs of the ensemble are now larger (see Fig. 12). This results in smaller regions where the DOI is close to unity compared to the first case.

The last ensemble, where the CME width and its perturbation are larger compared to the second case, is shown in Fig. 13. The DOI pattern is qualitatively similar to Fig. 12, but due to the larger values in the size of the CME and its perturbation, the regions with high DOI values (meaning the regions affected by CME propagation in at least one member of the ensemble) are slightly larger as well.

Additional analysis, not shown here, was carried out on subsets of the ensembles to ensure the ensemble size is sufficient. We found that in this case convergence was achieved if at least 25 ensemble members were used (as described in Section 2); however, this number may differ in other cases, depending on the specifics of the ensemble.

The DOI analysis applied to the simulations performed with PLUTO are indicative of the versatility of the method. In all PLUTO ensembles, we can monitor the influence of the CME during its propagation and the response of the system via the DOI. Moreover, we can identify certain CME components, such as the leading edge, from the DOI maps. Differences in the response of the system due to the choice of the perturbation or parameters are captured as well. The resolution used here was sufficient to capture the CME injection and propagation within reasonable computational cost; the typical run time for simulating a member of the ensemble was of the order of $\sim 10'$ minutes on 28 cores.

However, some limitations of the model must be considered. The **axisymmetric** assumption simplifies the problem and allows to reduce computation costs, but with the drawback of not accounting for the three-dimensional CME structure. The limited angular resolution imposes a weak constraint on both the perturbed and unperturbed size of the CME **that we can simulate**. Runs with a higher resolution can remove this constraint at additional computational cost. Simulations in 3D will be part of future work in order to capture the full system, where also differences in the polar direction can be examined. Finally, a more realistic model for the background magnetic field should be used, rather than a simple static dipole. We focused mainly on calculating the DOI at different times and locations, but a similar approach can be used to estimate the arrival time of the CME, as described in Owens et al. (2020).

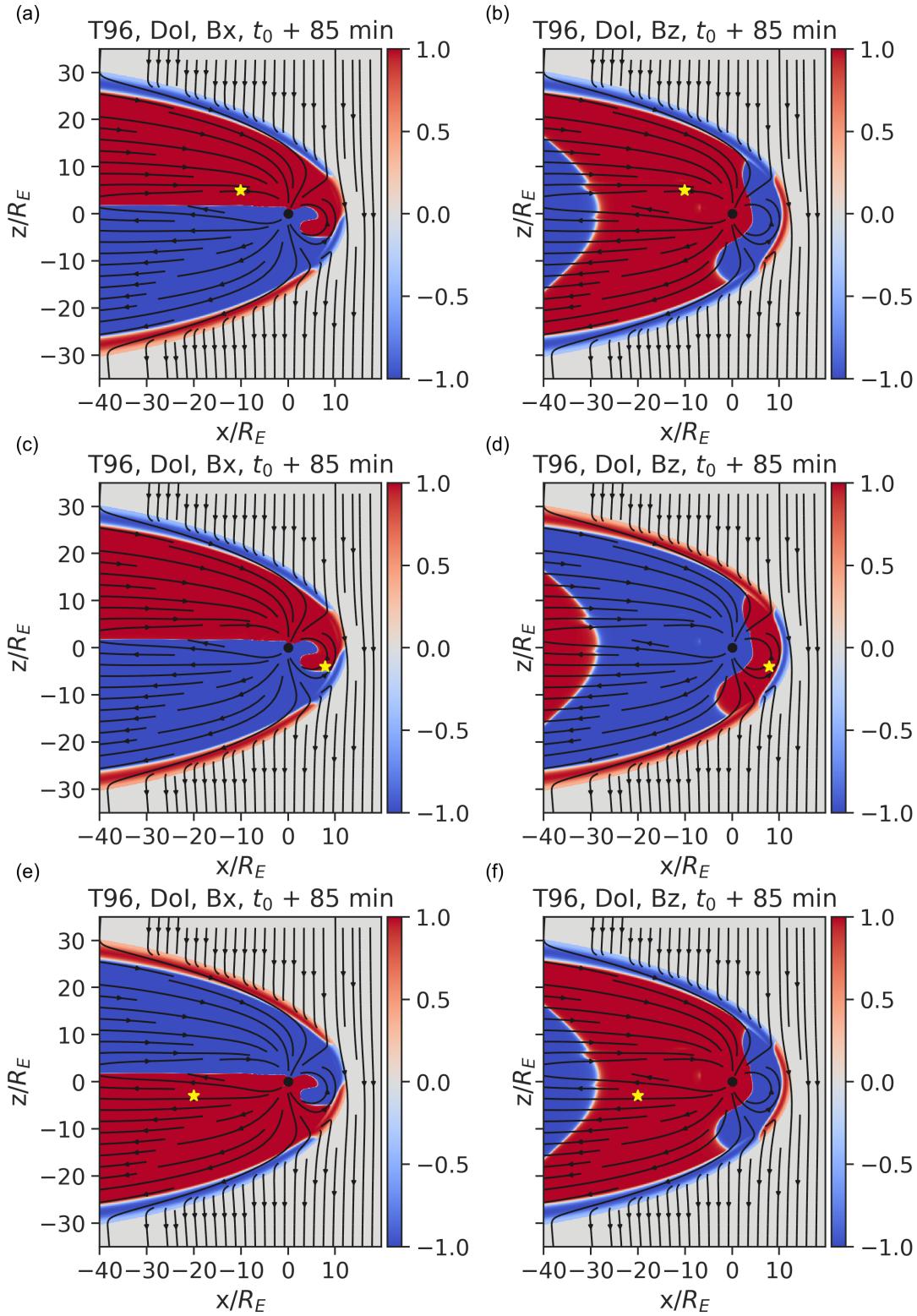


Figure 7. Each row displays the DOI maps for B_x and B_z , from an ensemble of Tsyganenko model T96 simulations with observation point in the plasma sheet, dayside magnetosphere and southern lobe tail, respectively.

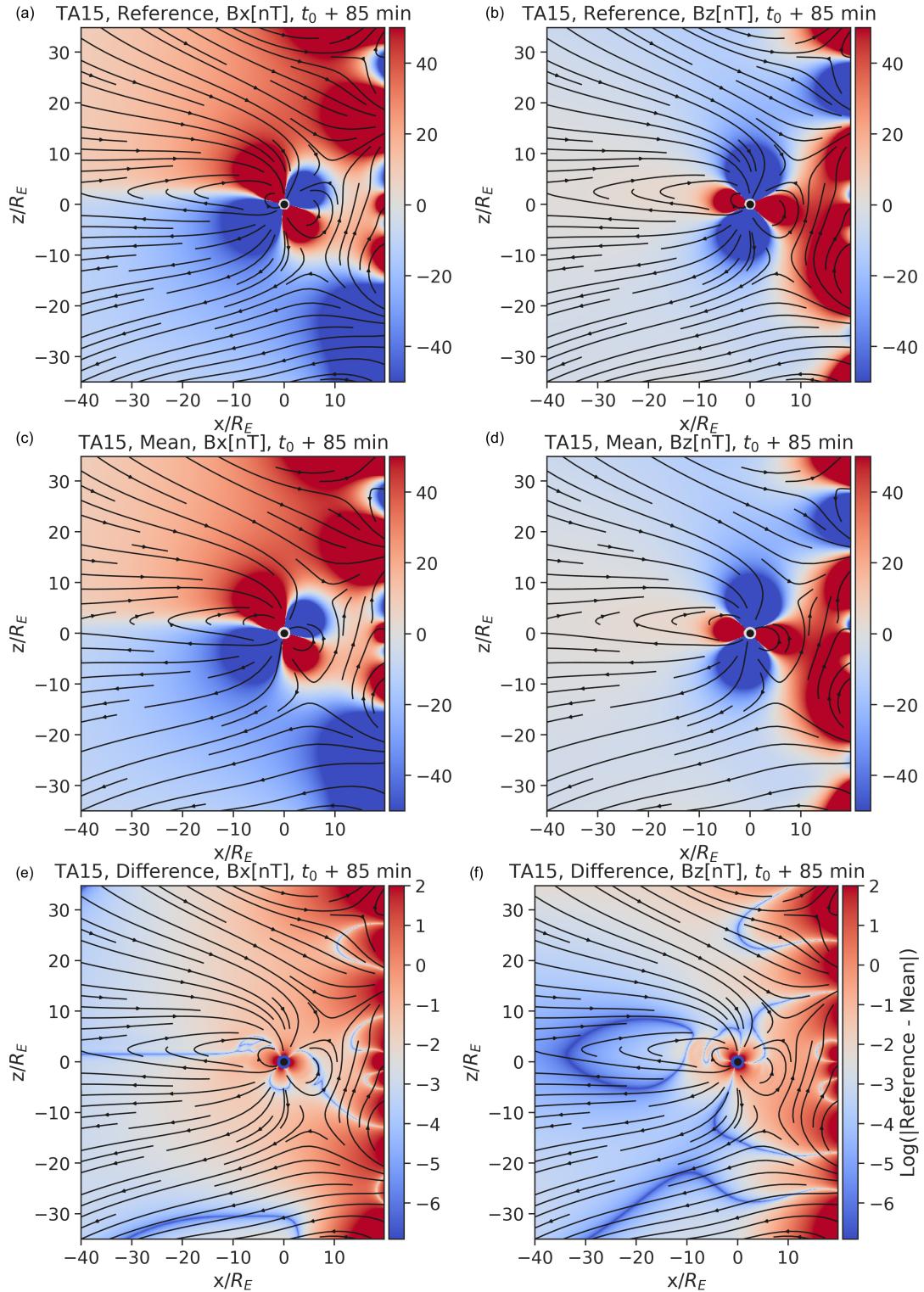


Figure 8. The reference and mean magnetic field at time $t_0 + 85$ min from the Tsyganenko TA15 model are displayed in the top two rows of figures. Notice the non-realistic high magnetic field values in the inflowing solar wind (at $x/R_E > 10$). We assume these unrealistic values are necessary for the model to construct the day-side magnetosphere. The values of the magnetic field have been cut off in the top two rows of figures at $|50nT|$, to make sure the variations in the tail are visible. The last row of figures shows the logarithm of the absolute error between the reference simulation and the mean of the ensemble.

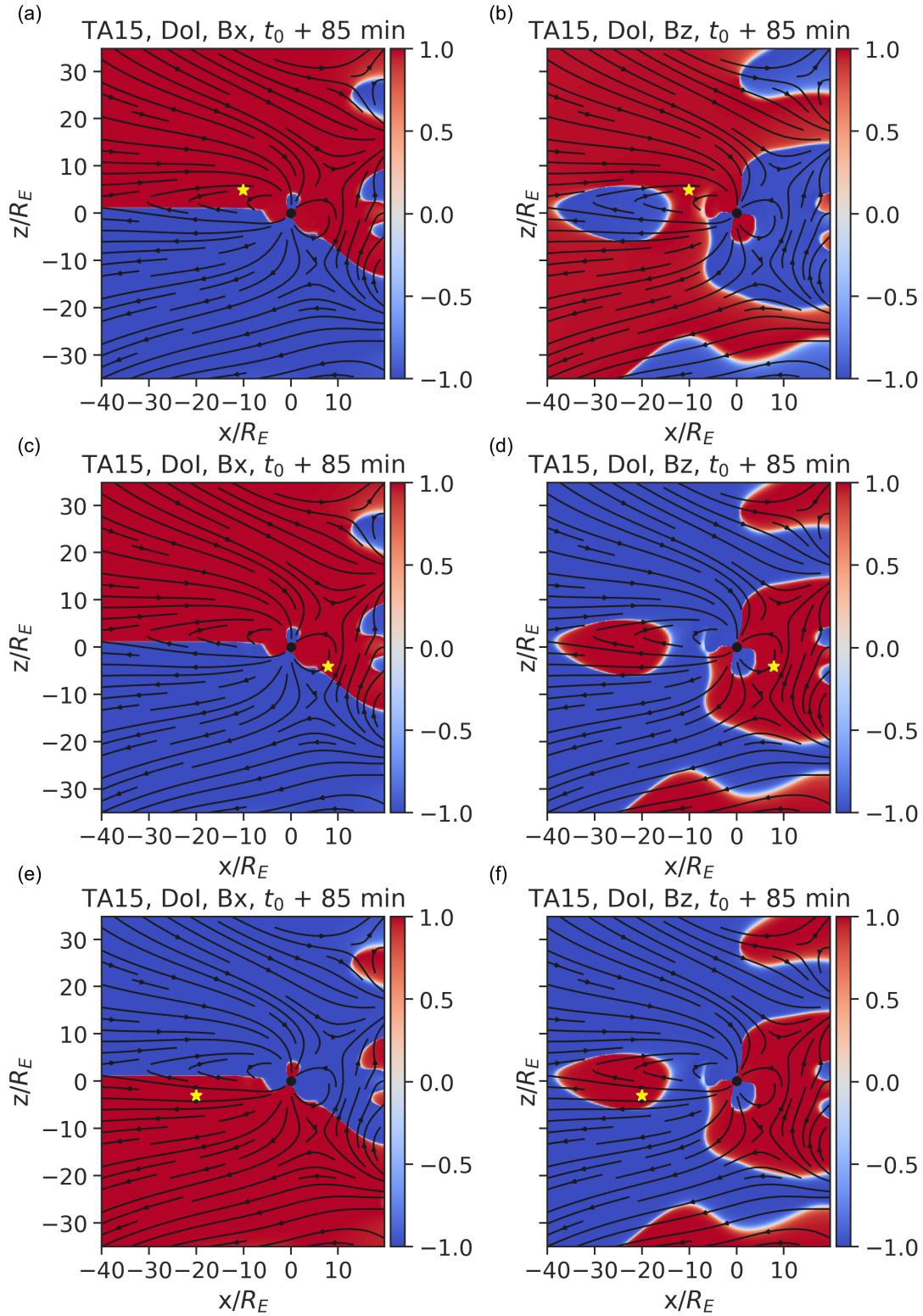


Figure 9. DOI maps for B_x and B_z , computed from an ensembles of Tsyganenko TA15 simulations at $t_0 + 85$ min with observation points at the same position as Figure 7.

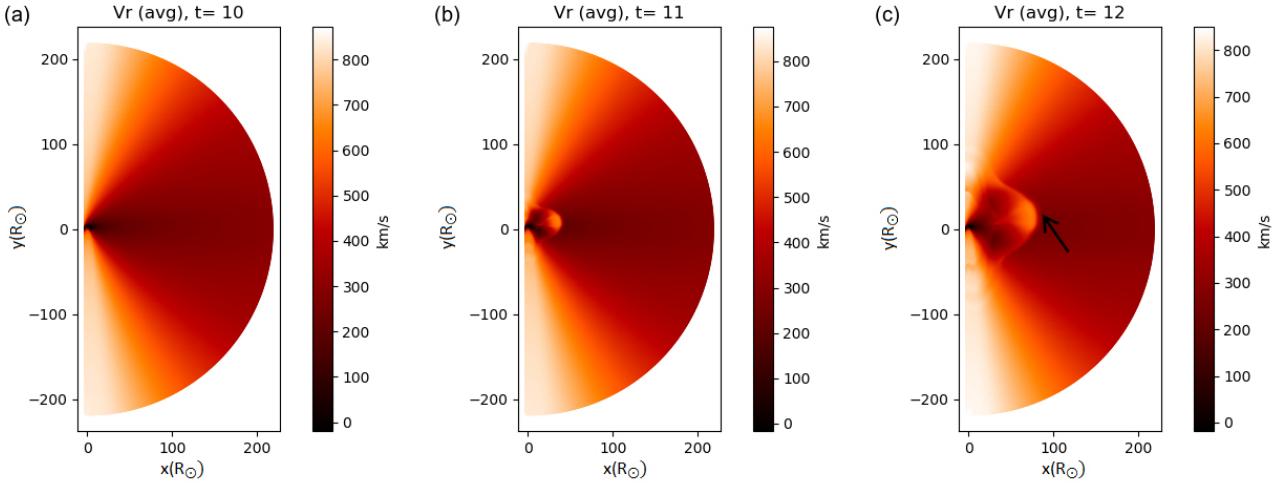


Figure 10. Injection and propagation of a CME via the average radial velocity in the ensemble (in km/s). From left (a) to right (c): relaxed solar wind ($t=10$), injection and propagation of the CME ($t = 11$, $t = 12$). The leading edge (front shock) is prominent in the last panel, marked by a black arrow.

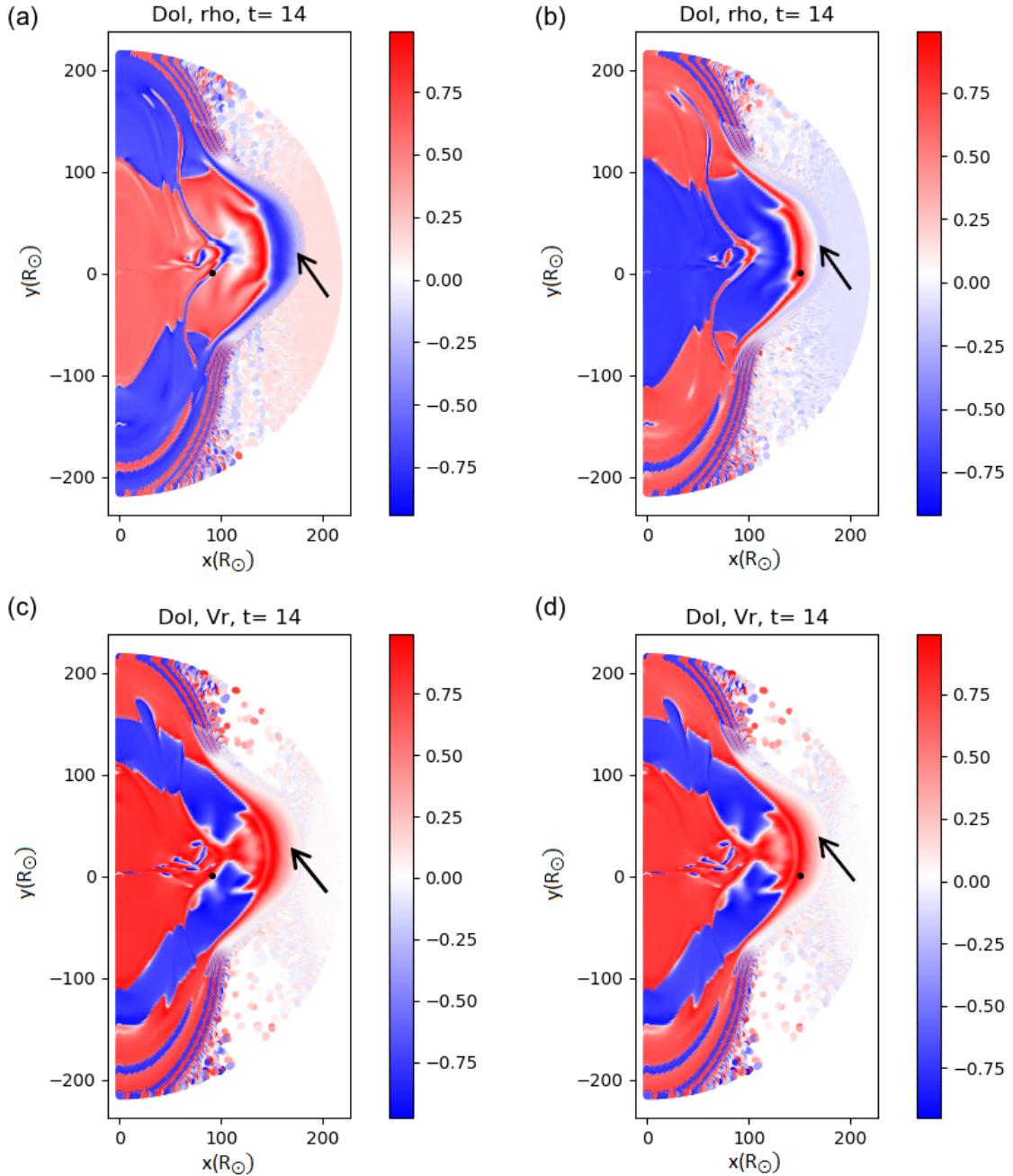


Figure 11. DOI maps for the density (first row) and the radial velocity (second row), from an ensemble of PLUTO simulations. The first column has the observation point at $R = 90R_{\odot}$ and the second column at $R = 150R_{\odot}$ (marked by the black dot). In this set we only perturb the radial velocity of the CME. The structure of the CME can be seen quite clearly in panels (a) and (b), where the leading edge is evident and marked by a black arrow in all panels.

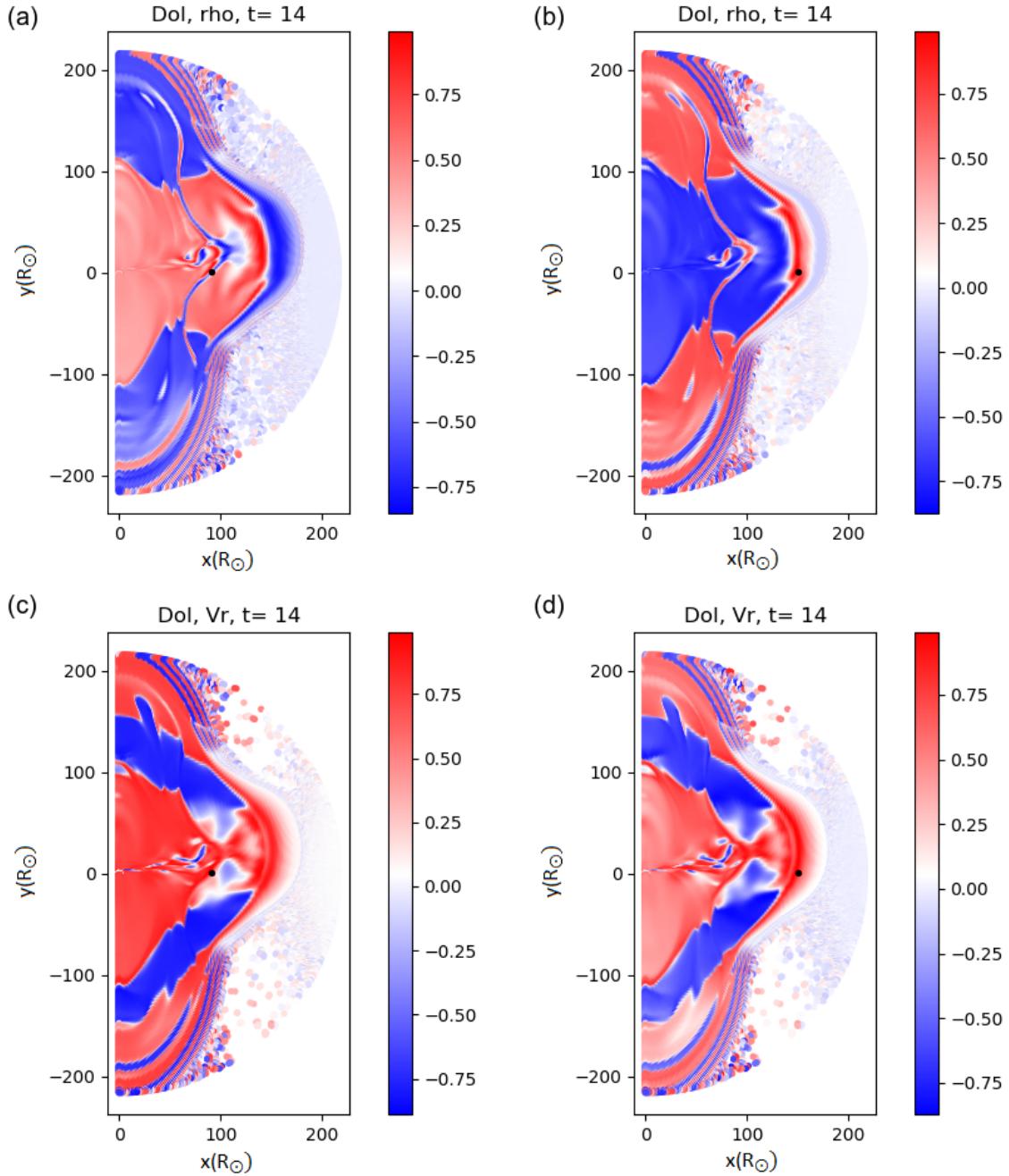


Figure 12. DOI maps for the density (first row) and the radial velocity (second row), computed from PLUTO simulation ensembles. The first column has a detection point at $R = 90R_{\odot}$ and the second column at $R = 150R_{\odot}$ (marked by the black dot). In this set we perturb two parameters, the radial velocity and the size of the CME.

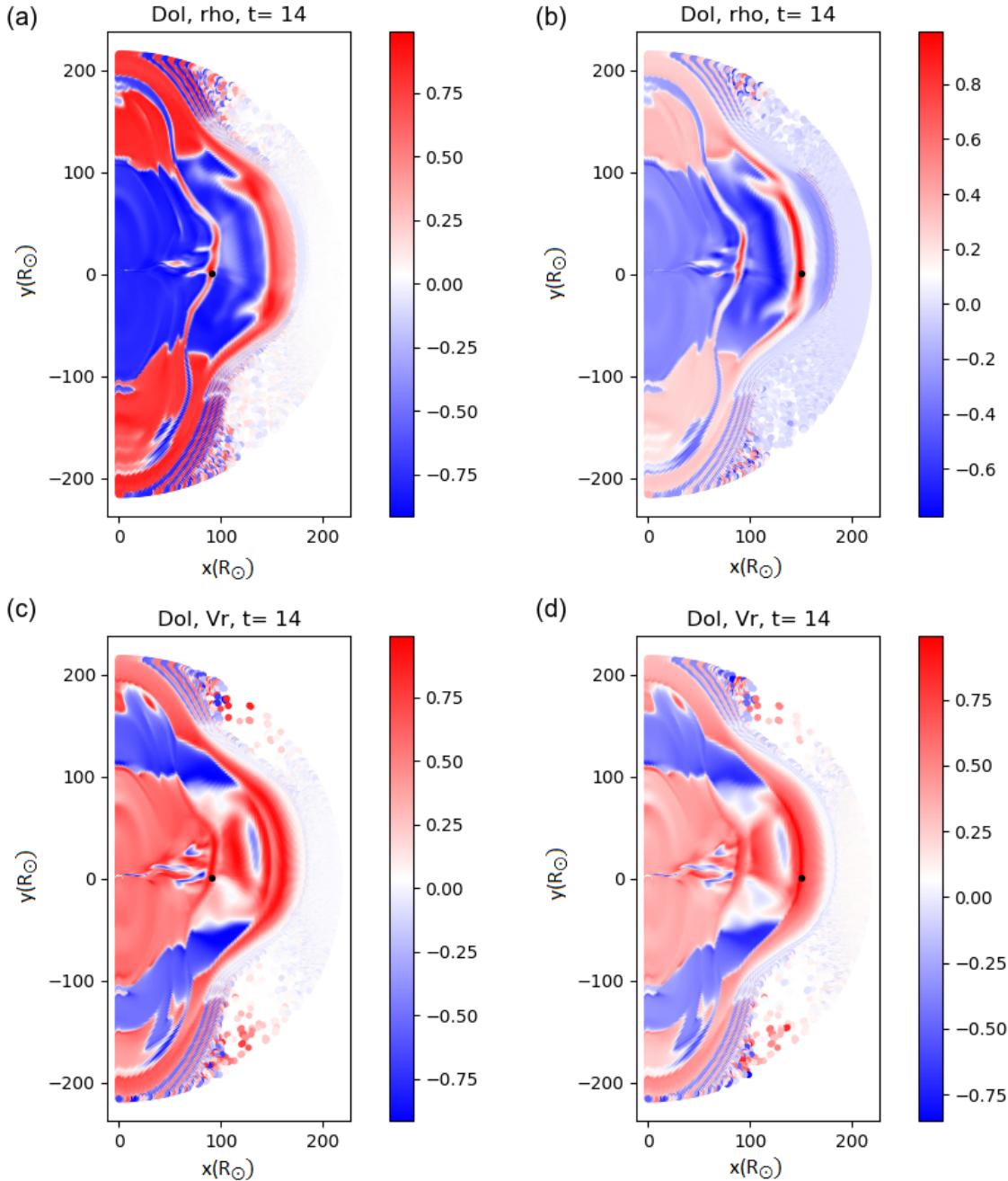


Figure 13. DOI maps from ensembles of PLUTO simulations for the density (first row) and the radial velocity (second row). The first column has a detection point at $R = 90R_{\odot}$ and the second column at $R = 150R_{\odot}$. In this set we perturb two parameters, the radial velocity and the size of the CME, with a larger perturbation in the size with respect to the second data set.

4 SUMMARY AND CONCLUSIONS

In this paper, we apply the Representer analysis and the Domain of Influence analysis to two fundamental components of the Sun-Earth connection: the interaction between the solar wind and the terrestrial magnetosphere, simulated with the OpenGGCM MHD code and with the empirical Tsyganenko models, and the propagation of CMEs in the background solar wind, simulated with the MHD PLUTO model.

In each case an ensemble is generated by appropriately perturbing initial/ boundary conditions. Subsequently, the DOI analysis is applied over the ensemble. Localisation methods, which can be used to reduce spurious correlations in the estimated prior covariance matrix (Anderson, 2007; Bishop and Hodyss, 2007; Sakov and Bertino, 2011), are not used at this stage.

Primarily, the DOI analysis is a first step in the application of Data Assimilation techniques to a model, and can be applied before assimilation itself to gain insight on the system and on the model. However, the DOI analysis can also be used to gain physical insight, and to devise optimized observation systems, as discussed below.

Our main results are as follows.

First, we have demonstrated that DOI analysis can provide useful information on the most appropriate locations for future observation points, such as solar wind and magnetospheric monitors. Large absolute values of the DOI, calculated with respect to an observation point, means that observations at that location would provide significant information of that field in the specific, large $|DOI|$ area, but less so in areas with lower $|DOI|$. This can be used in two different ways. On one hand, DOI analysis can help to find observations points that are connected to large $|DOI|$ areas, in order to increase the amount of information brought in by a single new observation. On the other hand, the same information can be used for a different objective. Given a particular location, one can ask where observations need to be obtained to improve knowledge of that area. A useful example here is the plasma sheet in the OpenGGCM analysis, Section 3.1. Figure 3 and 4 show that $|DOI|$ values in the plasma sheet are consistently low, notwithstanding the field which is examined (v_x or B_x) and the location of the observation point. $|DOI|$ values in the plasma sheet are low even if the observation point is in the plasma sheet itself: $|DOI|$ values, which are of course 1 at the observation point itself, quickly become smaller even a small distance away. Since the plasma sheet is a location of particular importance for space weather forecasting, or basic research for that matter, single s/c in the plasma sheet are of limited use, and rather a constellation of satellites, such as proposed in Angelopoulos et al. (1998); Raeder and Angelopoulos (1998) would be necessary.

Second, we have used the DOI analysis to improve our knowledge of the models we use, and in particular to investigate whether these models are appropriate for the implementation of Data Assimilation. The DOI analysis for the Tsyganenko models in Section 3.2 powerfully highlights the model evolution from version T96 to version TA15. In version T96, the magnetosphere is a closed system, and solar wind conditions are not correlated ($DOI \sim 0$ in Figure 7) with the magnetospheric region. In version TA15, the magnetosphere opens up to solar wind driving, and the correlation between the solar wind region and the magnetosphere becomes very high (Figure 9). One should remember that the Tsyganenko models are supposed to be used to investigate the magnetospheric system, and the solar wind configuration is artfully modified as to give the best representation of the magnetosphere under the specified conditions.

One common aspect of the two versions of the Tsyganenko models is that (with the exception of the solar wind region in T96) the DOI values are always either 1 or -1, for all fields and regions examined. These results appear less realistic than the OpenGGCM results obtained in Section 3.1, where DOI values have

larger variability. The Tsyganenko models differ with respect to OpenGGCM in two fundamental aspects, in that they (a) empirically reconstruct the magnetospheric magnetic field from an array of observations and (b) that they are not time-dependent. Either of these two aspects can contribute to the unrealistically high correlations we observe. Investigations on other models, and specifically on empirical, time-dependent models, will possibly help disentangle the role of these two aspects. At this stage of the investigation, we advance the hypothesis that time-dependent models may be better suited than time-independent models as background models for Data Assimilation techniques.

Third, with this analysis we have highlighted a possible path for future, targeted improvements of global heliospheric models used, among other things, for simulations of CME propagation in the heliosphere. It has long been known that one of the critical aspects of the simulation of CME arrival time is the estimation of the physical parameters to use as initial conditions in the simulations. While some parameters can be easily estimated from remote sensing, others are more difficult to determine properly and their variability affects the accuracy of the forecast (Falkenberg et al., 2010). In this paper, we have shown that DOI analysis could constitute an important stage of a model analysis effort aimed at clarifying which aspects of a model should be prioritized in order to obtain more accurate simulations of CME propagation.

In this study, as a first step, we show DOI maps obtained from the correlations of a single variable calculated between the variable at the observation point and the same variable in the domain under investigation. As demonstrated in Skandrani et al. (2014), cross-correlations can be used to find the influence of one variable upon another.

The results of a DOI cross-correlation analysis can then be used to determine which quantities and areas in a simulation are most relevant in determining a certain observational quantity (such as the radial velocity of a CME in the case of CME propagation simulations). This analysis can then guide modelers on deciding which aspects of a model could be improved for more realistic results. It could help understanding, for example, if CME propagation in a model is mainly controlled by the background magnetic field configuration or by the properties of the CME itself at launch. In the first case, modeling efforts could be directed into accurate high resolution representation of the magnetic field configuration in the lower corona. In the second case, instead, modeling improvements could be focused on extracting better estimates of CME launch parameters (e.g. CME density, velocity, internal magnetic field configuration with respect to the background wind) from available observations.

The spatial correlations provided by DOI can also be of particular interest in evaluating the effect of actual measurements done at positions different from the traditional L1, such as, for example, missions planned for L5 or missions closer to the Sun.

Future work will extend this study to include temporal and cross correlations between different field components. This will further increase our knowledge of the models used to simulate such critical space weather processes.

The DOI analysis presented here can also be combined with an Observing System Simulation Experiment (OSSE), an approach already used in ionospheric and solar dynamo studies (Hsu et al., 2018; Dikpati, 2017) to help provide a cost-effective approach to the evaluation of the potential impact of new observations. OSSE requires that DA is already implemented and uses independently simulated “data” that are ingested into a different model or a different instance of the same model. The effect of DA can then be investigated, albeit with caveats, since the “data” are not real. DOI analysis would obviate the need to have DA implemented, which can be very costly. Instead, only ensemble runs with an unmodified model are required, and can provide a measure of the usefulness of a model and the available data for a specific situation.

CONFLICT OF INTEREST STATEMENT

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

AUTHOR CONTRIBUTIONS

MEI, BL and DM created the ensembles using OpenGGCM, Tsyganenko and PLUTO respectively, as they appear in the main body of the article, processed the data and presented the results of the analysis and prepared the manuscript. JR provided OpenGGCM and contributed to the interpretation of the magnetosphere model results. JR, GL and SP provided support during the analysis and the preparation of the article.

FUNDING

DM acknowledges support from AFRL (Air Force Research Laboratory)/USAF (US Air Force) project (AFRL Award No. FA9550-14-1-0375, 2014-2019) and partial support by UK STFC (Science and Technology Facilities Council) Consolidated Grant ST/S000240/1 (UCL-MSSL, University College London-Mullard Space Science Laboratory, Solar System). M.E.I.'s work is supported by an FWO (Fonds voor Wetenschappelijk Onderzoek – Vlaanderen) postdoctoral fellowship. BL, MEI, GL, JR acknowledge funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 776262 (AIDA, Artificial Intelligence for Data Analysis, www.aida-space.eu). JR also acknowledges support through AFOSR grant FA9550-18-1-0483 and from the NASA/THEMIS mission through a subcontract from UC Berkeley. SP and GL acknowledge funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 870405. These results were also obtained in the framework of the projects C14/19/089 (C1 project Internal Funds KU Leuven), G.0A23.16N (FWO-Vlaanderen), C 90347 (ESA Prodex), and Belspo (Belgian Science Policy) BRAIN project BR/165/A2/CCSOM.

ACKNOWLEDGMENTS

The PLUTO simulations were performed using allocated time on the clusters Genius and Breniac. The computational resources and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by the Research Foundation Flanders (FWO) and the Flemish Government – department EWI. The OpenGGCM simulations were performed on the supercomputer Marconi-Broadwell (Cineca, Italy) under a PRACE allocation. We acknowledge the NASA National Space Science Data Center, the Space Physics Data Facility, and the ACE Principal Investigator, Edward C. Stone of the California Institute of Technology, for usage of ACE data. We acknowledge use of NASA/GSFC's Space Physics Data Facility's OMNIWeb service, and OMNI data.

SUPPLEMENTAL DATA

DATA AVAILABILITY STATEMENT

The OMNIWeb database was used to get the solar wind conditions for the Tsyganenko runs. The ACE database was used to get the solar wind conditions for the OpenGGCM runs. The data used in the analysis are available upon reasonable request.

REFERENCES

Allen, J., Sauer, H., Frank, L., and Reiff, P. (1989). Effects of the march 1989 solar activity. *Eos, Transactions American Geophysical Union* 70, 1479–1488

Altschuler, M. D. and Newkirk, G. (1969). Magnetic fields and the structure of the solar corona. *Solar Physics* 9, 131–149

Anderson, B. J., Korth, H., Welling, D. T., Merkin, V. G., Wiltberger, M. J., Raeder, J., et al. (2017). Comparison of predictive estimates of high-latitude electrodynamics with observations of global-scale birkeland currents. *Space Weather* 15, 352–373. doi:10.1002/2016sw001529

Anderson, J. L. (2007). Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter. *Physica D: Nonlinear Phenomena* 230, 99–111

Angelopoulos, V., Carlson, C. W., Curtis, D. W., Harvey, P., Lin, R. P., Mozer, F. S., et al. (1998). On the necessity and feasability of a equatorial magnetospheric constellation. In *Science Closure and Enabling Technologies for Constellation Class Missions*, eds. V. Angelopoulos and P. V. Panetta (University of California, Berkeley, and NASA Goddard Space Flight Center). 14

Arge, C. N., Henney, C. J., Koller, J., Compeau, C. R., Young, S., MacKenzie, D., et al. (2010). Air force data assimilative photospheric flux transport (adapt) model. In *AIP Conference Proceedings* (AIP), vol. 1216, 343–346

Bennett, A. F. (1992). *Inverse methods in physical oceanography* (Cambridge university press)

Berchem, J., Raeder, J., and Ashour-Abdalla, M. (1995). Reconnection at the magnetospheric boundary: Results from global MHD simulations. In *Physics of the Magnetopause*, eds. B. U. Sonnerup and P. Song. vol. 90 of *AGU Geophysical Monograph*, 205

Bishop, C. H. and Hodyss, D. (2007). Flow-adaptive moderation of spurious ensemble correlations and its use in ensemble-based data assimilation. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography* 133, 2029–2044

Bishop, G., Welch, G., et al. (2001). An introduction to the kalman filter. *Proc of SIGGRAPH, Course* 8, 41

Bothmer, V. and Daglis, I. A. (2007). *Space weather: physics and effects* (Springer Science & Business Media)

Bouttier, F. and Courtier, P. (1999). Data assimilation concepts and methods. meteorological training course lecture series. ecmwf. *Reading, UK*

Camporeale, E. (2019). The challenge of machine learning in space weather: Nowcasting and forecasting. *Space Weather* 17, 1166–1207

Chané, E., Jacobs, C., van der Holst, B., Poedts, S., and Kimpe, D. (2005). On the effect of the initial magnetic polarity and of the background wind on the evolution of CME shocks. *A&A* 432, 331–339. doi:10.1051/0004-6361:20042005

Chané, E., Poedts, S., and van der Holst, B. (2008). On the combination of ACE data with numerical simulations to determine the initial characteristics of a CME. *A&A* 492, L29–L32. doi:10.1051/0004-6361:200811022

Connor, H. K., Zesta, E., Fedrizzi, M., Shi, Y., Raeder, J., Codrescu, M. V., et al. (2016). Modeling the ionosphere-thermosphere response to a geomagnetic storm using physics-based magnetospheric energy input: OpenGGCM-CTIM results. *Journal of Space Weather and Space Climate* 6, A25. doi:10.1051/swsc/2016019

Cramer, W. D., Raeder, J., Toffoletto, F., Gilson, M., and Hu, B. (2017). Plasma sheet injections into the inner magnetosphere: Two-way coupled openggcm-rcm model results. *Journal of Geophysical Research: Space Physics* 122, 5077–5091

Dikpati, M. (2017). Ensemble Kalman Filter Data Assimilation in a Solar Dynamo Model. In *AGU Fall Meeting Abstracts*. vol. 2017, SM14A–05

Dorelli, J. C. (2004). A new look at driven magnetic reconnection at the terrestrial subsolar magnetopause. *Journal of Geophysical Research* 109. doi:10.1029/2004ja010458

Eastwood, J. P., Biffis, E., Hapgood, M. A., Green, L., Bisi, M. M., Bentley, R. D., et al. (2017). The economic impact of space weather: Where do we stand? *Risk Analysis* 37, 206–218. doi:10.1111/risa.12765

Echevin, V., De Mey, P., and Evensen, G. (2000). Horizontal and vertical structure of the representer functions for sea surface measurements in a coastal circulation model. *Journal of physical oceanography* 30, 2627–2635

[Dataset] Egbert, G. and Bennett, A. (1996). Data assimilation methods for ocean tides, modern approaches to data assimilation in ocean modeling, edited by: Malonotte-rizzoli, p

Evensen, G. (2009). *Data assimilation: the ensemble Kalman filter* (Springer Science & Business Media)

Fortin, V., Abaza, M., Anctil, F., and Turcotte, R. (2014). Why should ensemble spread match the rmse of the ensemble mean? *Journal of Hydrometeorology* 15, 1708–1713

Ge, Y. S., Raeder, J., Angelopoulos, V., Gilson, M. L., and Runov, A. (2011). Interaction of dipolarization fronts within multiple bursty bulk flows in global MHD simulations of a substorm on 27 February 2009. *J. Geophys. Res.* 116, A00I23

Ghil, M. and Malanotte-Rizzoli, P. (1991). Data assimilation in meteorology and oceanography. In *Advances in geophysics* (Elsevier), vol. 33. 141–266

Hsu, C. T., Matsuo, T., and Liu, J. Y. (2018). Impact of Assimilating the FORMOSAT-3/COSMIC and FORMOSAT-7/COSMIC-2 RO Data on the Midlatitude and Low-Latitude Ionospheric Specification. *Earth and Space Science* 5, 875–890. doi:10.1029/2018EA000447

Huang, C.-L., Spence, H. E., Lyon, J. G., Toffoletto, F. R., Singer, H. J., and Sazykin, S. (2006). Storm-time configuration of the inner magnetosphere: Lyon-fedder-mobarry mhd code, tsyganenko model, and goes observations. *Journal of Geophysical Research: Space Physics* 111

Innocenti, M. E., Lapenta, G., Vršnak, B., Crespon, F., Skandrani, C., Temmer, M., et al. (2011). Improved forecasts of solar wind parameters using the kalman filter. *Space Weather* 9, 1–15

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of basic Engineering* 82, 35–45

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

Keppens, R. and Goedbloed, J. P. (2000). Stellar Winds, Dead Zones, and Coronal Mass Ejections. *ApJ* 530, 1036–1048. doi:10.1086/308395

King, J. and Papitashvili, N. (2005). Solar wind spatial scales in and comparisons of hourly wind and ace plasma and magnetic field data. *Journal of Geophysical Research: Space Physics* 110

Kondrashov, D., Shprits, Y., Ghil, M., and Thorne, R. (2007). A kalman filter technique to estimate relativistic electron lifetimes in the outer radiation belt. *Journal of Geophysical Research: Space Physics* 112

Lang, M., Owens, M., and Lawless, A. (2020). Improving solar wind forecasts using data assimilation. In *EGU General Assembly Conference Abstracts*. EGU General Assembly Conference Abstracts, 10909

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

Lang, M. and Owens, M. J. (2019). A variational approach to data assimilation in the solar wind. *Space Weather* 17, 59–83

Laperre, B., Amaya, J., and Lapenta, G. (2020). Dynamic time warping as a new evaluation for dst forecast with machine learning. *arXiv preprint arXiv:2006.04667*

Lavraud, B., Liu, Y., Segura, K., He, J., Qin, G., Temmer, M., et al. (2016). A small mission concept to the sun–earth lagrangian 15 point for innovative solar, heliospheric and space weather science. *Journal of Atmospheric and Solar-Terrestrial Physics* 146, 171 – 185. doi:<https://doi.org/10.1016/j.jastp.2016.06.004>

Le Dimet, F.-X. and Talagrand, O. (1986). Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus A: Dynamic Meteorology and Oceanography* 38, 97–110

Luhmann, J. G., Solomon, S. C., Linker, J. A., Lyon, J. G., Mikic, Z., Odstrcil, D., et al. (2004). Coupled model simulation of a sun-to-earth space weather event. *Journal of atmospheric and solar-terrestrial physics* 66, 1243–1256

McComas, D. J., Bame, S. J., Barraclough, B. L., Feldman, W. C., Funsten, H. O., Gosling, J. T., et al. (1998). Ulysses’ return to the slow solar wind. *Geophysical Research Letters* 25, 1–4. doi:[10.1029/97GL03444](https://doi.org/10.1029/97GL03444)

Mendoza, O. B., De Moor, B., and Bernstein, D. (2006). Data assimilation for magnetohydrodynamics systems. *Journal of Computational and Applied Mathematics* 189, 242–259

Mignone, A., Bodo, G., Massaglia, S., Matsakos, T., Tesileanu, O., Zanni, C., et al. (2007). PLUTO: A Numerical Code for Computational Astrophysics. *ApJS* 170, 228–242. doi:[10.1086/513316](https://doi.org/10.1086/513316)

Mignone, A., Zanni, C., Tzeferacos, P., van Straalen, B., Colella, P., and Bodo, G. (2012). The PLUTO Code for Adaptive Mesh Computations in Astrophysical Fluid Dynamics. *ApJS* 198, 7. doi:[10.1088/0067-0049/198/1/7](https://doi.org/10.1088/0067-0049/198/1/7)

Moretto, T., Vennerstrom, S., Olsen, N., Rastaetter, L., and Raeder, J. (2006). Using global magnetospheric models for simulation and interpretation of SWARM external field measurements. *Earth, Planets, Space* 58, 439–449

Nikolić, L. (2019). On solutions of the pfss model with gong synoptic maps for 2006–2018. *Space Weather* 17, 1293–1311. doi:[10.1029/2019SW002205](https://doi.org/10.1029/2019SW002205)

Owens, M., Lang, M., Barnard, L., Riley, P., Ben-Nun, M., Scott, C. J., et al. (2020). A Computationally Efficient, Time-Dependent Model of the Solar Wind for Use as a Surrogate to Three-Dimensional Numerical Magnetohydrodynamic Simulations. *Sol. Phys.* 295, 43. doi:[10.1007/s11207-020-01605-3](https://doi.org/10.1007/s11207-020-01605-3)

Pevtsov, A. A., Petrie, G., MacNeice, P., and Virtanen, I. I. (2020). Effect of additional magnetograph observations from different lagrangian points in sun-earth system on predicted properties of quasi-steady solar wind at 1 au. *Space Weather* 18, e2020SW002448. doi:[10.1029/2020SW002448](https://doi.org/10.1029/2020SW002448)

Plunkett, S. (2005). *The extreme solar storms of October to November 2003*. Tech. rep., NAVAL RESEARCH LAB WASHINGTON DC SPACE SCIENCE DIV

Pulkkinen, A., Lindahl, S., Viljanen, A., and Pirjola, R. (2005). Geomagnetic storm of 29–31 october 2003: Geomagnetically induced currents and their relation to problems in the swedish high-voltage power transmission system. *Space Weather* 3

Raeder, J. (2003). Global Magnetohydrodynamics – A Tutorial. In *Space Plasma Simulation*, eds. J. Büchner, C. T. Dum, and M. Scholer (Berlin Heidelberg New York: Springer Verlag). doi:[10.1007/3-540-36530-3_11](https://doi.org/10.1007/3-540-36530-3_11)

Raeder, J. (2006). Flux transfer events: 1. generation mechanism for strong southward IMF. *Annales Geophysicae* 24, 381–392. doi:[10.5194/angeo-24-381-2006](https://doi.org/10.5194/angeo-24-381-2006)

Raeder, J. and Angelopoulos, V. (1998). Using global simulations of the magnetosphere for multi-satellite mission planning and analysis,. In *Science Closure and Enabling Technologies for Constellation Class Missions*, eds. V. Angelopoulos and P. V. Panetta (University of California, Berkeley, and NASA Goddard Space Flight Center). 78

Raeder, J., Cramer, W. D., Germaschewski, K., and Jensen, J. (2017). Using OpenGGCM to Compute and Separate Magnetosphere Magnetic Perturbations Measured on Board Low Earth Orbiting Satellites. *Space Sci. Rev.* 206, 601–620. doi:10.1007/s11214-016-0304-x

Raeder, J. and Lu, G. (2005). Polar cap potential saturation during large geomagnetic storms. *Advances in Space Research* 36, 1804–1808. doi:10.1016/j.asr.2004.05.010

Raeder, J., McPherron, R. L., Frank, L. A., Paterson, W. R., Sigwarth, J. B., Lu, G., et al. (2001a). Global simulation of the geospace environment modeling substorm challenge event. *J. Geophys. Res.* 106, 381

Raeder, J., Wang, Y. L., Fuller-Rowell, T. J., and Singer, H. J. (2001b). Global simulation of space weather effects of the Bastille Day storm. *Sol. Phys.* 204, 325

Raeder, J., Zhu, P., Ge, Y., and Siscoe, G. L. (2010). OpenGGCM simulation of a substorm: Axial tail instability and ballooning mode preceding substorm onset. *J. Geophys. Res.* 115, A00116

Richardson, I. G. and Cane, H. V. (2010). Near-earth interplanetary coronal mass ejections during solar cycle 23 (1996–2009): Catalog and summary of properties. *Solar Physics* 264, 189–237

Rigler, E., Baker, D., Weigel, R., Vassiliadis, D., and Klimas, A. (2004). Adaptive linear prediction of radiation belt electrons using the kalman filter. *Space Weather* 2, 1–9

Sakov, P. and Bertino, L. (2011). Relation between two common localisation methods for the enkf. *Computational Geosciences* 15, 225–237

Schatten, K. H. (1971). Current sheet magnetic model for the solar corona

Schatten, K. H., Wilcox, J. M., and Ness, N. F. (1969). A model of interplanetary and coronal magnetic fields. *Solar Physics* 6, 442–455

Schrijver, C. J. and DeRosa, M. L. (2003). Photospheric and heliospheric magnetic fields. *Solar Physics* 212, 165–200

Schunk, R. W., Scherliess, L., Sojka, J. J., Thompson, D. C., Anderson, D. N., Codrescu, M., et al. (2004). Global assimilation of ionospheric measurements (gaim). *Radio Science* 39

Shi, Q. Q., Hartinger, M., Angelopoulos, V., Tian, A., Fu, S., Zong, Q.-G., et al. (2014). Solar wind pressure pulse-driven magnetospheric vortices and their global consequences. *Journal of Geophysical Research: Space Physics* 119, 4274–4280. doi:10.1002/2013ja019551

Siscoe, G., Baker, D., Weigel, R., Hughes, J., and Spence, H. (2004). Roles of empirical modeling within cism. *Journal of atmospheric and solar-terrestrial physics* 66, 1481–1489

Skandran, C., Innocenti, M. E., Bettarini, L., Crespon, F., Lamouroux, J., and Lapenta, G. (2014). Flip-mhd-based model sensitivity analysis. *Nonlinear processes in geophysics* 21, 539–553

Stone, E. C., Frandsen, A., Mewaldt, R., Christian, E., Margolies, D., Ormes, J., et al. (1998). The advanced composition explorer. *Space Science Reviews* 86, 1–22

Thomsen, M., McComas, D., Reeves, G., and Weiss, L. (1996). An observational test of the tsyganenko (t89a) model of the magnetospheric field. *Journal of Geophysical Research: Space Physics* 101, 24827–24836

Toffoletto, F., Spiro, R., Wolf, R., Birn, J., and Hesse, M. (2001). Modeling inner magnetospheric electrodynamics. *Washington DC American Geophysical Union Geophysical Monograph Series* 125, 265–272

Tóth, G., Sokolov, I. V., Gombosi, T. I., Chesney, D. R., Clauer, C. R., De Zeeuw, D. L., et al. (2005). Space weather modeling framework: A new tool for the space science community. *Journal of Geophysical Research: Space Physics* 110

Tsurutani, B. T., Gonzalez, W. D., Gonzalez, A. L., Guarnieri, F. L., Gopalswamy, N., Grande, M., et al. (2006). Corotating solar wind streams and recurrent geomagnetic activity: A review. *Journal of Geophysical Research: Space Physics* 111

Tsyganenko, N. (1987). Global quantitative models of the geomagnetic field in the cislunar magnetosphere for different disturbance levels. *Planetary and space science* 35, 1347–1358

Tsyganenko, N. (1989). A magnetospheric magnetic field model with a warped tail current sheet. *Planetary and Space Science* 37, 5–20

Tsyganenko, N. (2002a). A model of the near magnetosphere with a dawn-dusk asymmetry 1. mathematical structure. *Journal of Geophysical Research: Space Physics* 107, SMP-12

Tsyganenko, N. (2002b). A model of the near magnetosphere with a dawn-dusk asymmetry 2. parameterization and fitting to observations. *Journal of Geophysical Research: Space Physics* 107, SMP-10

Tsyganenko, N. and Andreeva, V. (2015). A forecasting model of the magnetosphere driven by an optimal solar wind coupling function. *Journal of Geophysical Research: Space Physics* 120, 8401–8425

Tsyganenko, N., Singer, H., and Kasper, J. (2003). Storm-time distortion of the inner magnetosphere: How severe can it get? *Journal of Geophysical Research: Space Physics* 108

Tsyganenko, N. and Sitnov, M. (2005). Modeling the dynamics of the inner magnetosphere during strong geomagnetic storms. *Journal of Geophysical Research: Space Physics* 110

Tsyganenko, N. A. (1995). Modeling the earth's magnetospheric magnetic field confined within a realistic magnetopause. *Journal of Geophysical Research: Space Physics* 100, 5599–5612

Tsyganenko, N. A. (1996). Effects of the solar wind conditions in the global magnetospheric configurations as deduced from data-based field models. In *International conference on substorms*. vol. 389, 181

Vennerstrom, S., Moretto, T., Rastaetter, L., and Raeder, J. (2005). Field-aligned currents during northward interplanetary field: Morphology and causes. *J. Geophys. Res.* 110, A06205, doi: 10.1029/2004JA010802

Vourlidas, A. (2015). Mission to the sun-earth 15 lagrangian point: An optimal platform for space weather research. *Space Weather* 13, 197–201. doi:10.1002/2015SW001173

Wang, Y.-M. and Sheeley Jr, N. (1992). On potential field models of the solar corona. *The Astrophysical Journal* 392, 310–319

Woodfield, E., Dunlop, M., Holme, R., Davies, J., and Hapgood, M. (2007). A comparison of cluster magnetic data with the tsyganenko 2001 model. *Journal of Geophysical Research: Space Physics* 112

Zhou, X.-Z., Ge, Y. S., Angelopoulos, V., Runov, A., Liang, J., Xing, X., et al. (2012). Dipolarization fronts and associated auroral activities: 2. acceleration of ions and their subsequent behavior. *Journal of Geophysical Research: Space Physics* 117, 1. doi:10.1029/2012ja017677

Zhu, P., Raeder, J., Germaschewski, K., and Hegna, C. C. (2009). Initiation of ballooning instability in the near-Earth plasma sheet prior to the 23 march 2007 THEMIS substorm expansion onset. *Annales Geophysicae* 27, 1129–1138. doi:10.5194/angeo-27-1129-2009