Application of data assimilation techniques to heliospheric modelling: two preliminary studies

M.E. Innocenti[1], G. Lapenta[1], B. Vrsnak[2], F. Crespon[3], C. Skandrani[3], T. Rotter[4], M. Temmer[4], A. Veronig[4], L. Bettarini[5], S. Markidis[1], M. Skender[2]

1: KU Leuven, Belgium
2: Hvar Observatory, Croatia
3: Noveltis, France
4: University of Graz, Austria
5: Royal Observatory of Belgium

Solar Storm Early Forecasting - "SolarCast-1",
Copenaghen, Denmark
03/12/2015

Outline

1. Data Assimilation (DA) techniques: Kalman filtering
2. Application of Kalman filtering to an empirical solar wind forecasting model
3. Application of ensemble method techniques to an MHD model for background solar wind propagation
4. Conclusions

1. Data Assimilation

- consists in steering the evolution of a model through comparison with observations [Welch & Bishop, 2001; Kalman, 1960]

- is commonplace in domains where observations are available at multiple locations in the system, i.e. ionosphere [Schunk et al., 2003], electrons dynamics in the radiation belts [Kondrashov et al., 2007; Rigler et al., 2004], atmospheric and oceanic studies [Ghil and Malanotte−Rizzoli, 1991], …

- is rarely applied to the heliosphere for relative lack of observations; [Schrijver & DeRosa, 2003] is an exception

1. Kalman filter algorithms for system identification purposes

the state x:

time index

state noise

state noise covariance

$$x_k \in \Re^n \qquad x_k = x_{k-1} + w_{k-1} \qquad p(w) \approx N(0, Q)$$

state value, unknown

system identification: time evolution matrix A=I

state noise distribution

the observations z:

observations

measurement noise

measurement noise covariance

$$z_k \in \Re^m \qquad z_k = H x_k + \nu_k \qquad p(\nu) \approx N(0, R)$$

relation between observations z and state x

measurement noise distribution

1. Kalman filter algorithms for system identification purposes

prediction phase

a priori state estimate

$$\hat{x}_k^- = \hat{x}_{k-1}$$

a posteriori state estimate

a priori error covariance

$$P_k^- = P_{k-1} + Q$$

a posteriori error covariance

correction phase

Kalman gain

$$K_k = P_k^- H^T \left(H P_k^- H^T + R\right)^{-1}$$

"innovation": difference between observations and 'pseudo-observations' generated with the a priori estimate

$$\hat{x}_k = \hat{x}_k^- + K_k \left(z_k - H\hat{x}_k^-\right)$$

$$P_k = (I - K_k H)\, P_k^-$$

if A!=I (the method is not used for system identification, the state evolve in time) the prediction phase includes time evolution; the correction phase is unaltered

2. Application of Kalman filtering to a solar wind forecasting model

We applied Kalman filtering techniques to a baseline empirical model for the forecast of solar wind parameters (proton density n, magnetic field magnitude B, proton temperature T, proton velocity v)

The aim is to understand if Data Assimilation can be of benefit to solar wind forecasting models; we started with an extremely simple one and tried to push its limits

[Innocenti et al, 2011]

2.1. The baseline model

An empirical model…

The baseline model exploits the well-known connection between Coronal Holes, High Speed Streams and Solar Wind dynamics

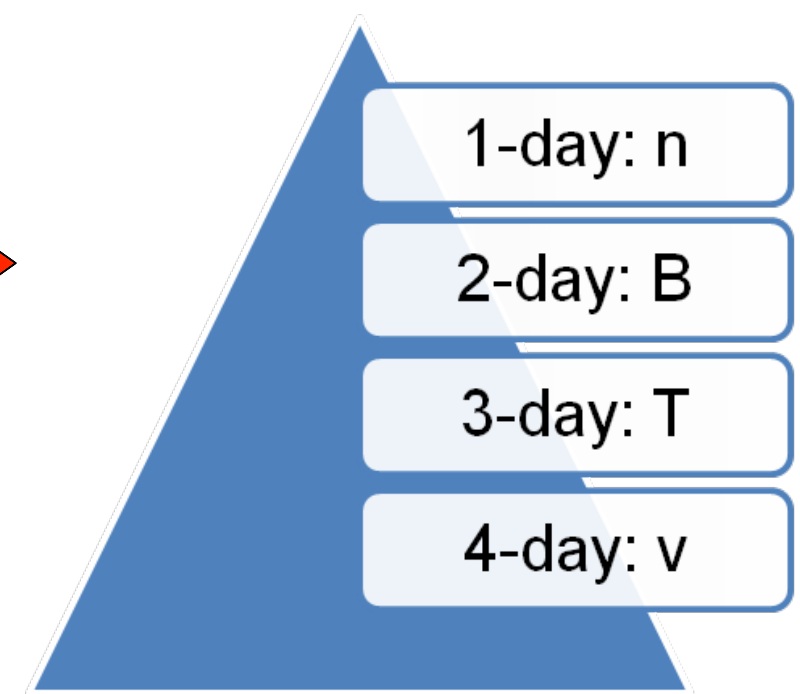
GOES-12 observations of CH area in the meridional slice of the sun

Linear relations with coefficients statistically determined from ACE measurements (LMS analysis of CH coverage and ACE measurements from DOYs 25-125, year 2005)

H

CH observations and SW forecasts are related by time lags which differ with the quantity to forecast

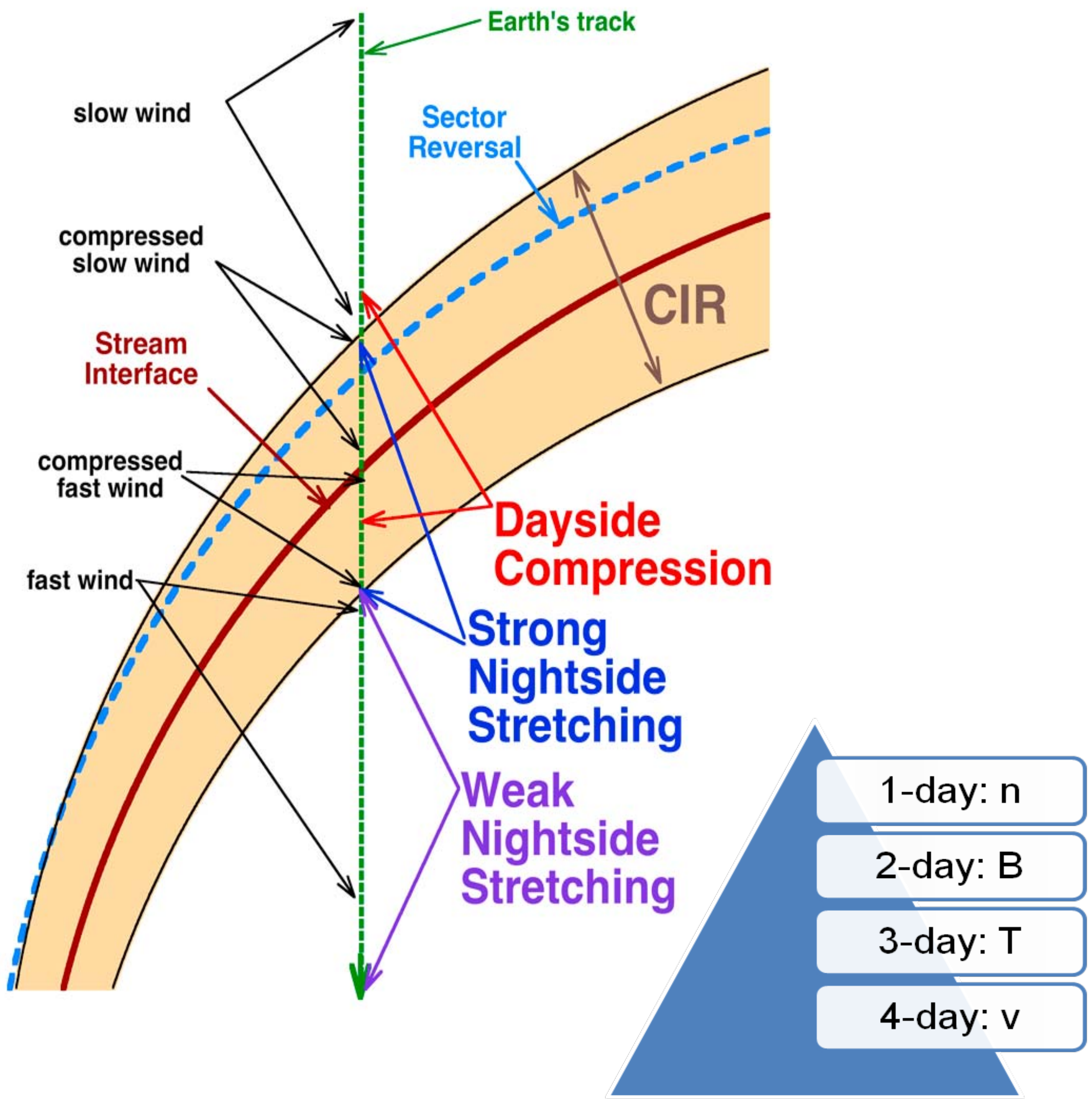
1-day resolution nowcast/ forecast of SW parameters at 1 AU with different lead times



Levine77, Sheeley81, Nolte76, Robbins06, Vrsnak07

2.1. The baseline model

… with a strong physical background



Earth intersecting a Corotating Interaction Region (CIR)

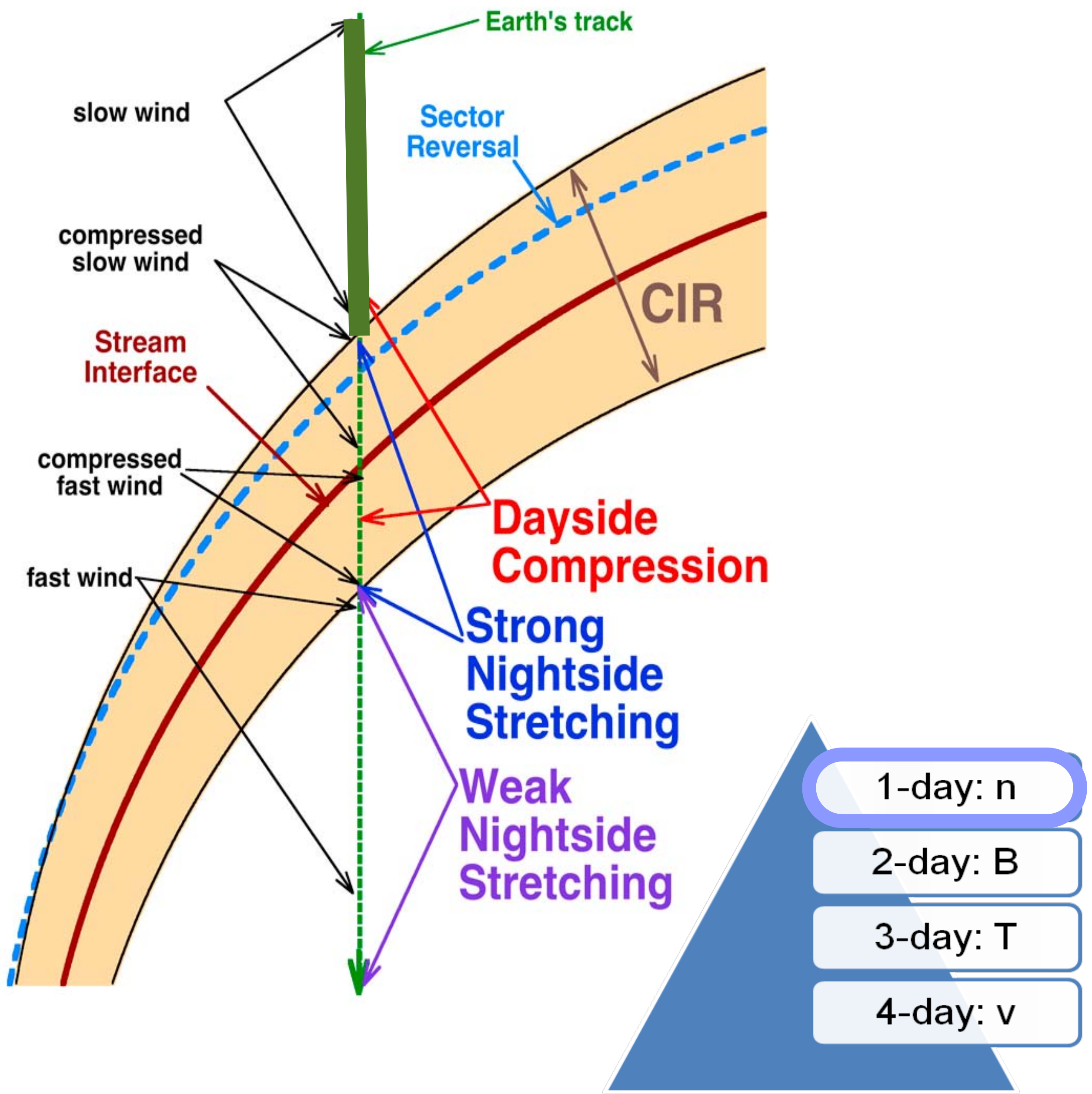
The origin of the time- lagged peaks in the SW quantities is the entrance of the Earth in the area affected by an incoming CIR

The density peak maps to the leading edge of the CH, the velocity peak to the center of the CH (peak in the CH area)

Gosling99, Borovsky10, Tsurutani06

2.1. The baseline model

… with a strong physical background



Earth intersecting a Corotating Interaction Region (CIR)

The origin of the time- lagged peaks in the SW quantities is the entrance of the Earth in the area affected by an incoming CIR

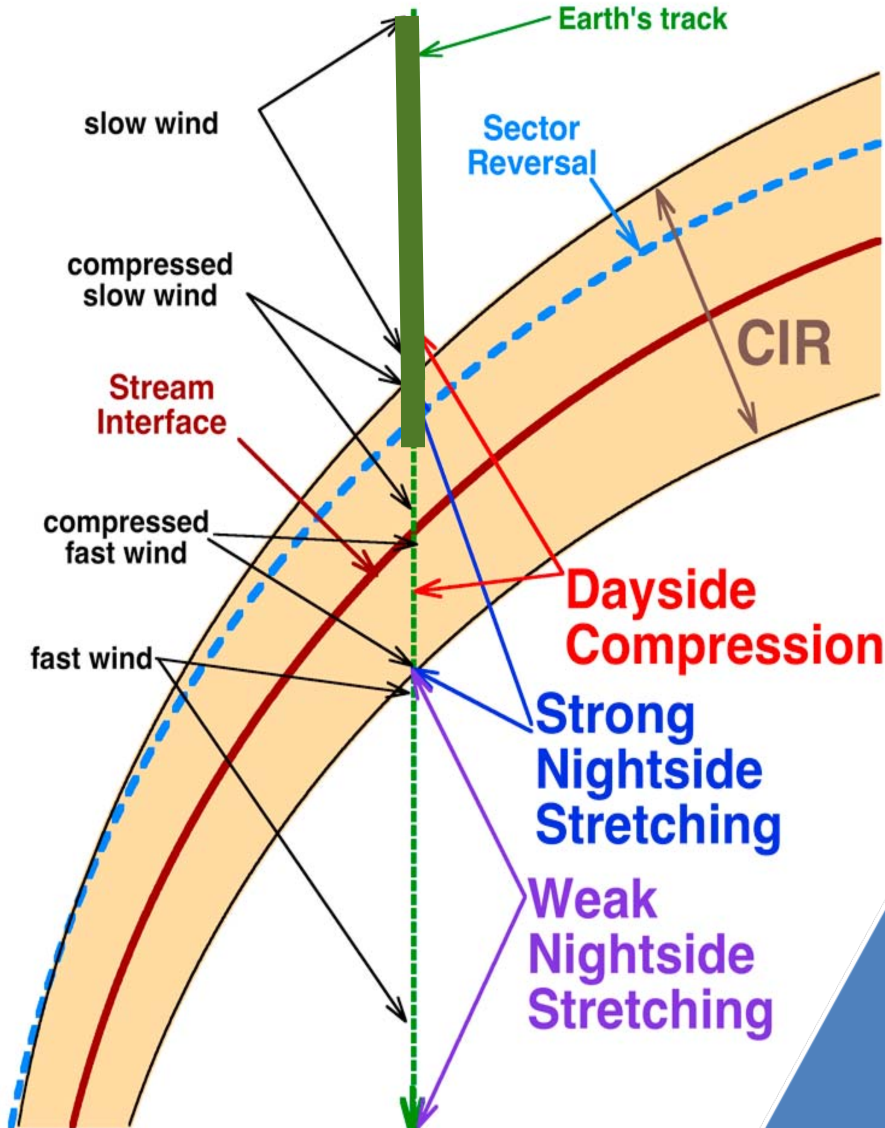
1. the Earth enters the area of compressed slow solar wind ahead of the CIR →peak in n

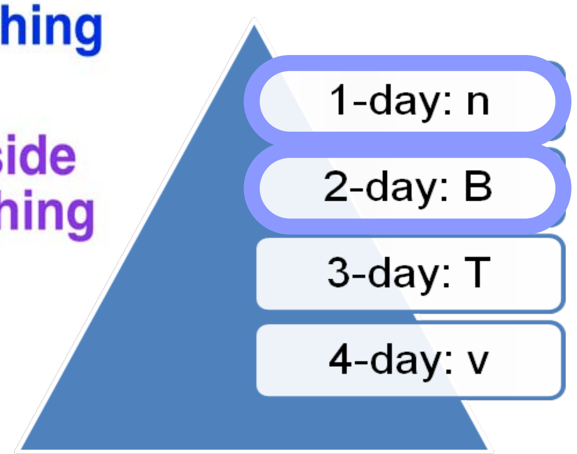
The density peak maps to the leading edge of the CH, the velocity peak to the center of the CH (peak in the CH area)

Gosling99, Borovsky10, Tsurutani06

2.1. The baseline model

… with a strong physical background





Earth intersecting a Corotating Interaction Region (CIR)

The origin of the time- lagged peaks in the SW quantities is the entrance of the Earth in the area affected by an incoming CIR

1. the Earth enters the area of compressed slow solar wind ahead of the CIR →peak in n

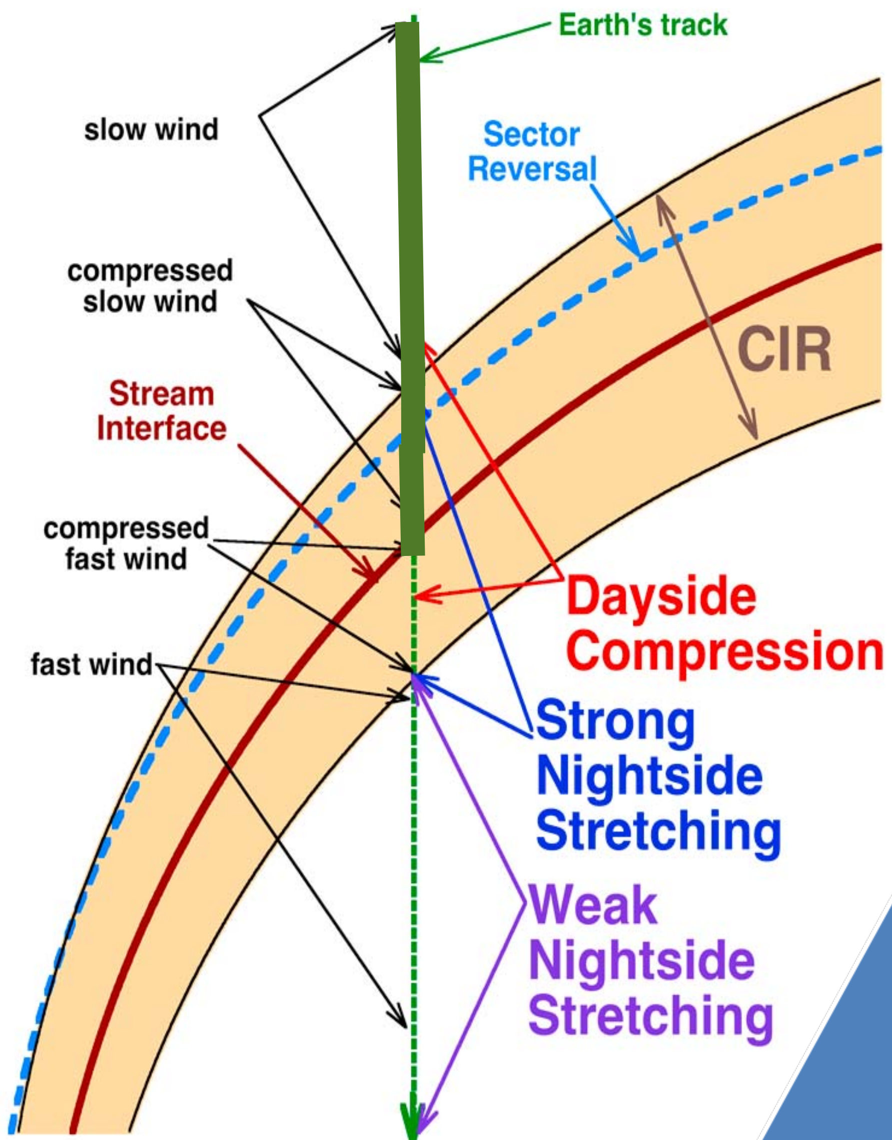
2. magnetospheric field compression around the time of passage of the Interplanetary Magnetic Field (IMF) sector reversal →peak in B

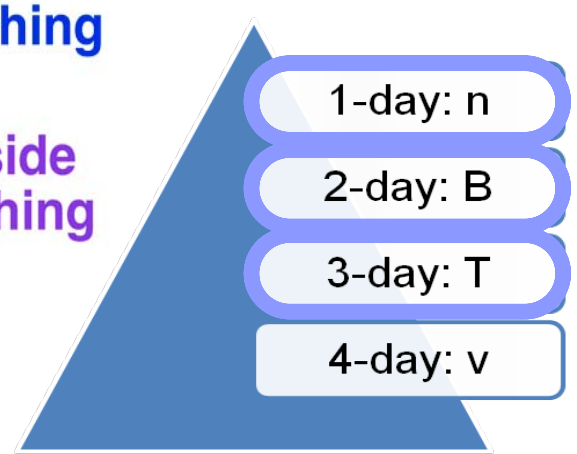
The density peak maps to the leading edge of the CH, the velocity peak to the center of the CH (peak in the CH area)

Gosling99, Borovsky10, Tsurutani06

2.1. The baseline model

… with a strong physical background





Earth intersecting a Corotating Interaction Region (CIR)

The origin of the time- lagged peaks in the SW quantities is the entrance of the Earth in the area affected by an incoming CIR

1. the Earth enters the area of compressed slow solar wind ahead of the CIR →peak in n

2. magnetospheric field compression around the time of passage of the Interplanetary Magnetic Field (IMF) sector reversal →peak in B

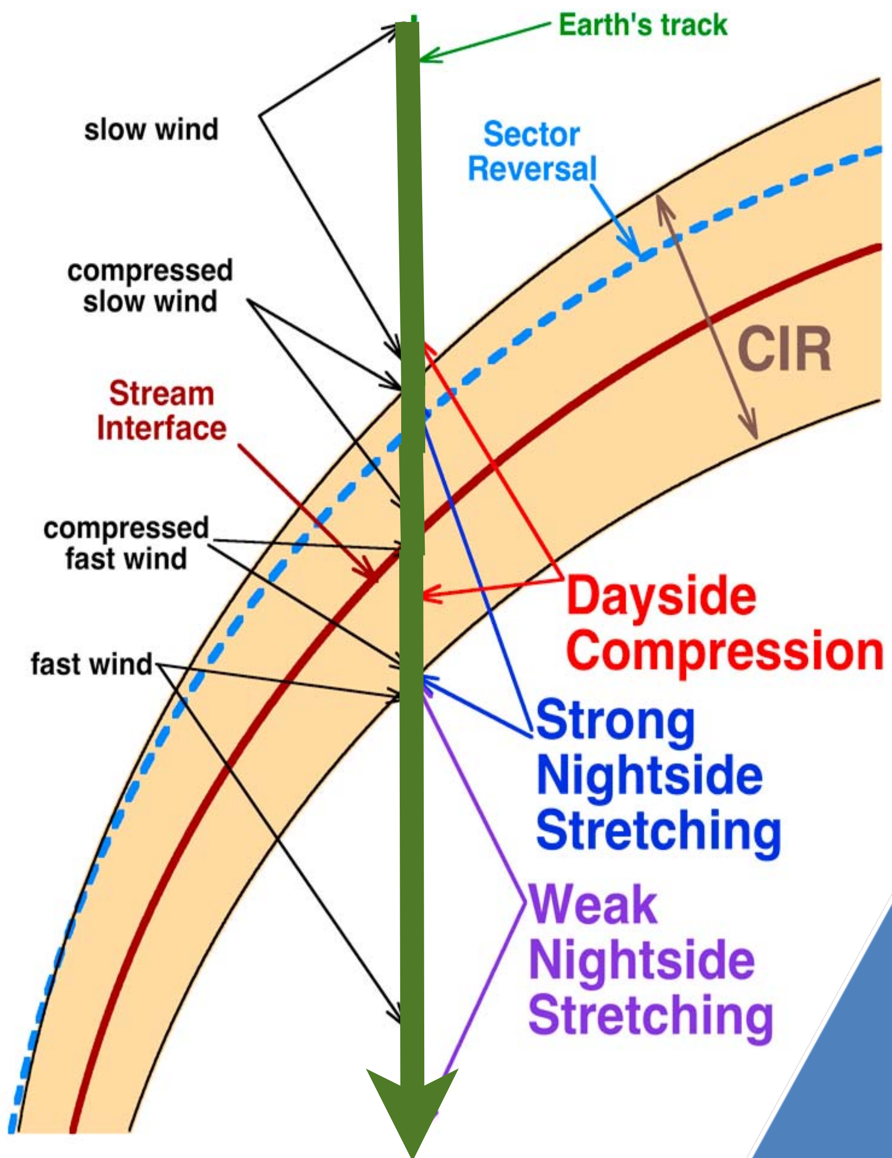
3. crossing of the stream interface →peak in T

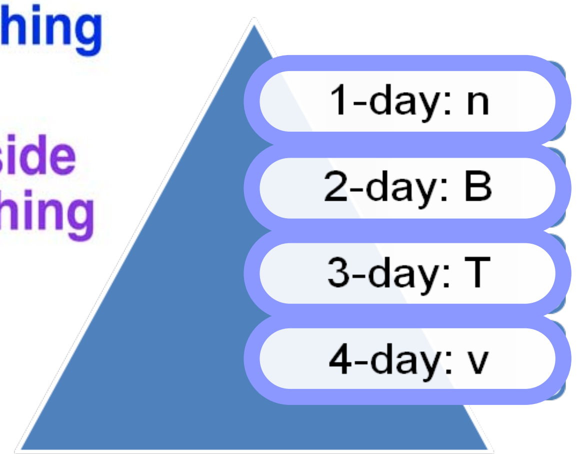
The density peak maps to the leading edge of the CH, the velocity peak to the center of the CH (peak in the CH area)

Gosling99, Borovsky10, Tsurutani06

2.1.The baseline model

… with a strong physical background





Earth intersecting a Corotating Interaction Region (CIR)

The origin of the time- lagged peaks in the SW quantities is the entrance of the Earth in the area affected by an incoming CIR

1. the Earth enters the area of compressed slow solar wind ahead of the CIR →peak in n

2. magnetospheric field compression around the time of passage of the Interplanetary Magnetic Field (IMF) sector reversal →peak in B

3. crossing of the stream interface →peak in T

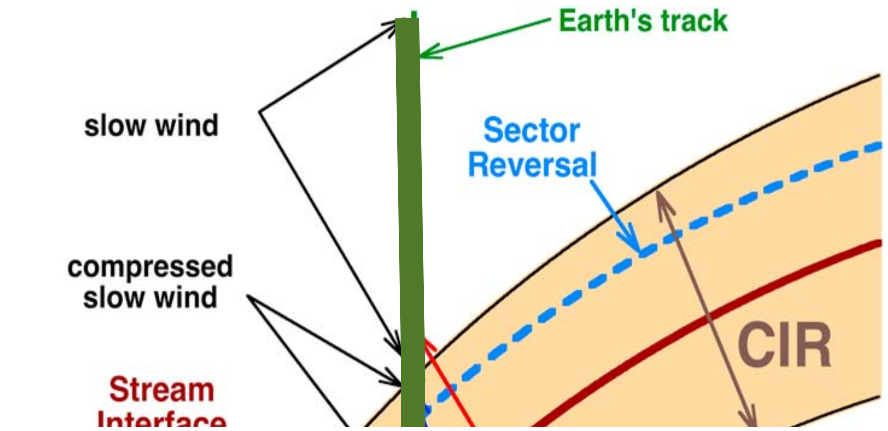
4. the Earth emerges into the fast wind region behind the CIR →peak in v

The density peak maps to the leading edge of the CH, the velocity peak to the center of the CH (peak in the CH area)

Gosling99, Borovsky10, Tsurutani06

2.1. The baseline model

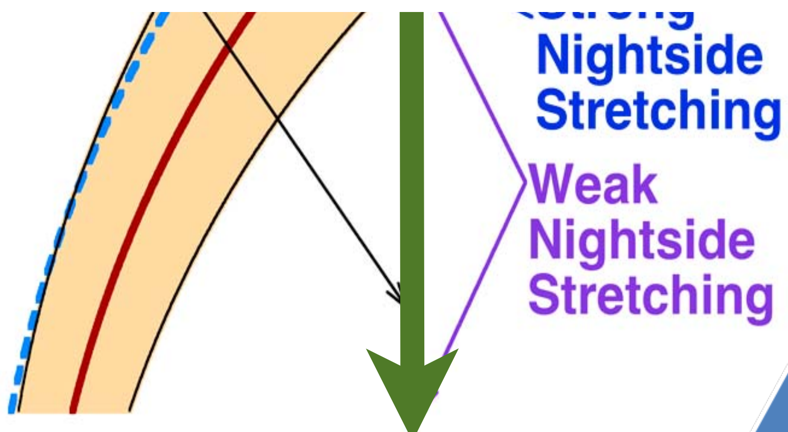
… with a strong physical background

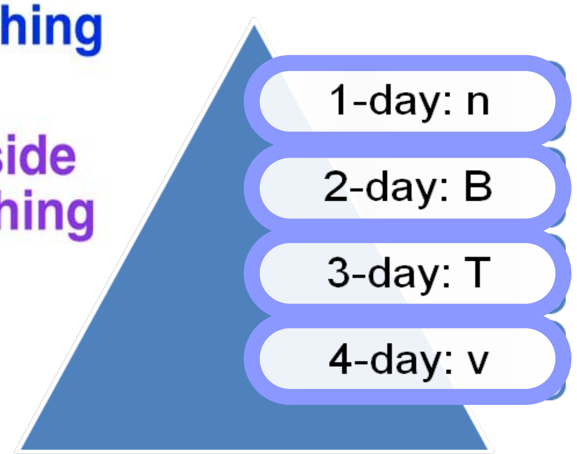


The origin of the time- lagged peaks in the SW quantities is the entrance of the Earth in the area affected by an incoming CIR

1. the Earth enters the area of compressed slow solar wind

the baseline model is physically more relevant during the declining phase of the solar cycle, when HSS/CIR dynamics are the main drivers of geomagnetic activity at Earth





Earth intersecting a Corotating Interaction Region (CIR)

→peak in B

3. crossing of the stream interface →peak in T

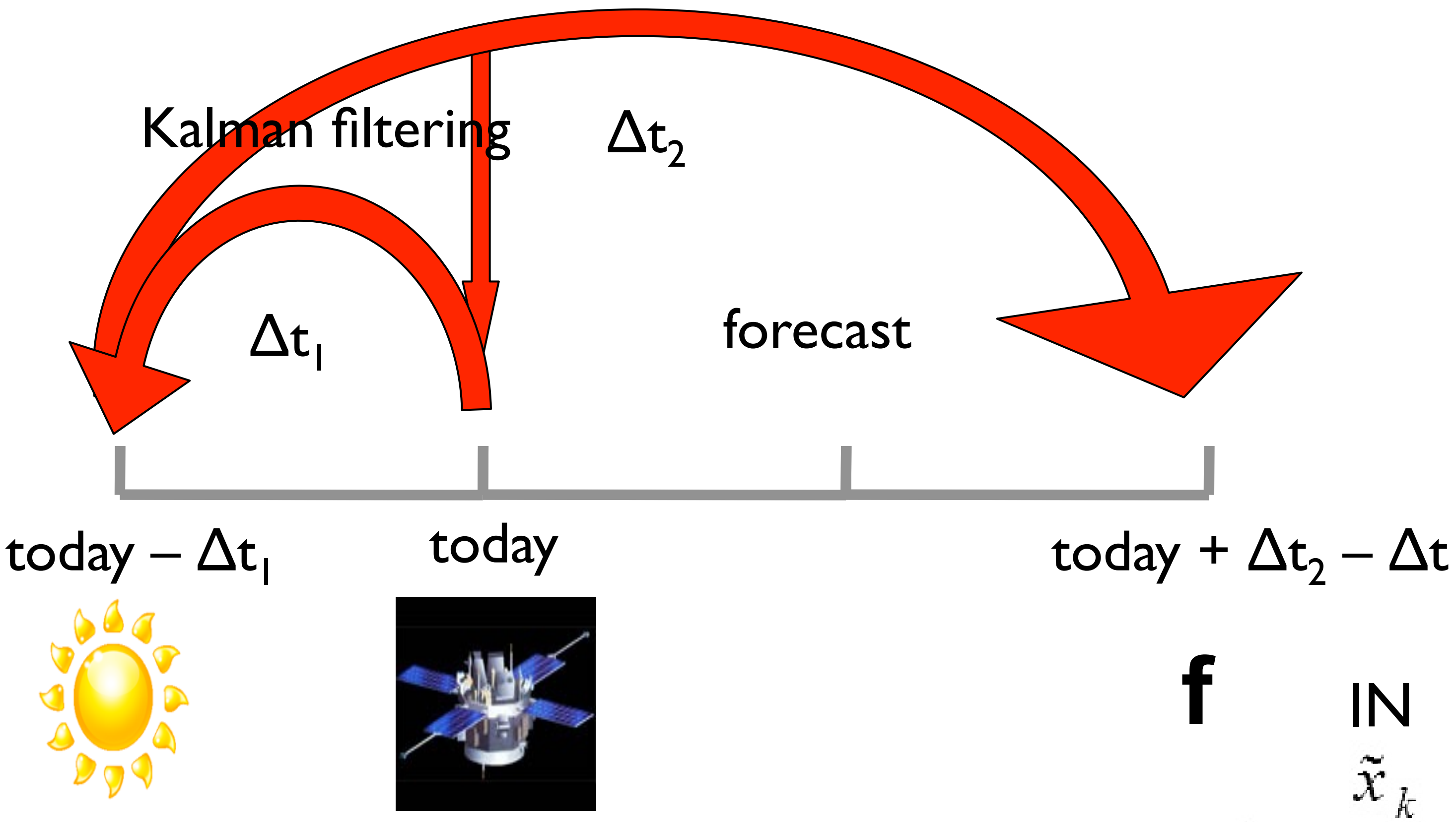
4. the Earth emerges into the fast wind region behind the CIR →peak in v

The density peak maps to the leading edge of the CH, the velocity peak to the center of the CH (peak in the CH area)

Gosling99, Borovsky10, Tsurutani06

2. The assimilation method

Kalkan filtering techniques have been applied in a rather unconventional way to exploit a very convenient characteristic of the baseline model, i.e. the time delayed correlation between the observables

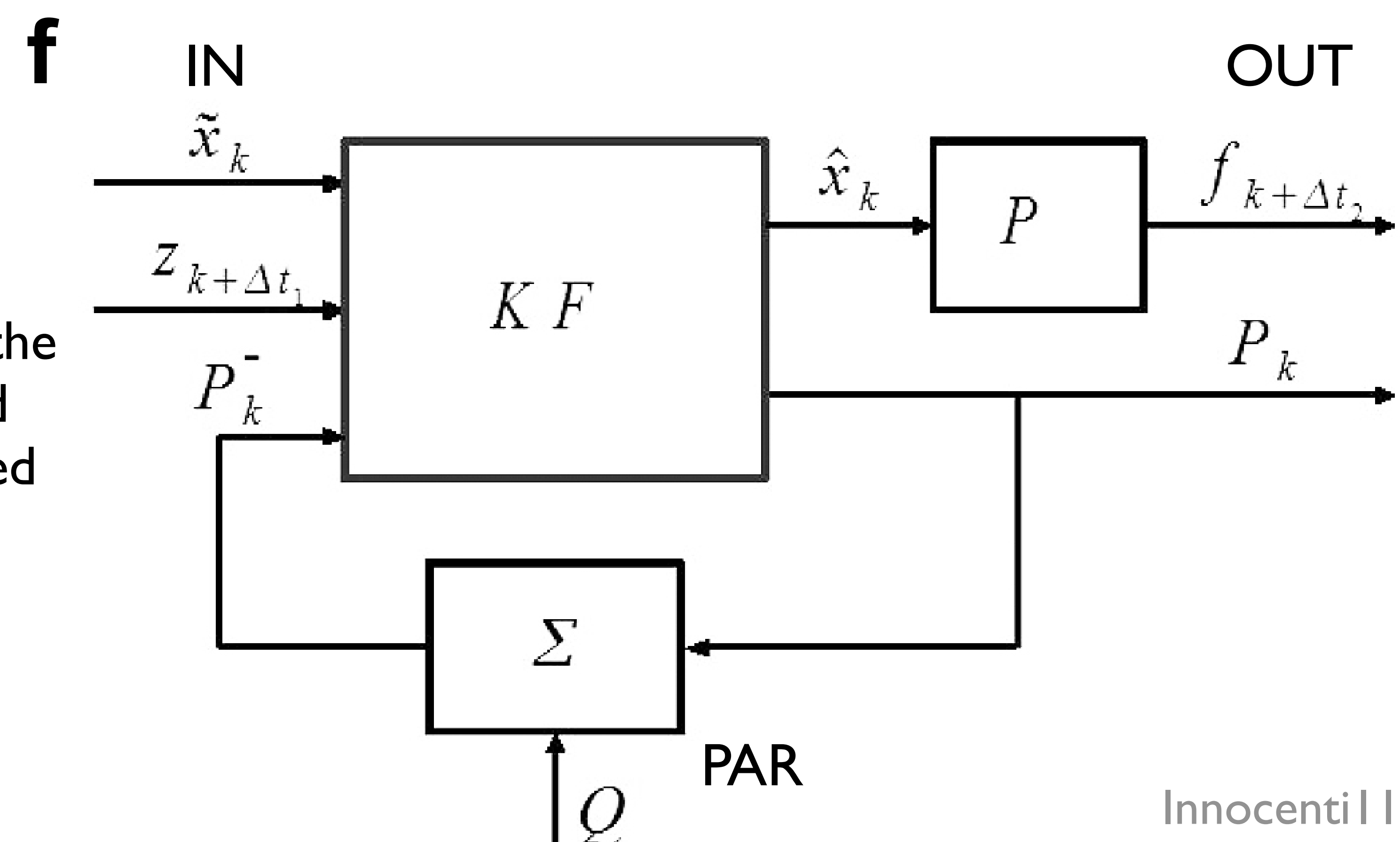


IN: solar observation x_k and time-lagged ACE observation $z_{k+\Delta t1}$

OUT: improved solar observation $\hat{x}_k$, used for forecasting the "later" quantity

PAR: tunable filter parameter: process noise covariance (a measure of the model to measurements confidence ratio)

the different time lags between GOES12 observations and the forecasts allow to correct the fractional CH area with "earlier" quantities and than forecast "later" quantities with the improved state



2. Comparison of forecast quality with alternative forecasting methods

Mean Absolute Error (MAE) for the year 2005 (panel a) and 2006 (panel b) for the forecast of the proton velocity with assimilation of the proton temperature.

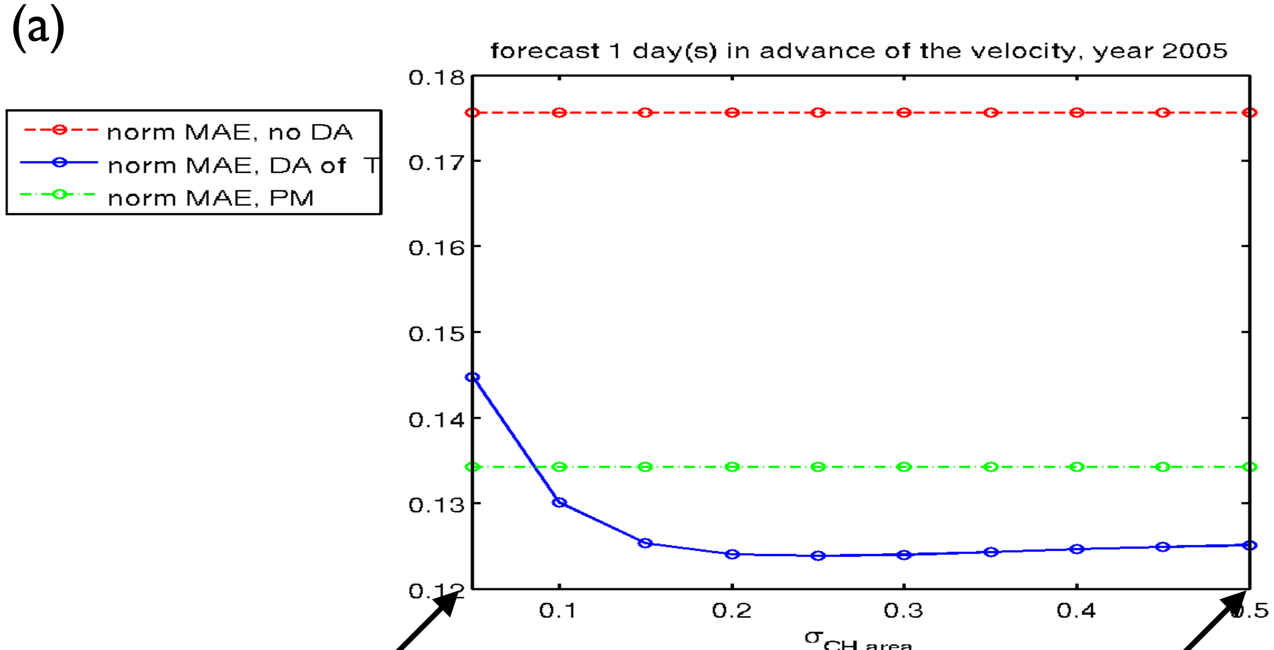
The MAEs are as a function of the tunable filter parameter, the process noise covariance $\sigma_{CH\ area}$

forecast with the baseline method

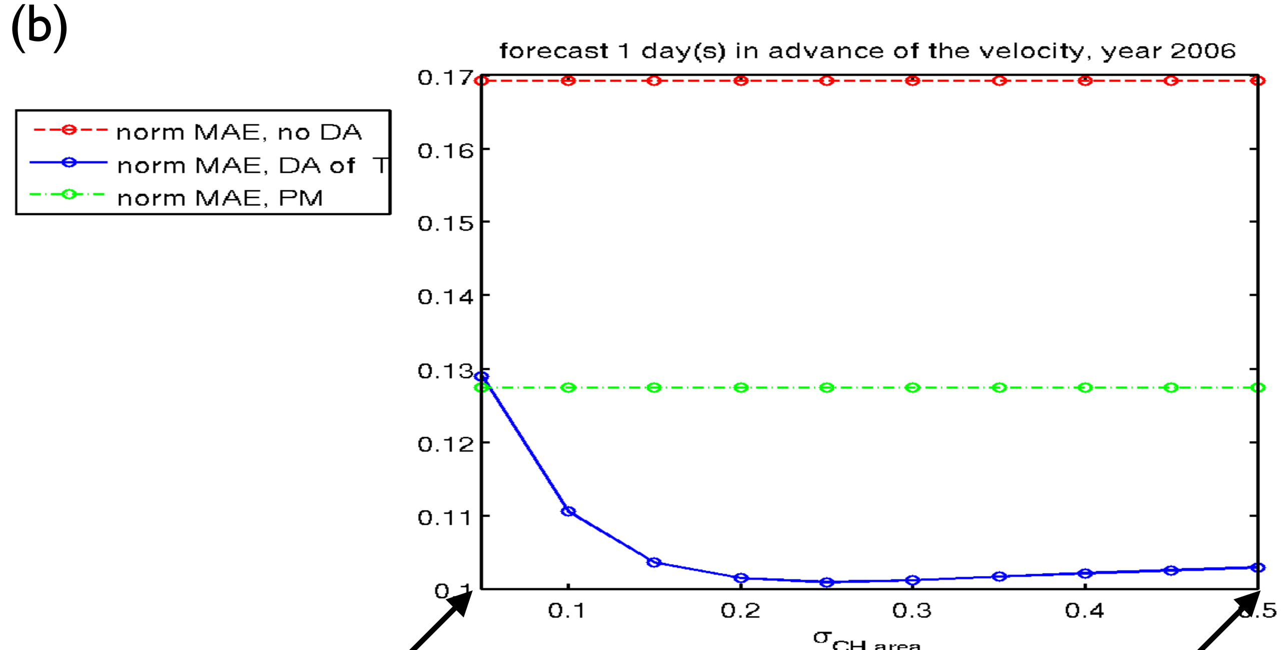
forecast with Data Assimilation

forecast with persistence method; i.e., use the current value as forecast; it often outperforms more "refined" models and is often used as benchmark for new forecasting models, see the CISM website

(a)



(b)



CH observations are considered reliable

CH observations are NOT considered reliable

CH observations are considered reliable

CH observations are NOT considered reliable

The model with DA outperforms both the baseline and the persistence method for years 2005 and 2006, with the baseline model coefficients being calculated for a fraction of year 2005 (performances are evaluated with the Mean Absolute Error - MAE)

2. Performance during periods of geomagnetic activity

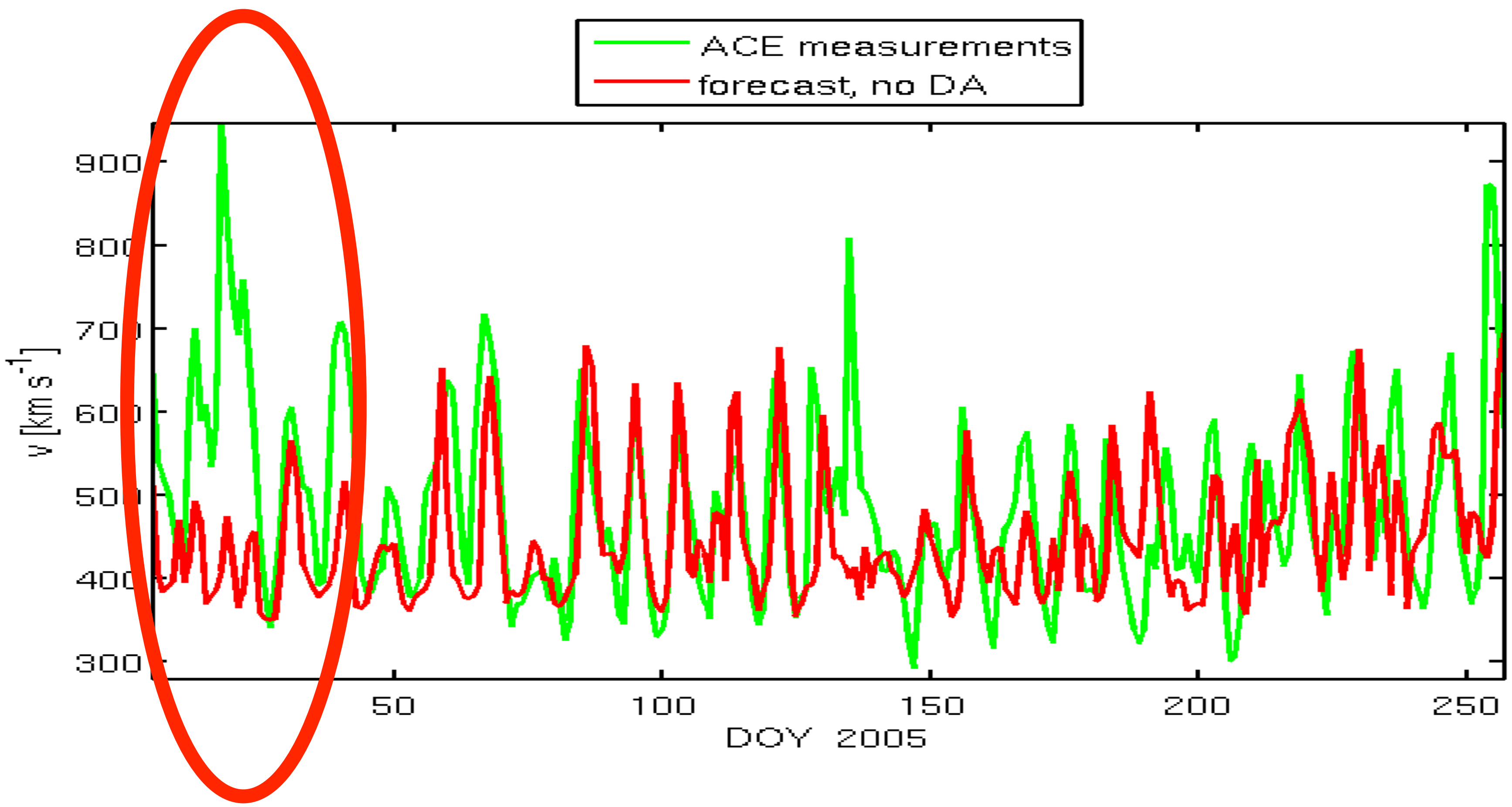
On paper, one of the advantages of DA is to include in a system processes *not included* in the baseline model through assimilation of observation; we tested this with CME activity (the baseline model covers HSS, not CMEs)

year 2005

ACE measurement of proton velocity

forecast of the proton velocity with the baseline model (-4 days)

forecast of the proton velocity with assimilation of proton temperature (-1 day)



The model with DA improves the quality of the forecast during the period of moderate geomagnetic activity on January 7-27 (successive flares, several halo CME emitted, fast halo CME emitted on January 20)

Du08, Foullon07

2. Performance during periods of geomagnetic activity

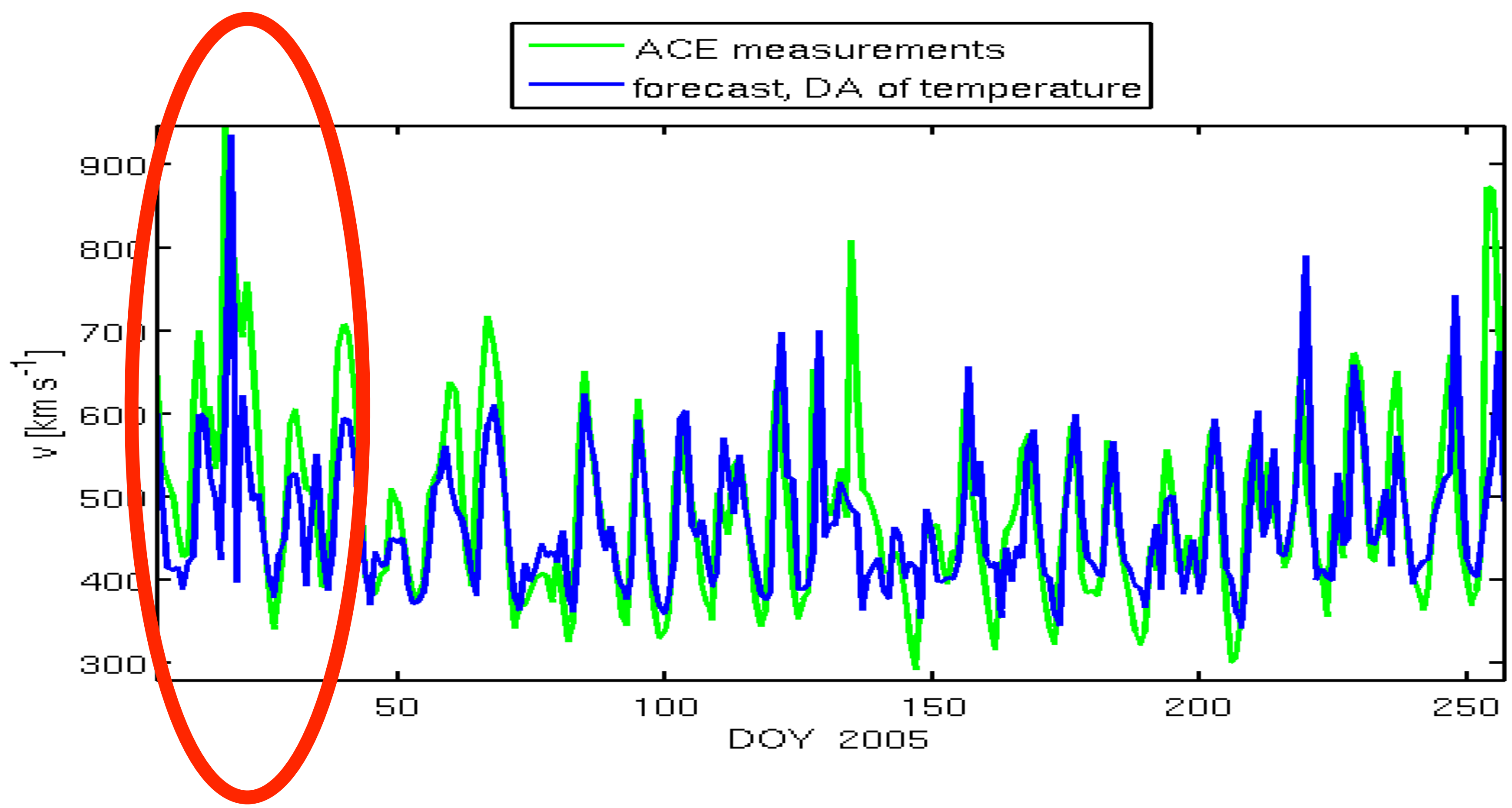
On paper, one of the advantages of DA is to include in a system processes *not included* in the baseline model through assimilation of observation; we tested this with CME activity (the baseline model covers HSS, not CMEs)

year 2005

ACE measurement of proton velocity

forecast of the proton velocity with the baseline model (-4 days)

forecast of the proton velocity with assimilation of proton temperature (-1 day)



The model with DA improves the quality of the forecast during the period of moderate geomagnetic activity on January 7-27 (successive flares, several halo CME emitted, fast halo CME emitted on January 20)

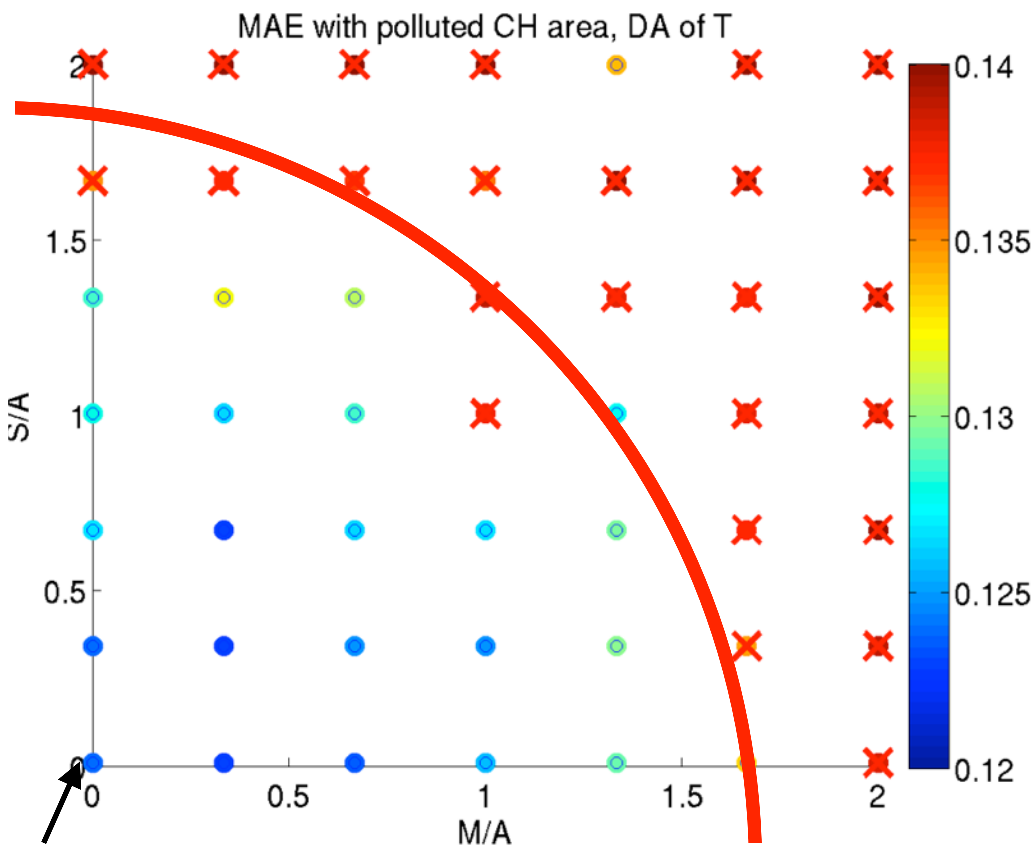
Du08, Foullon07

2. Robustness towards corrupted inputs

We tested the robustness of the model with DA by feeding it with “corrupted inputs”, i.e. superposing a normally distributed noise (with prescript mean and standard deviation) to the observed CH covarage

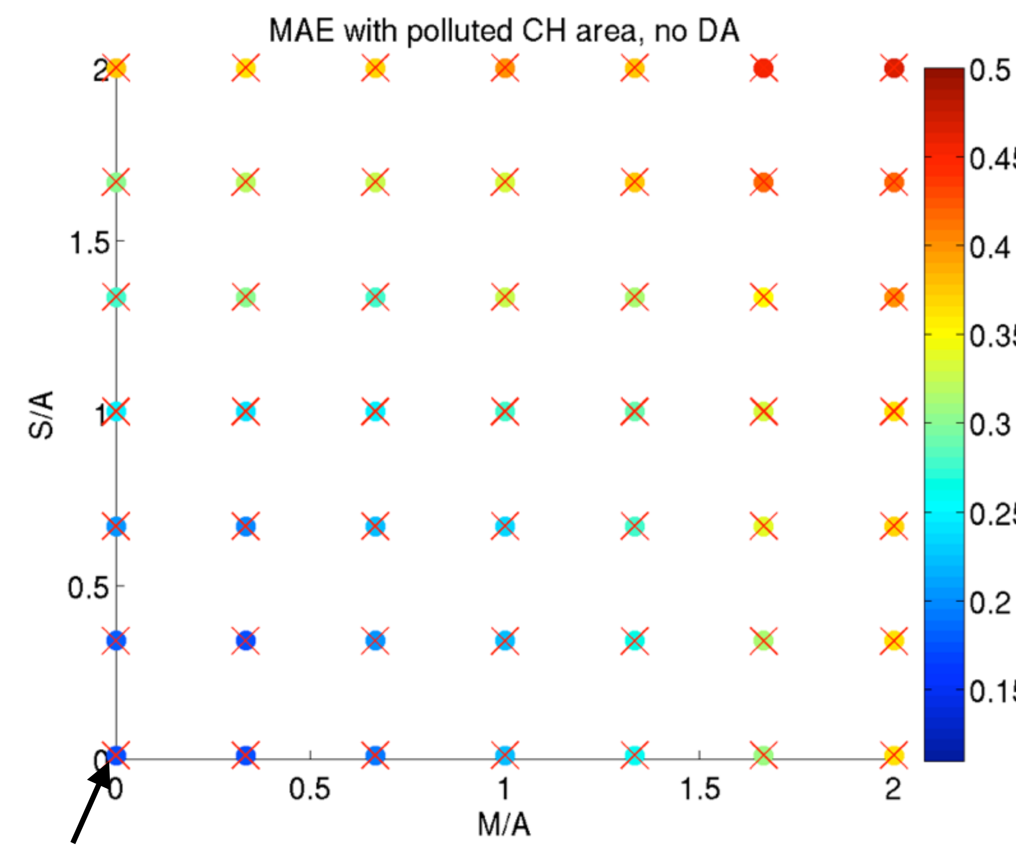
MAE for the forecast of the solar wind speed for the year 2005 with corrupted CH inputs, classified according to the mean and standard deviation of the normally distributed noise superimposed to the input (with respect to original values)

forecast with DA of T



non-corrupted input

forecast with baseline method



non-corrupted input

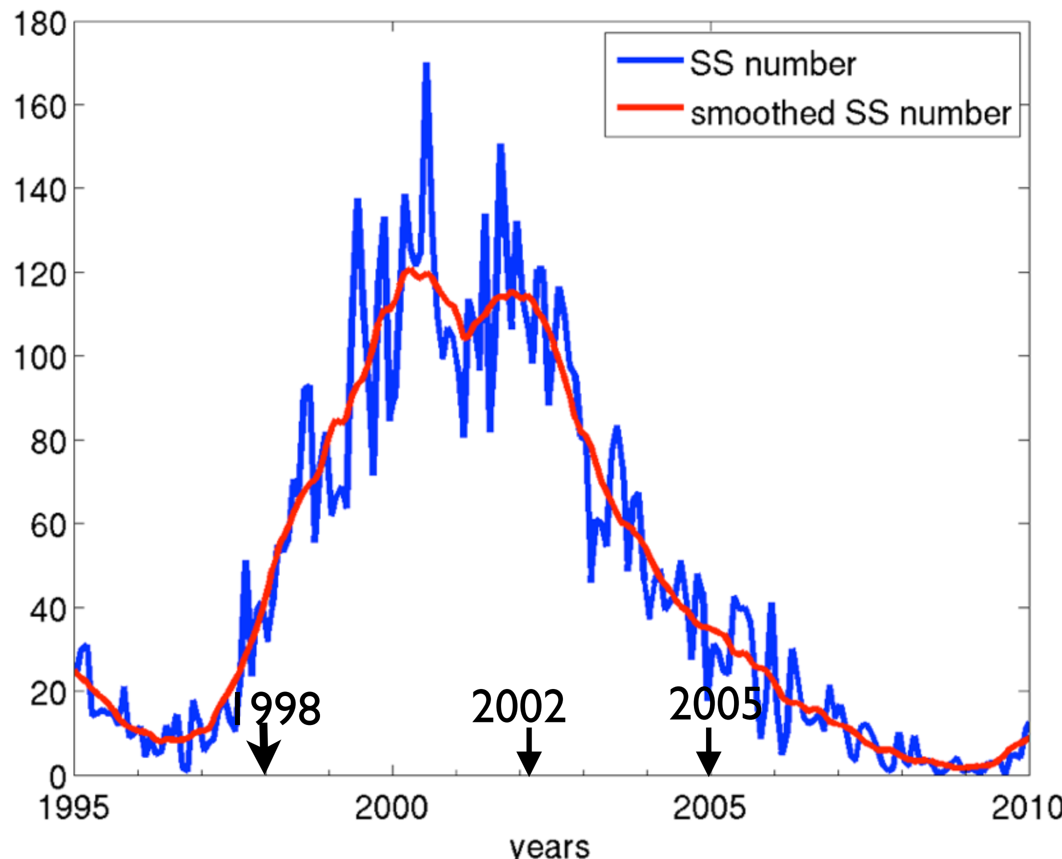
O: the current method with corrupted inputs outperforms the persistence method (PM) with uncorrupted inputs (MAE= 0.1342)

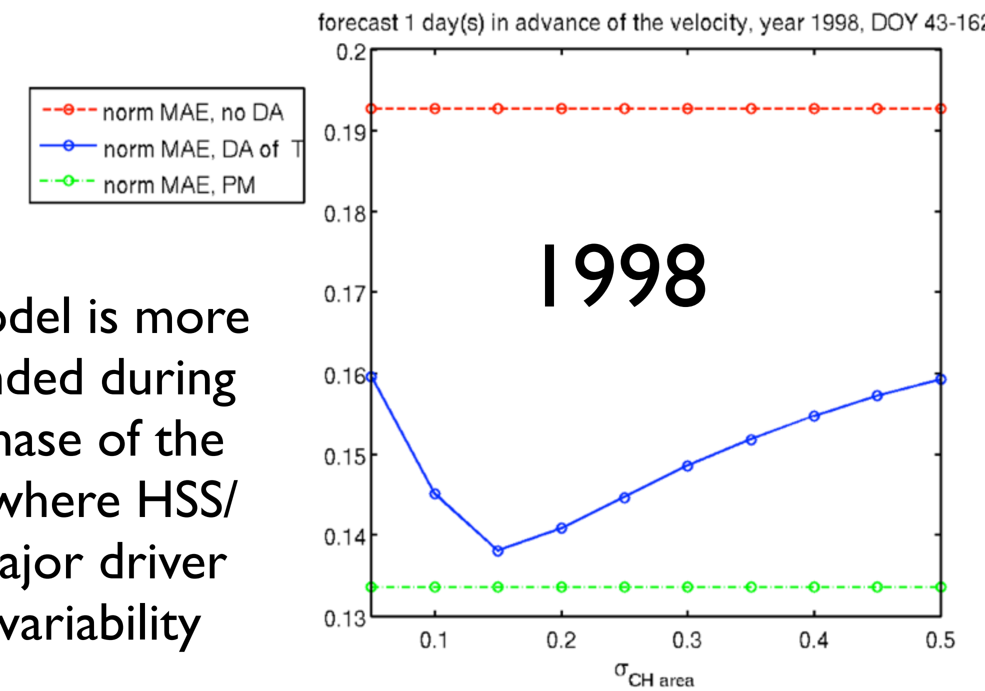
X: the current method with corrupted inputs performs less than the PM

→ the assimilation of T measurements grants a high tolerance of the forecast to input corruption

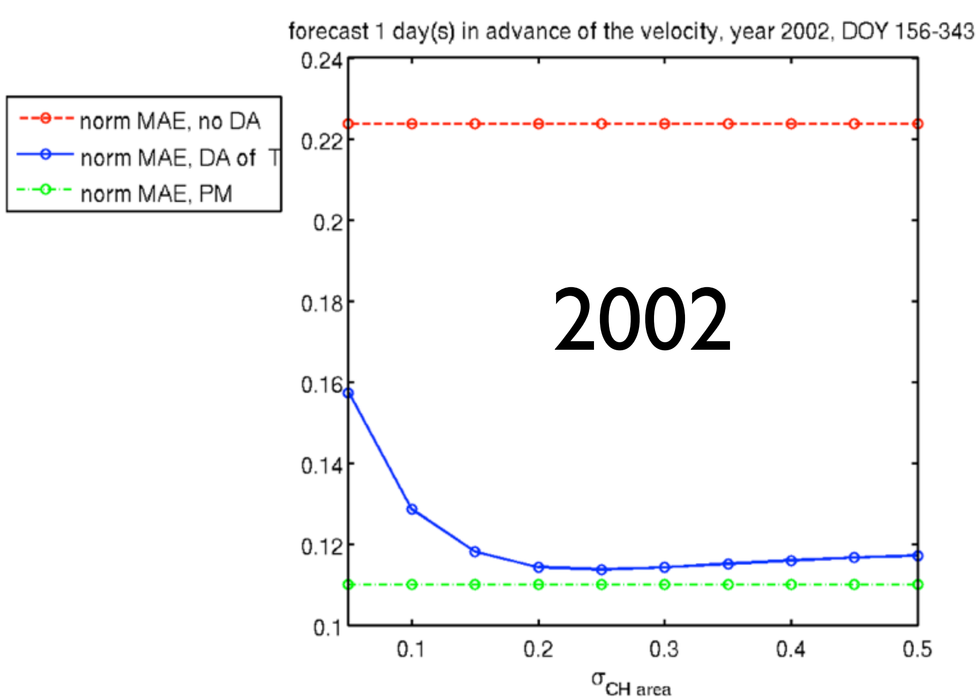
2. Performance outside of the period of applicability

We tested the performance of the method in periods of the cycle where the baseline model does *not* capture the major sources of solar wind variability

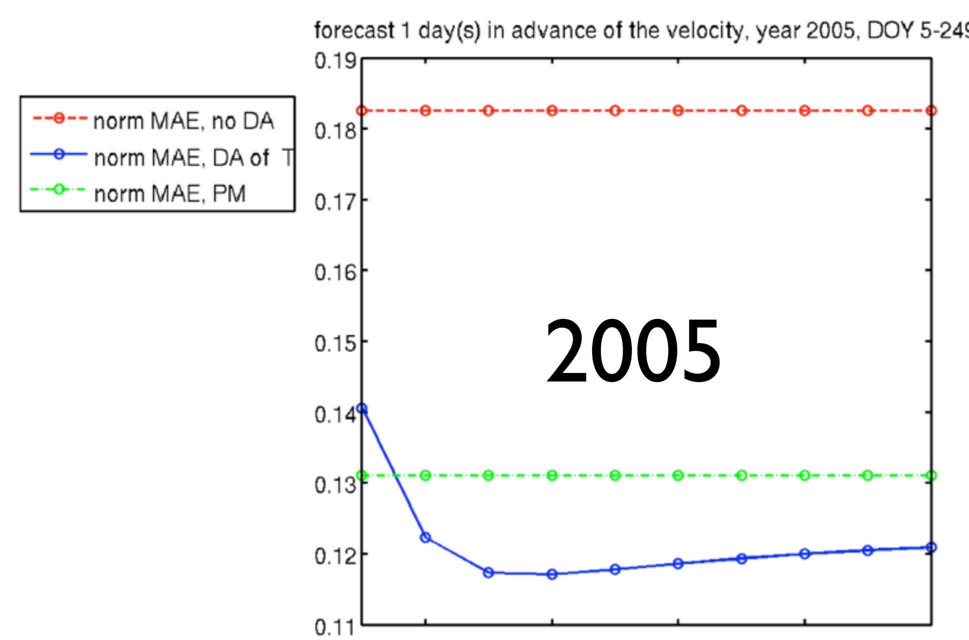




The baseline model is more physically grounded during the declining phase of the cycle (y. 2005), where HSS/ CIRs are the major driver of solar wind variability



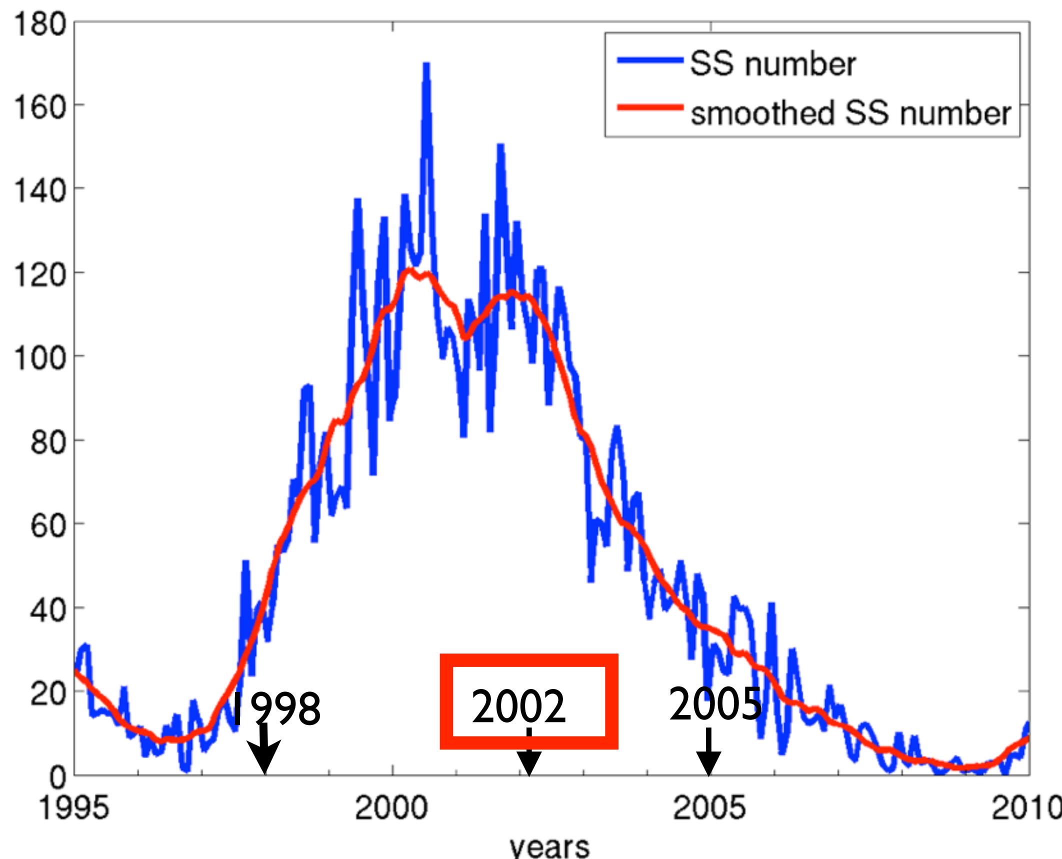
The forecasts with DA always outperform (lower MAEs) the forecasts with the baseline model. They outperform the persistence model also in the period (declining phase of the cycle) where the baseline model is more physically significant



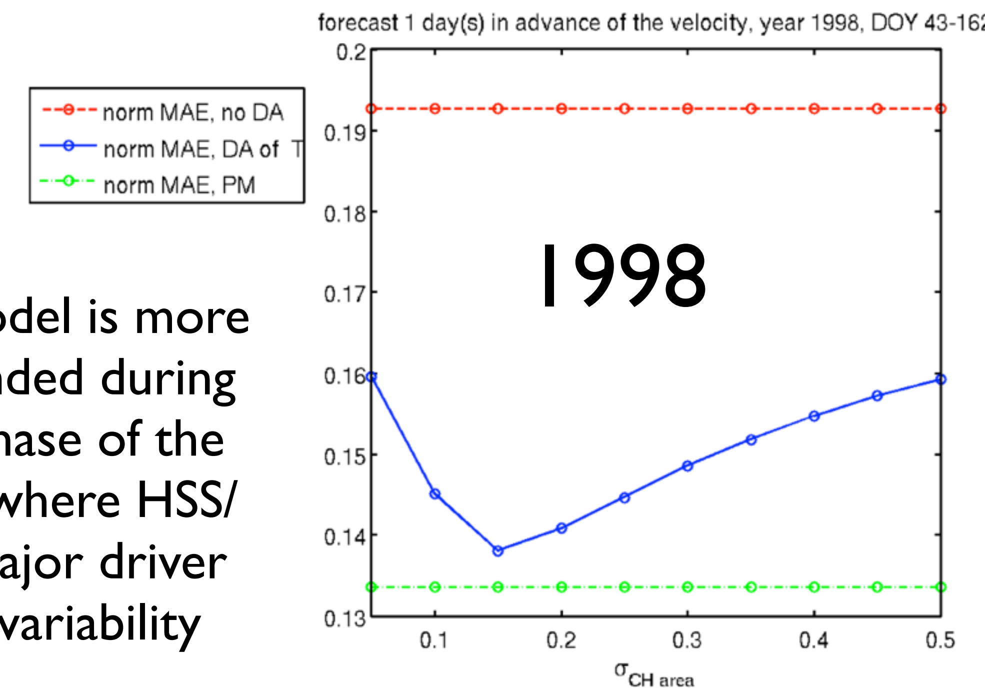
http://sidc.be/sunspot-data/

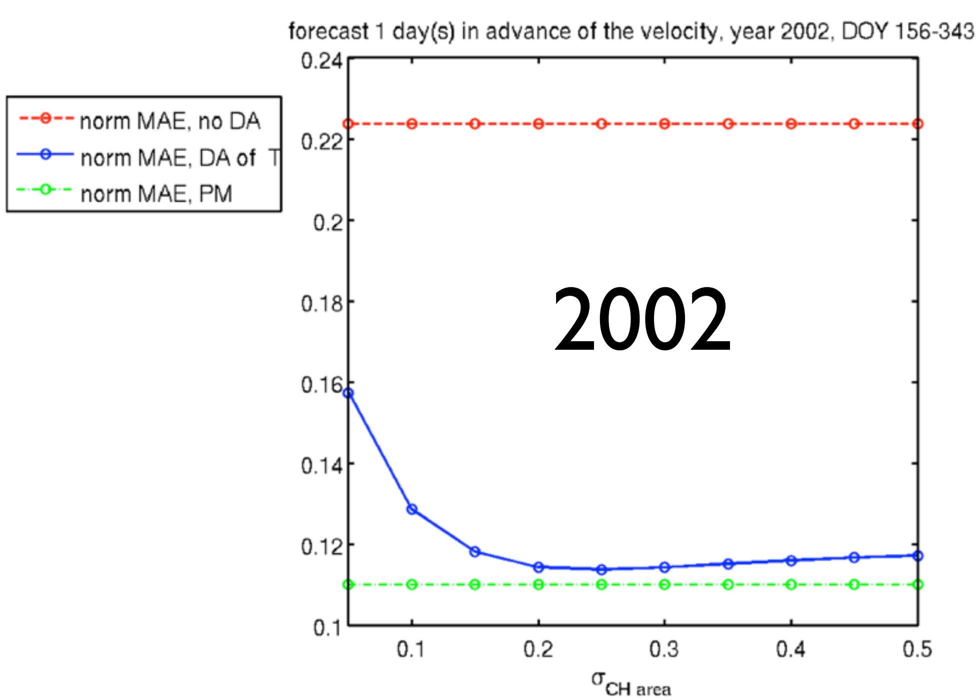
2. Performance outside of the period of applicability

We tested the performance of the method in periods of the cycle where the baseline model does *not* capture the major sources of solar wind variability

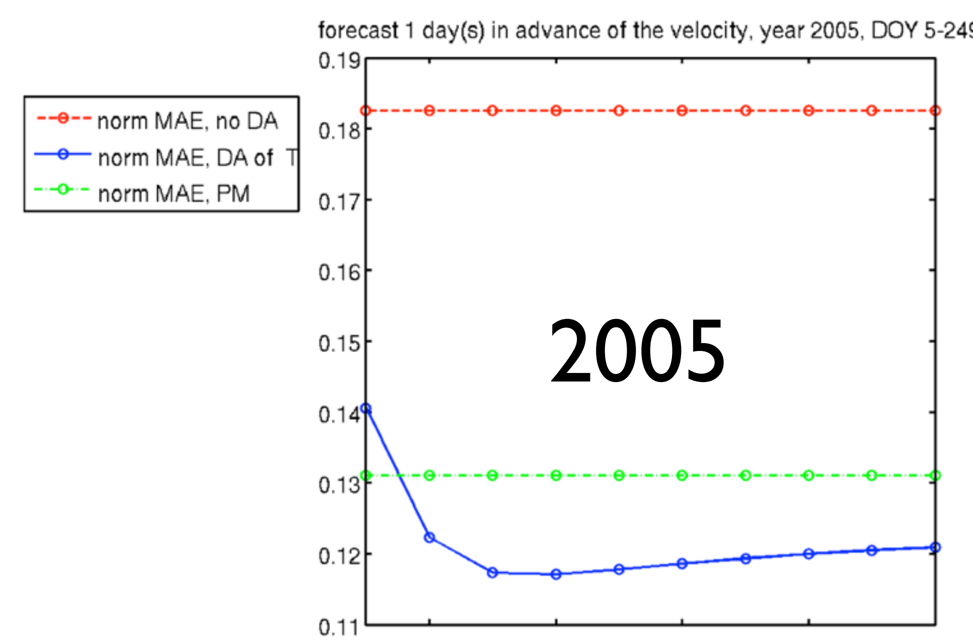


The baseline model is more physically grounded during the declining phase of the cycle (y. 2005), where HSS/CIRs are the major driver of solar wind variability





The forecasts with DA always outperform (lower MAEs) the forecasts with the baseline model. They outperform the persistence model also in the period (declining phase of the cycle) where the baseline model is more physically significant



http://sidc.be/sunspot-data/

Conclusions of part 2

We have applied Kalman filtering techniques to a very simple empirical model for the forecast of solar wind parameters. The aim is to understand if solar wind models can benefit from Data Assimilation. We have relied chiefly on the very strong (and time-delayed) connection between ion T and v in the solar wind. Less strong correlation have been explored and can be exploited. We have obtained that

- the model with DA always outperforms the baseline model; it outperforms the persistence model as well when the baseline model is more physically grounded (declining phase of the cycle) → DA is not effective per se, a good baseline model is needed
- processes not included in the baseline model (i.e., CMEs) are included in the forecast with DA thanks to the assimilation; this is the main result of the test
- DA improves the robustness of the baseline model to corrupted inputs (e.g., instrument noise)

Caveats: the baseline model used is very simple and very prone to improvements. Also, the assimilation technique used relies on a particular characteristics of the model (the time-delayed correlation between observables) which may not be available for other models

3. Preliminary study of DA benefits for an MHD model for the simulation of the heliosphere

Heliospheric models usually rely on boundary conditions at the source surface, which constitutes a major source of model errors. We used the ensemble method [Evensen09] to

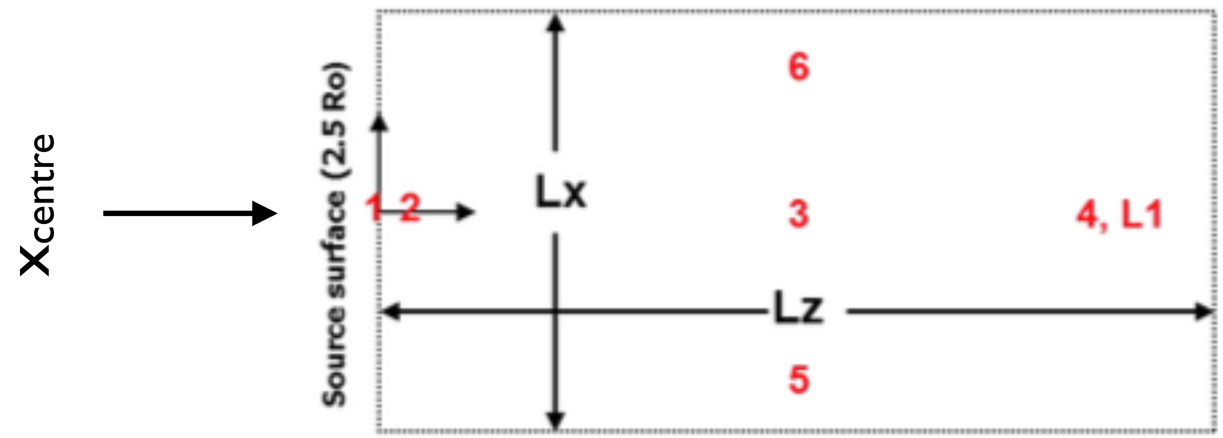
1. understand how sensitive the model is to input variation through the analysis of ensemble variances.
   In practice: how far from reality will my model go if there is an error (i.e., instrumental noise) in the boundary conditions at the source surface?

2. identify optimal locations for DA through the representer analysis
   In practice: where should my spacecraft be to get the best out of the assimilation of an observable? Which observable should I assimilate?

[Skandrani et al, 2014]

3. The baseline model

FLIP-MHD [Brackbill 1991] is used as baseline model for the simulation of the background solar wind from the source surface to 1 AU



red numbers: diagnosis locations

boundary conditions at the Source Surface:

- velocity is constant and equal to Alfven speed v_A
- magnetic field:

thickness of the inversion layer

$$b_x = 0,$$

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

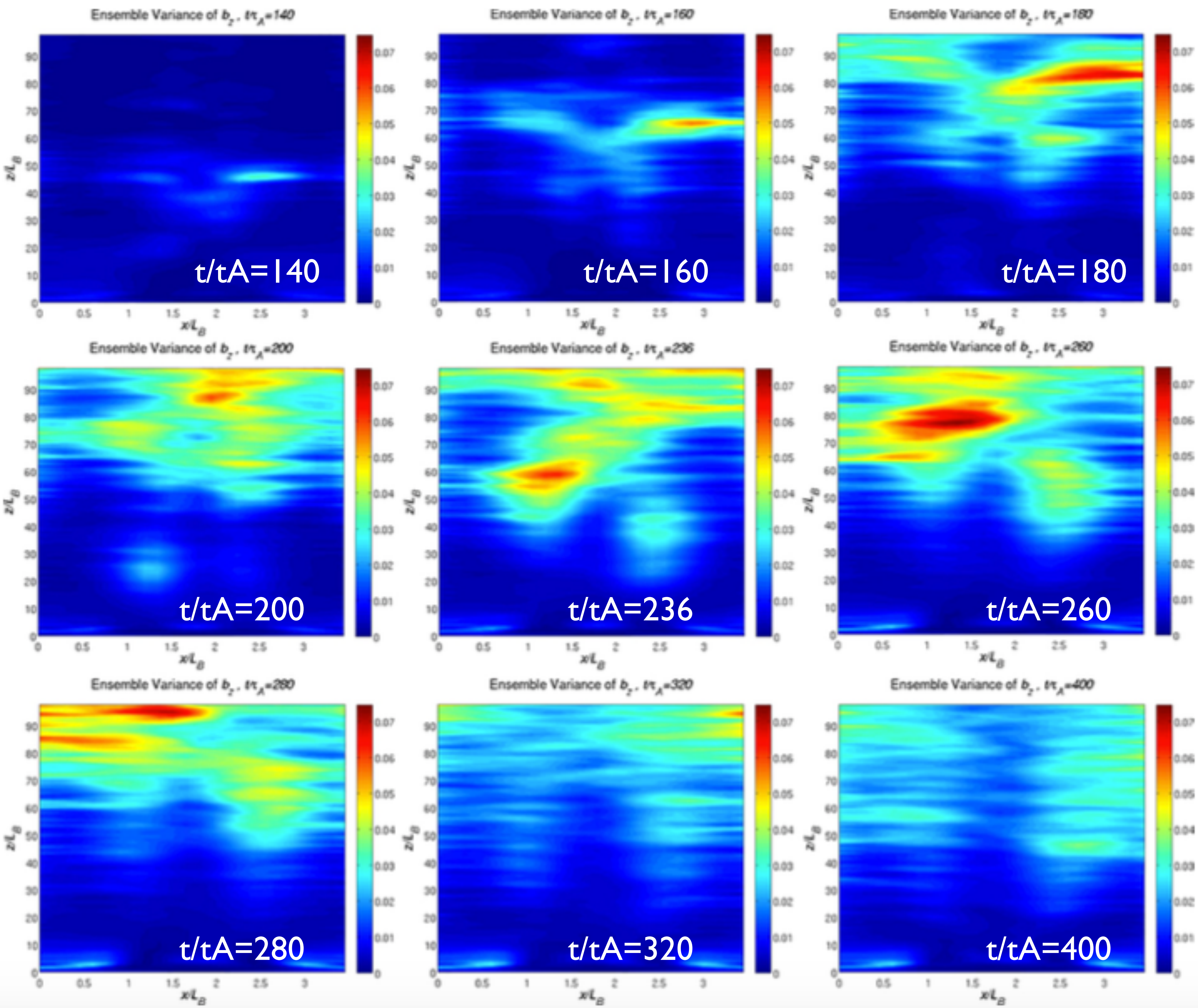
$$b_z = b_0 \tanh(x - x_{\text{centre}}) + \frac{b_0}{f(t)} \sin\left(\frac{2\pi(x - x_{\text{centre}})}{L_B}\right)$$

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

the ensemble is obtained by perturbing these initial conditions with a time dependent coefficient generated from a Gaussian distribution with mean= 0, s.d.=0.05 (small perturbation); ensemble-wide statistic is then collected

the boundary is driven at $t < t_\alpha = 0.4\ t_{final}$

3. Analysis of the ensemble variance



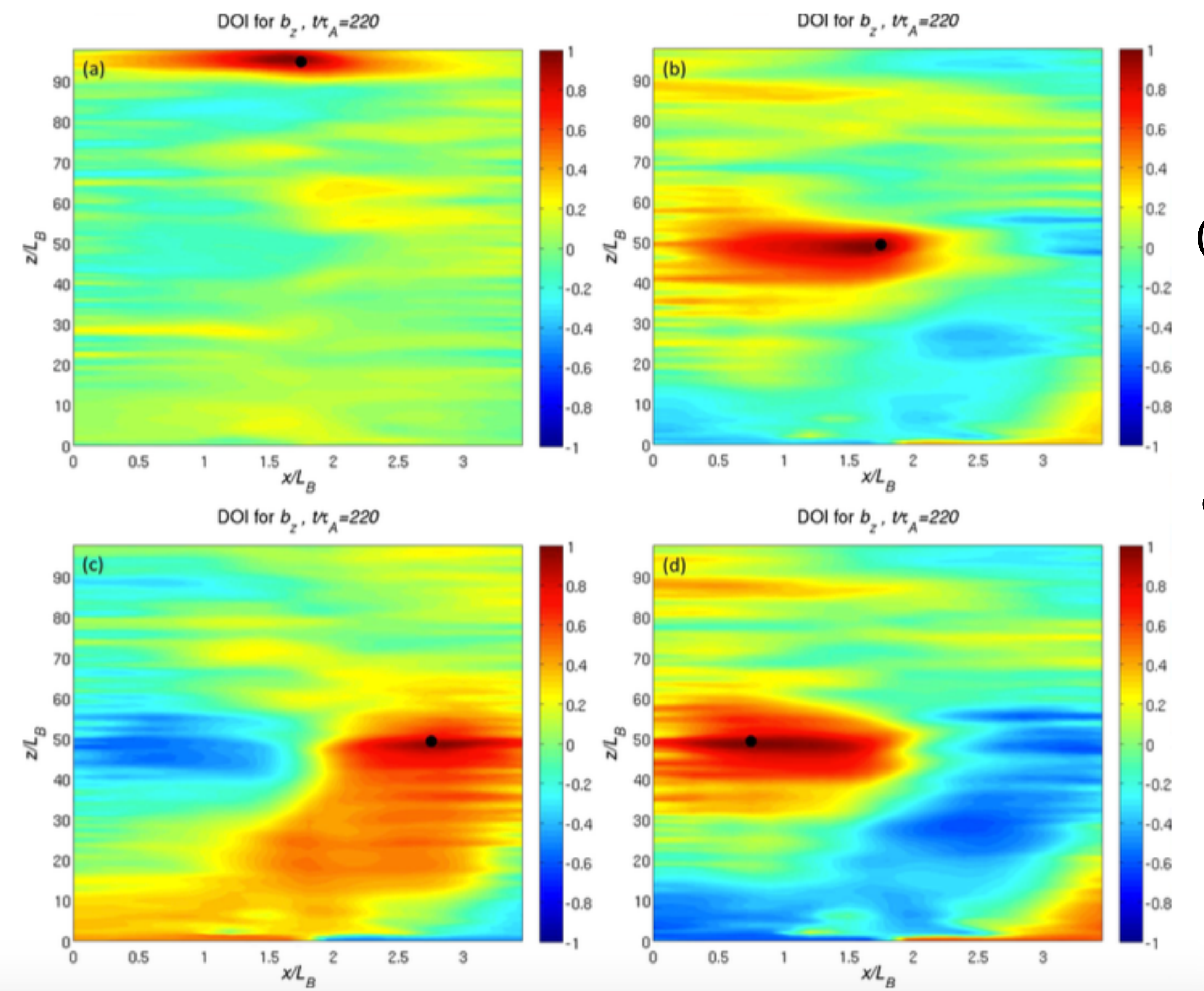
Ensemble variance of bz at differences times (tA is one Alfven transit time); blobs of higher variance (red) appear when the instances in the ensemble behave very differently e.g.: some reconnect, some not

Conclusion: a quite small perturbation of the boundary conditions result in rather different system evolution → it is crucial to have accurate boundary conditions

3. Analysis of the domain of influence

The representer analysis allows to calculate the "domain of influence" of an observation: meaning the areas in the system where assimilation of an observation at a particular location gives a result closer (doi>0) or further away (doi <0) from 'reality', where 'reality' is calculated from the ensemble

The math behind the doi is heavy, refer to [Skandrani 2014]



Effect of the assimilation of bz observations at different locations (black dots) on the bz field at a fixed time (t/tA=200)

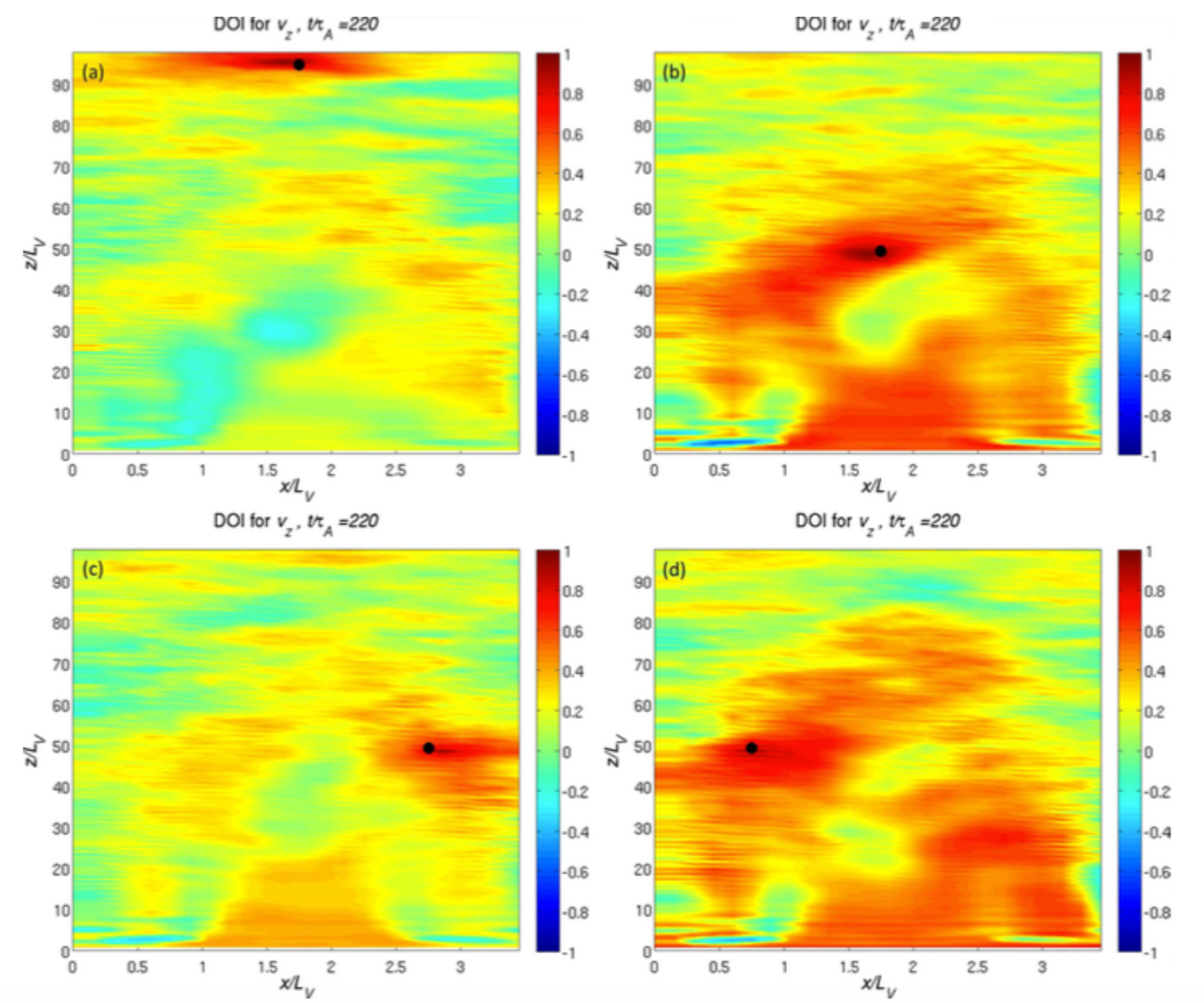
The red areas benefit from the assimilation; the blue areas are damaged → a perimeter around the assimilation location should be established for bz

Reconnection magnetically isolates certain areas of the domain and prevent benefits from the DA to spread

3. Analysis of the domain of influence

The representer analysis allows to calculate the "domain of influence" of an observation: meaning the areas in the system where assimilation of an observation at a particular location gives a result closer (doi>0) or further away (doi <0) from 'reality', where 'reality' is calculated from the ensemble

The math behind the doi is heavy, refer to [Skandrani 2014]



Effect of the assimilation of vz observations at different locations (red dots) on the vz field at a fixed time (t/tA=200)

The red areas benefit from the assimilation; no "damaged" areas → velocity observations are a better choice for assimilation than the magnetic field

Conclusions of part 2

We have applied ensemble techniques to an MHD model for the propagation of background solar wind from the source surface to 1 AU. The aim is again to understand if solar wind models can benefit from Data Assimilation. Very slight variation of the boundary conditions at the source surface are used to produce the instances of the ensemble.

- from the analysis of the ensemble variance, we have understood that quite small variations of the source surface input produce large variations in the behaviour of the ensemble members. For example, reconnection happen at very different locations. Accurate inputs are then needed

- from the analysis of the domain of influence, we have understood that velocity observations are better candidate for assimilation than magnetic field. For assimilation of velocity observations, benefits in the state can be expected quite far from the assimilation location, not so for assimilation of magnetic field observations

General conclusions

We have tried to assess the impact of Data Assimilation on solar wind forecasting models using two rather different approaches.

Both advocate for more extensive use of Data Assimilation in heliospheric modelling

References

M. E. Innocenti, G. Lapenta, B. Vrsnak, F. Crespon, C. Skandrani, M. Temmer, A. Veronig, L. Bettarini, S. Markidis, and M. Skender. Improved forecasts of solar wind parameters using the Kalman filter. Space Weather, 9(10), 10 2011.

C. Skandrani, M. E. Innocenti, L. Bettarini, F. Crespon, J. Lamouroux, and G. Lapenta. FLIP-MHD-based model sensitivity analysis. Nonlinear Processes in Geophysics, 21(2):539–553, 2014.