WP 3 Year 2 Report

A. M. Jorgensen
Electrical Engineering Department
New Mexico Tech
USA

Outline

- Overview of Milestones and Status
- Overview of Deliverables and Status
- Details of Milestones and Deliverables
- Work Plan for Year 3

Science Results and Lessons Interspersed...

Milestones and Status

#	Due Date	Description	Status
MS6	6 months	Basic Particle Filter and Ensemble Kalman Filter. Modified to Basic Ensemble Kalman Filter	Complete
MS7	12 months	Compare Particle Filter and Ensemble Kalman Filter with Simulated Data. Modified to demonstrate Ensemble Kalman Filter with Simulated Data and In-Situ Data	Complete
MS8	24 months	Demonstrate use of data assimilation with Whistler and FLR data	Complete
MS9	24 months	Begin delivering plasma density maps to WP 4	Complete

Deliverables and Status

#	Due Date	Description	Status
D3.1	24 months	Data assimilation code written in C++ with MPI which implements Ensemble Kalman Filter with DGCPM, dipole magnetic field, and adjustable electric field.	Complete – but still adding features and improving interface
D3.2	24 months	Plasma density maps as a function of time for interesting study periods selected by the team	Complete
D3.3	42 months	A set of instructions that will allow non-developers to run the assimilation code.	Complete

Other Items not in MS and D

- Near Real-time connection to data stream

Details of Milestones and Deliverables

- MS 6 and D3.1: Basic Ensemble Kalman Filter
- MS 7: Demonstrate Ensemble Kalman Filter with Simulated and In-Situ data
- MS 8: Demonstrate Ensemble Kalman Filter with VLF Whistler and Magnetometer FLR data
- MS 9/D3.2: Begin delivering plasma density maps to WP4

MS 6/D3.1: Ensemble Kalman Filter

- Architecture of code and all its components described here
- DGCPM
  - Data model comparison
- Data assimilation
  - EnKF
  - Parallelization with MPI
  - MPI
  - ScaLAPACK
  - BLACS
- Electric Potential Model

Sequential Bayesian Probability

- System state

$$\psi = \{\psi^0, \ldots, \psi^k\} \qquad \psi^i = f\left(\psi^{i-1}\right) \qquad \psi^i = f\left(\psi^{i-1}, q^i\right)$$

- Bayesian probability

$$f\left(\psi|d\right) = \frac{f\left(d|\psi\right) f\left(\psi\right)}{\int \left(\ldots\right)\, d\psi} = f\left(\psi_0, \ldots, \psi_k | d_1, \ldots, d_k\right)$$

- Sequential evaluation of probability

$$f\left(\psi_0, \ldots, \psi_k | d_1, \ldots, d_k\right)$$
$$= \frac{\boxed{f\left(\psi_0, \ldots, \psi_{k-1} | d_1, \ldots, d_{k-1}\right)}\, f\left(\psi_k | \psi_{k-1}\right) f\left(d_k, \psi_k\right)}{\int \left(\ldots\right)\, d\psi}$$

Probability density at
previous time-step

Kalman Filter

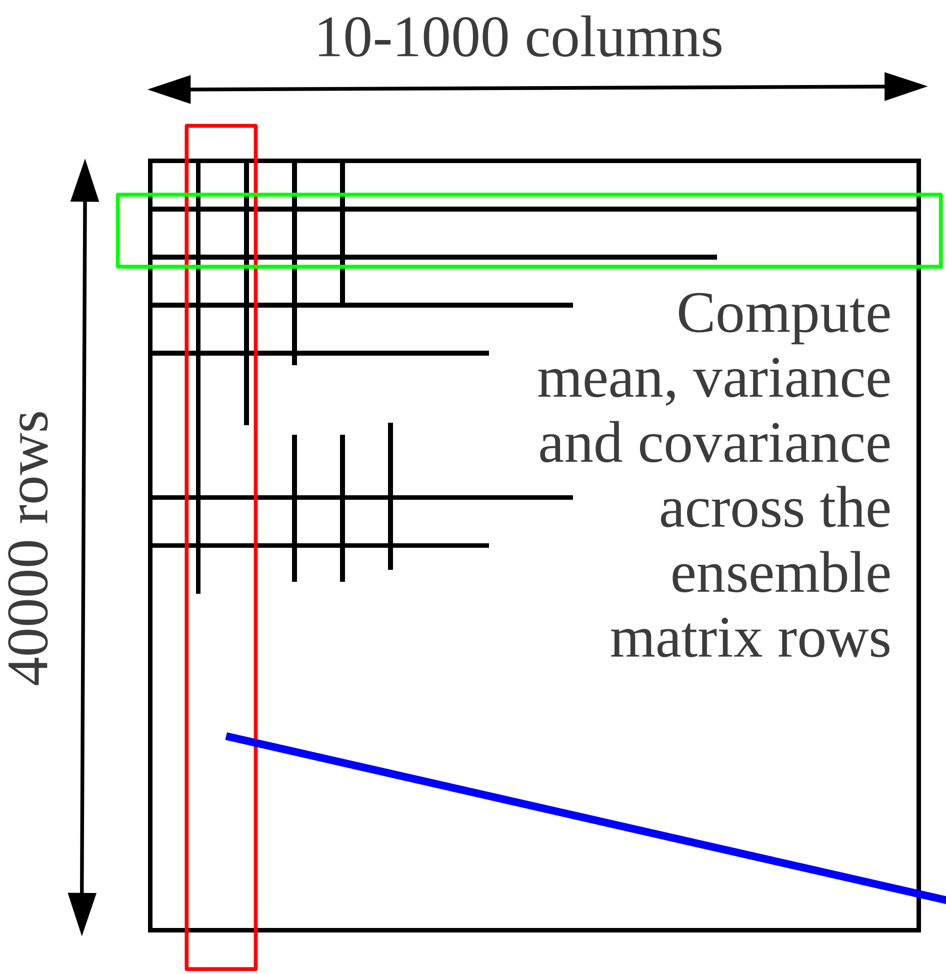
- If the system is linear and obeys Gaussian statistics,

  $$\boldsymbol{\alpha}\psi_{\boldsymbol{a}}^{\boldsymbol{i}} + \boldsymbol{\beta}\psi_{\boldsymbol{b}}^{\boldsymbol{i}} = \boldsymbol{\alpha f}\left(\psi_{\boldsymbol{a}}^{\boldsymbol{i-1}}\right) + \boldsymbol{\beta f}\left(\psi_{\boldsymbol{b}}^{\boldsymbol{i-1}}\right)$$

  then everything is expressed matrix form $\overline{\psi}^{\,\boldsymbol{i}} = \overline{\overline{\boldsymbol{X}}}\,\overline{\psi}^{\,\boldsymbol{i-1}}$

  with a covariance matrix $\overline{\overline{\boldsymbol{C}}}$ with similar update equation.
- Problem is that $\overline{\overline{\boldsymbol{C}}}$ is huge for a realistic system. DGCPM has $\mathbf{4 \times 10^4}$ grid points so $\overline{\overline{\boldsymbol{C}}}$ has $\mathbf{1.6 \times 10^9}$. Update becomes prohibitive.
- Also, many elements of $\overline{\overline{\boldsymbol{C}}}$ are of no interest of (nearly) zero.
- Instead use a ensemble representation in which several models are run in parallel

Ensemble Kalman Filter (EnKF)



Each column is a
complete model state

Evolve each column as before

$$\psi^i = f\left(\psi^{i-1}\right)$$

Each column evolves differently because of model noise. Here is "red noise"

$$q_{k+1} = \alpha q_k + \sqrt{1-\alpha^2}\, w_k \quad \alpha = \frac{1}{\tau}$$

Noise can be applied to all elements of the state, but we choose to add more variables as drivers, forming a "enhanced" state

$$\psi = \begin{pmatrix} q \\ \psi_N \\ \vdots \\ \psi_0 \end{pmatrix}$$

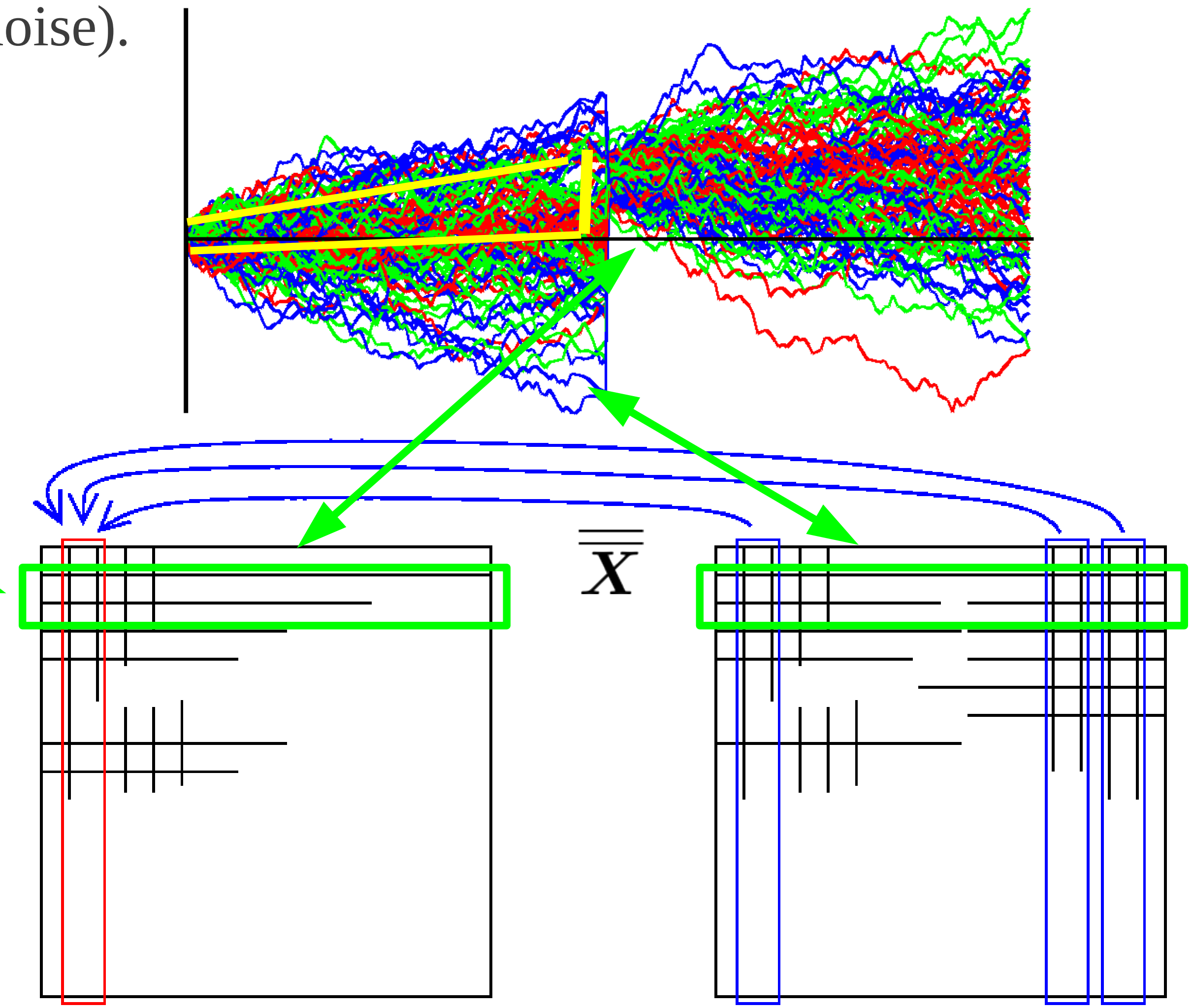
The Filter of the EnKF

Left alone the ensemble members will diverge over time due to the noise. “Analysis” at times of observations is linear transformation which reduces noise to observation uncertainty (if smaller than model noise).

$$\overline{\overline{\boldsymbol{\psi}}}_{\boldsymbol{a}} = \overline{\overline{\boldsymbol{\psi}}}_{\boldsymbol{f}}\,\overline{\overline{\boldsymbol{X}}}$$

or

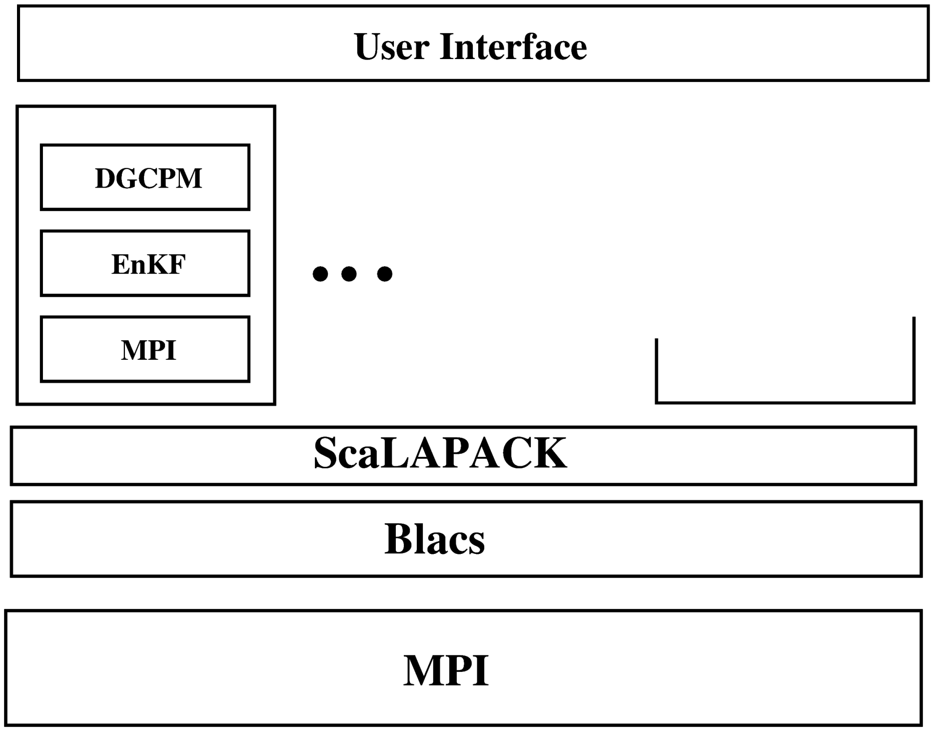
$$\overline{\boldsymbol{\psi}}_{\boldsymbol{ai}} = \sum_{\boldsymbol{j}} \boldsymbol{x}_{\boldsymbol{ji}}\,\overline{\boldsymbol{\psi}}_{\boldsymbol{fj}}$$



Post analysis variance in observed cells is (approximately) variance of observation.

Variance in the rest of the state space is reduced also

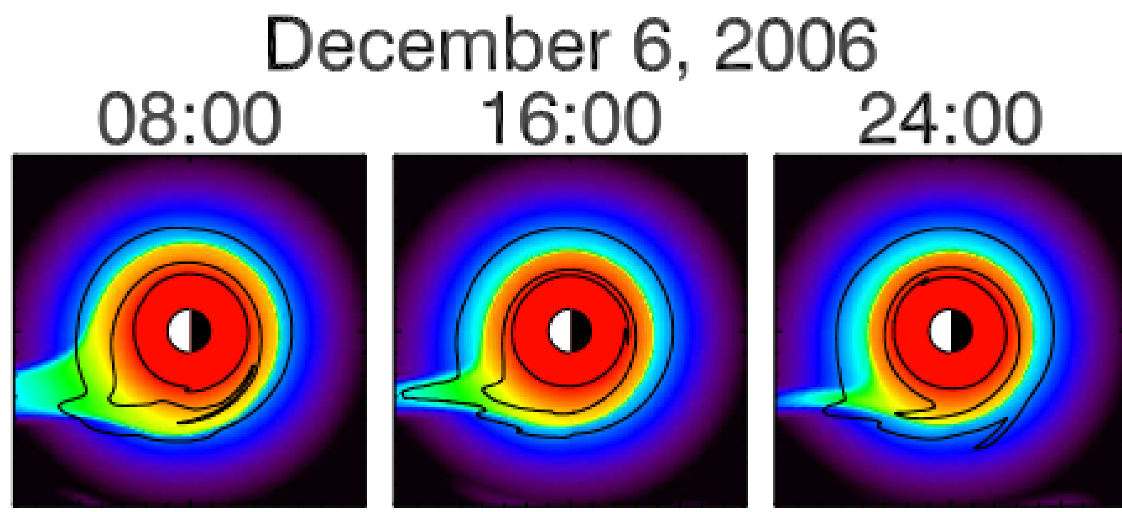
Code Architecture and Components



- Top-level C++ interface
- DGCPM in Fortran with C++ wrapper
- EnKF as C++ object
- ScaLAPACK – linear algebra package, in C
- BLACS – basic matrix operations, in C
- MPI parallelization, in C
- Runs on single machine or cluster or any size
- Currently on 16-thread Workstation

DGCPM

- 2D single species model of the plasmasphere (e.g. Ober et al. [1997])



$$\vec{B}\left(\vec{r}\right)$$

$$F_n = -\frac{NB_i}{\tau}$$

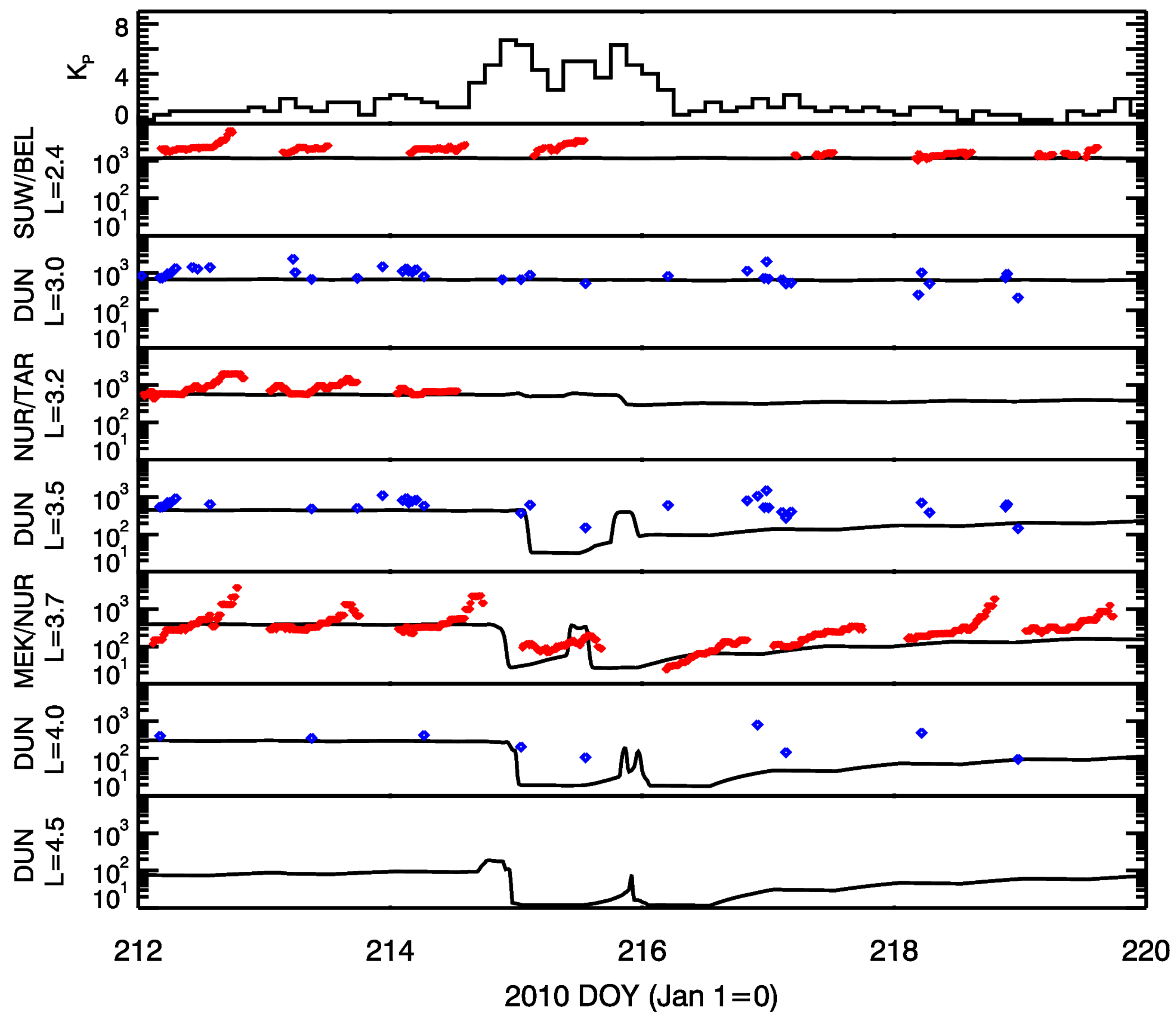
$$\frac{D_\perp N}{Dt} = \frac{F_N + F_S}{B_i}$$

$$\vec{E}\left(\vec{r}\right)$$

$$F_d = \frac{n_{\mathrm{sat}} - n}{n_{\mathrm{sat}}} F_{\max}$$

Data Model Comparison

- August 2010 event (more on this event later)
- Conclusion from comparison:
- DGCPM is good enough to model the variations that we see in the data, and can be used for assimilation



Data Assimilation

- We make use of the “enhanced state” formulation of the Ensemble Kalman filter

  Parameters of the model are included in the state and are updated in time according to a red-noise model (but could be any other model)

$$\boldsymbol{\psi} = \begin{pmatrix} \boldsymbol{q} \\ \boldsymbol{\psi_N} \\ \vdots \\ \boldsymbol{\psi_0} \end{pmatrix}$$

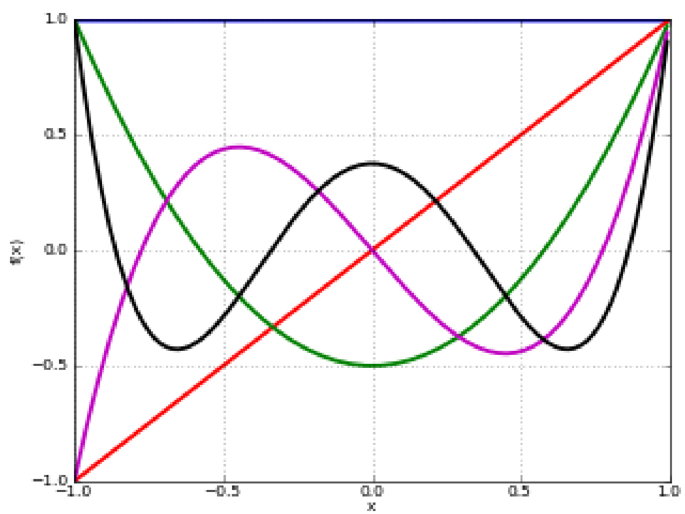
$$\boldsymbol{q_{k+1} = \alpha q_k + \sqrt{1-\alpha^2}\, w_k}$$

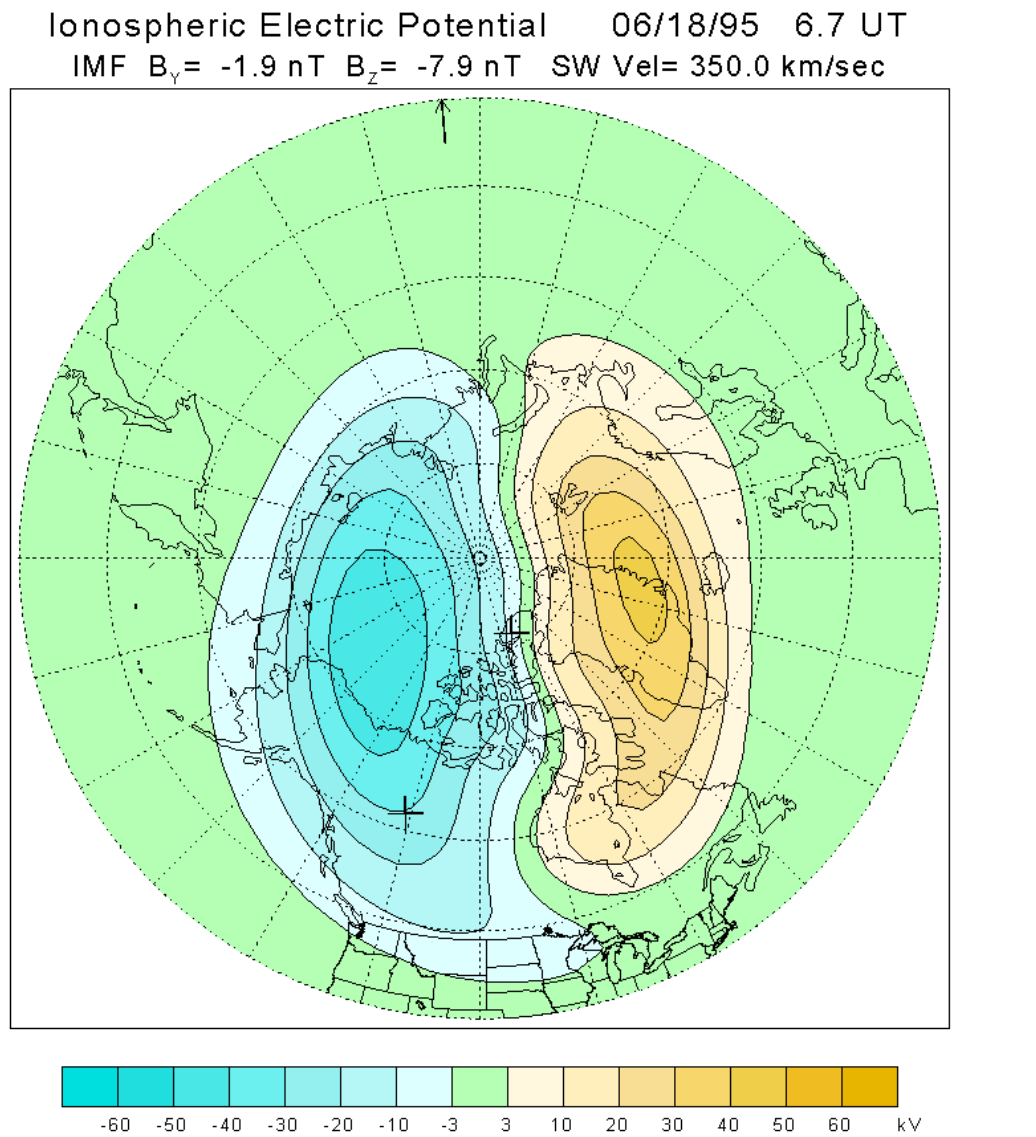
Electric Potential Model

- Based on the Weimer (2002) basis function set

$$\Phi(\theta,\phi) = \sum_{l=0}^{4} \sum_{m=0}^{min(l,3)} (A_{lm}\cos m\phi + B_{lm}\sin m\phi)P_l^m(\cos\theta),$$

- Assoc. Legendre Polynomials
- Periodic angular functions





MS 6: Model Input

- Time
- Density and uncertainty estimate
- Position (L, MLT)

MS 6: Model Output

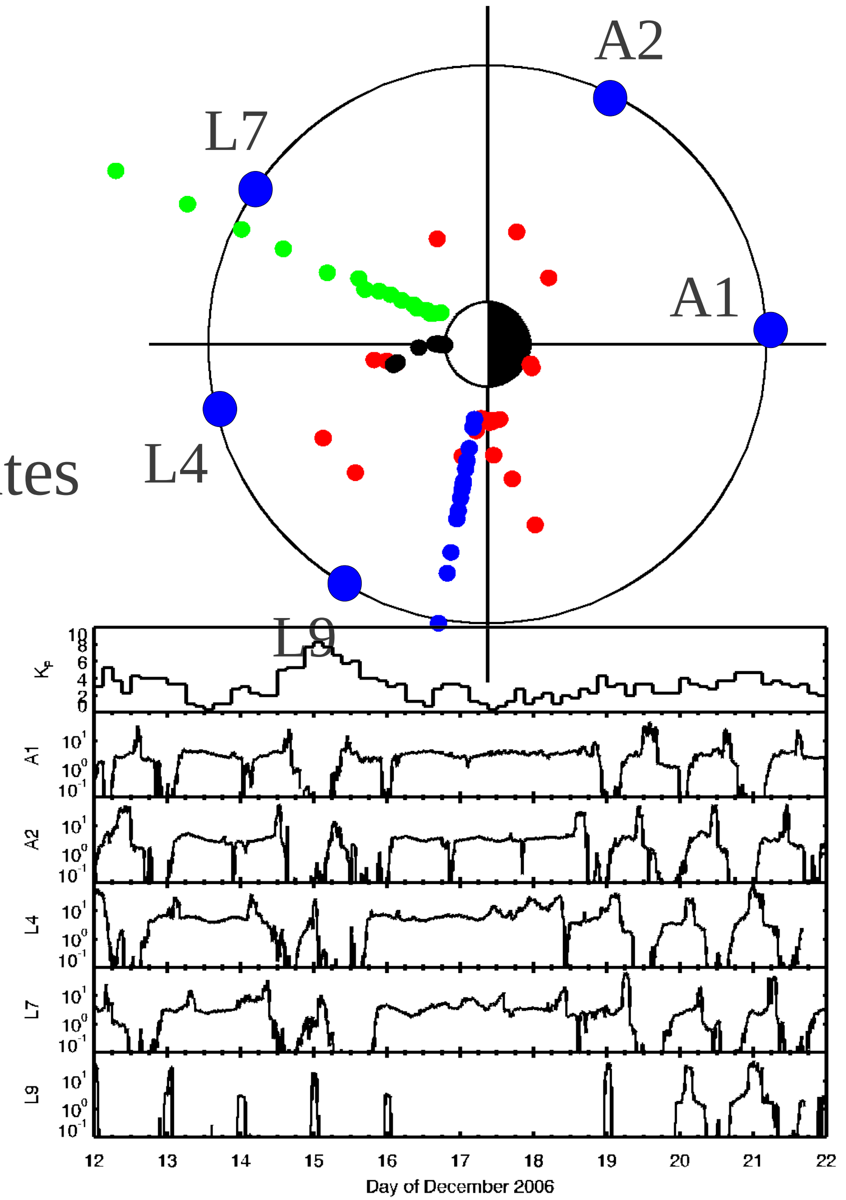
- The assimilation runs 100's of models internally, updating on a 10-minute (or other) model time-step
- Saving all this is a very large data volume.
- Instead we compute mean and variance of the model state at each time-step and save that information.
- The model state is the entire “enhanced state” which includes the plasma density and the values of the driver variables.

Details of Milestones and Deliverables

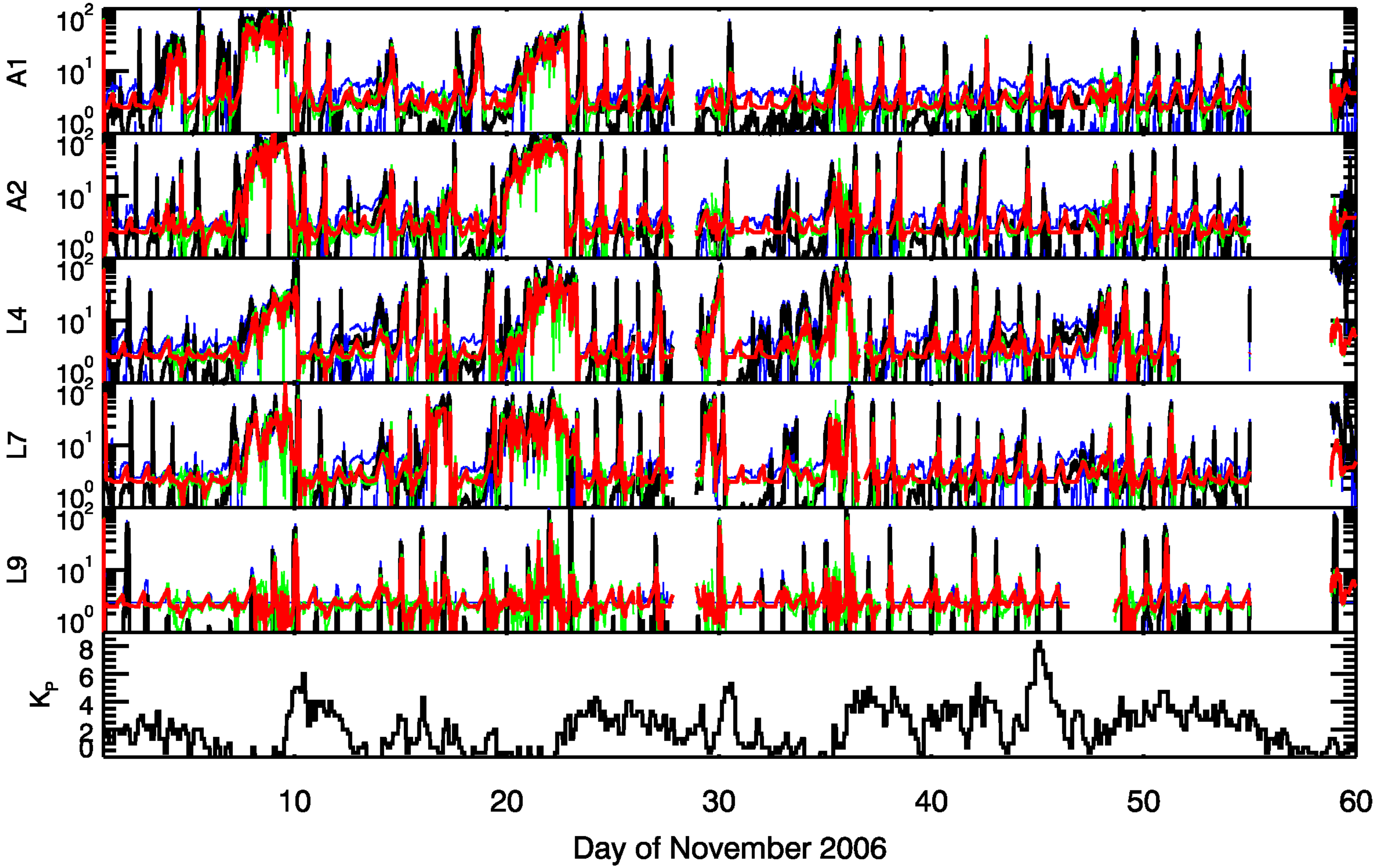
- MS 6 and D3.1: Basic Ensemble Kalman Filter
- MS 7: Demonstrate Ensemble Kalman Filter with Simulated and In-Situ data
- MS 8: Demonstrate Ensemble Kalman Filter with VLF Whistler and Magnetometer FLR data
- MS 9/D3.2: Begin delivering plasma density maps to WP4

MS 7: Demonstration With in-situ Data

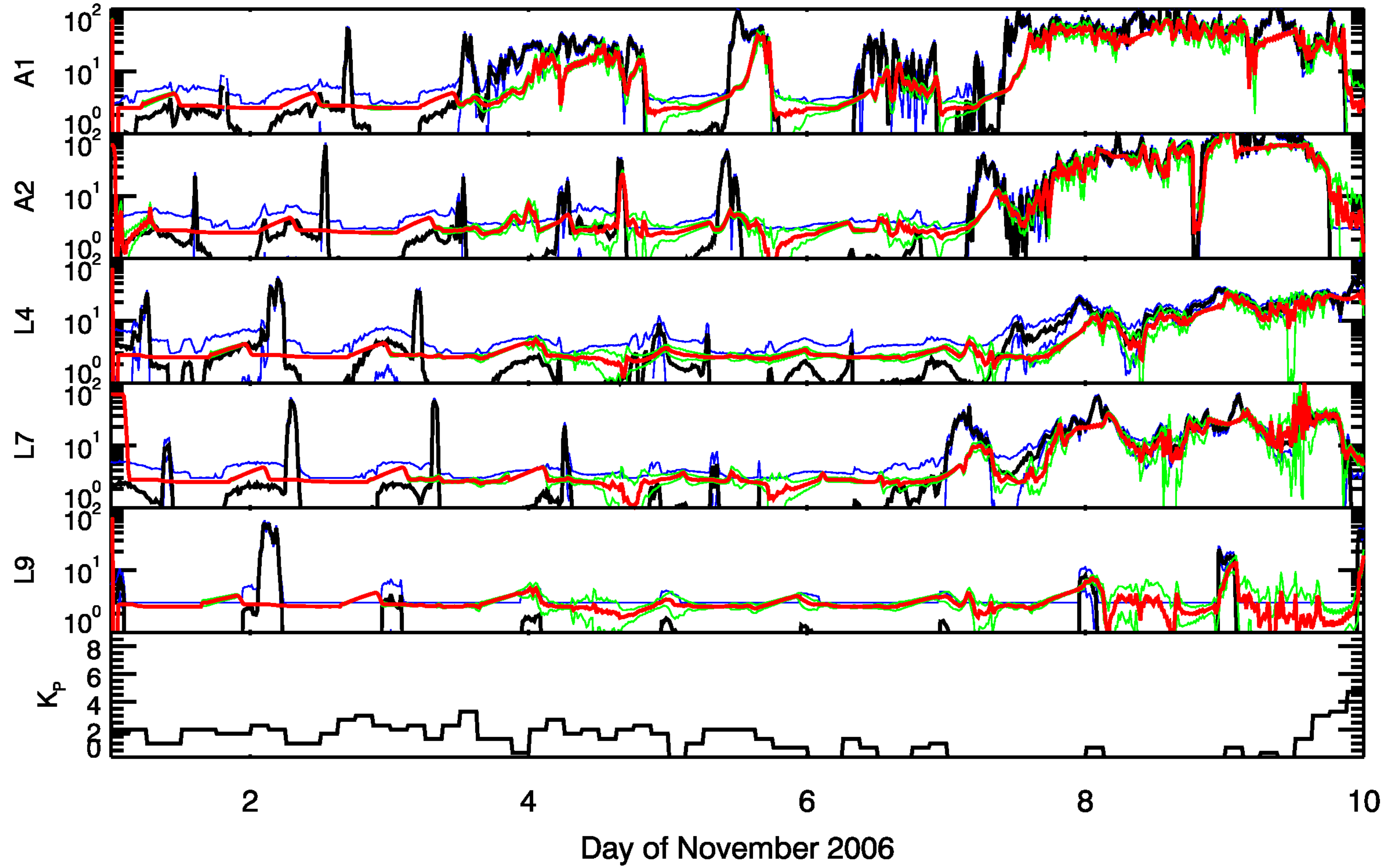
- LANL satellites
- Mostly outside plasmasphere
- Often cross dayside plume
- Reference Interval November and December 2006 when 5 LANL satellites were providing observations
- 16 CPUs
- 400 ensemble members
- 60 days
- 50x100 grid
- 10 parameters, tau=1 hour



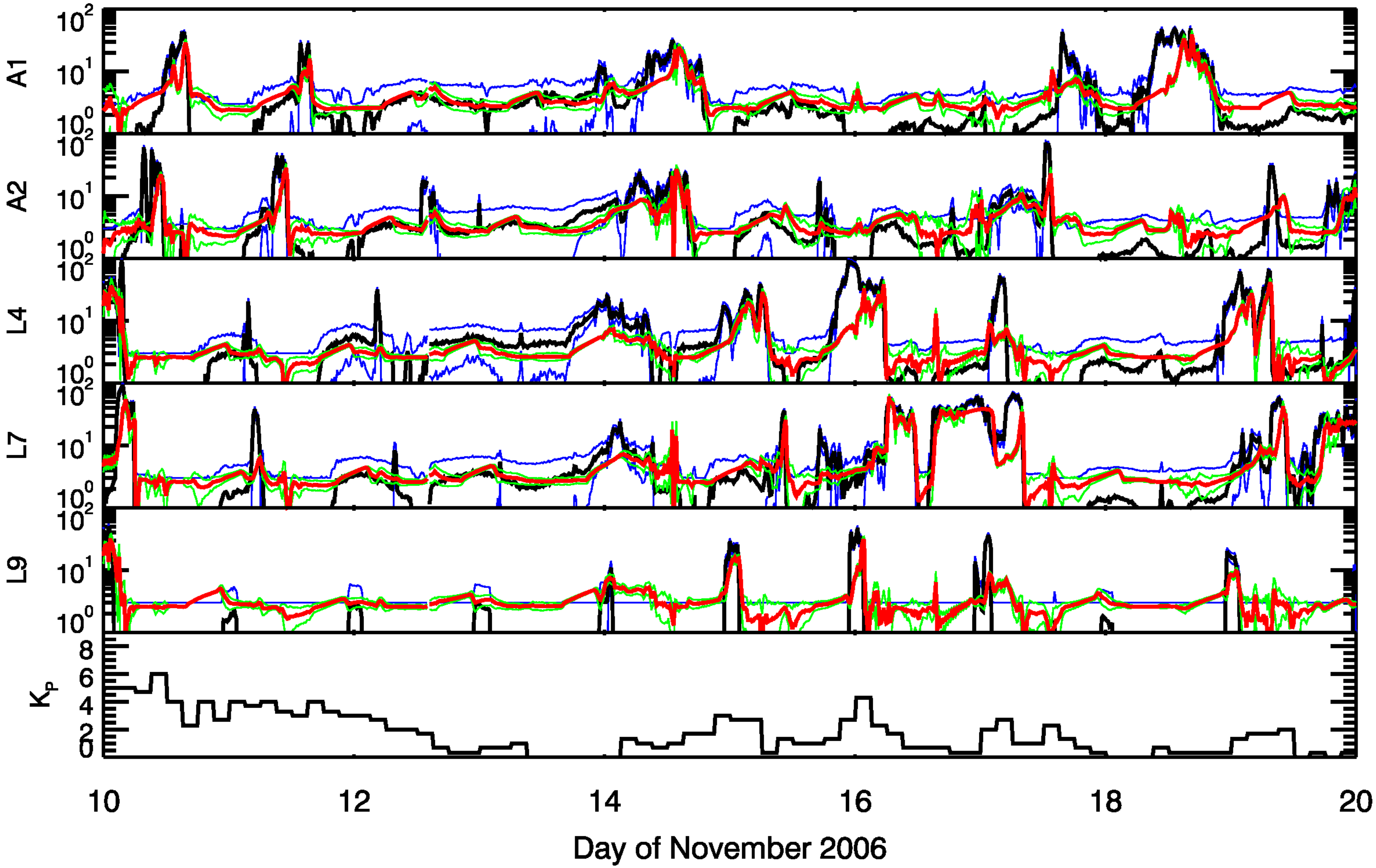
MS 7: LANL Geo Data Assimilation



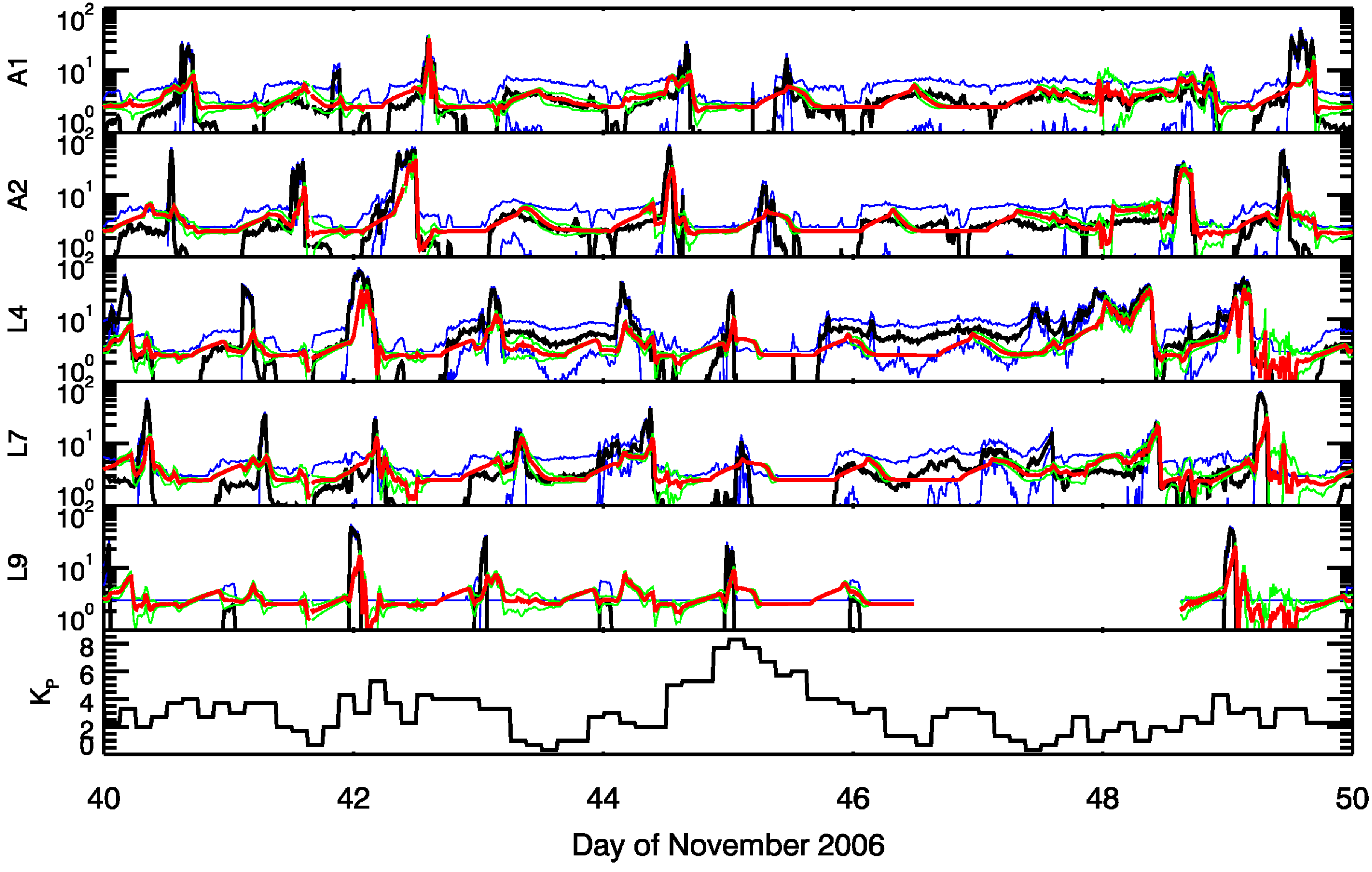
MS 7: LANL Geo Data Assimilation



MS 7: LANL Geo Data Assimilation



MS 7: LANL Geo Data Assimilation



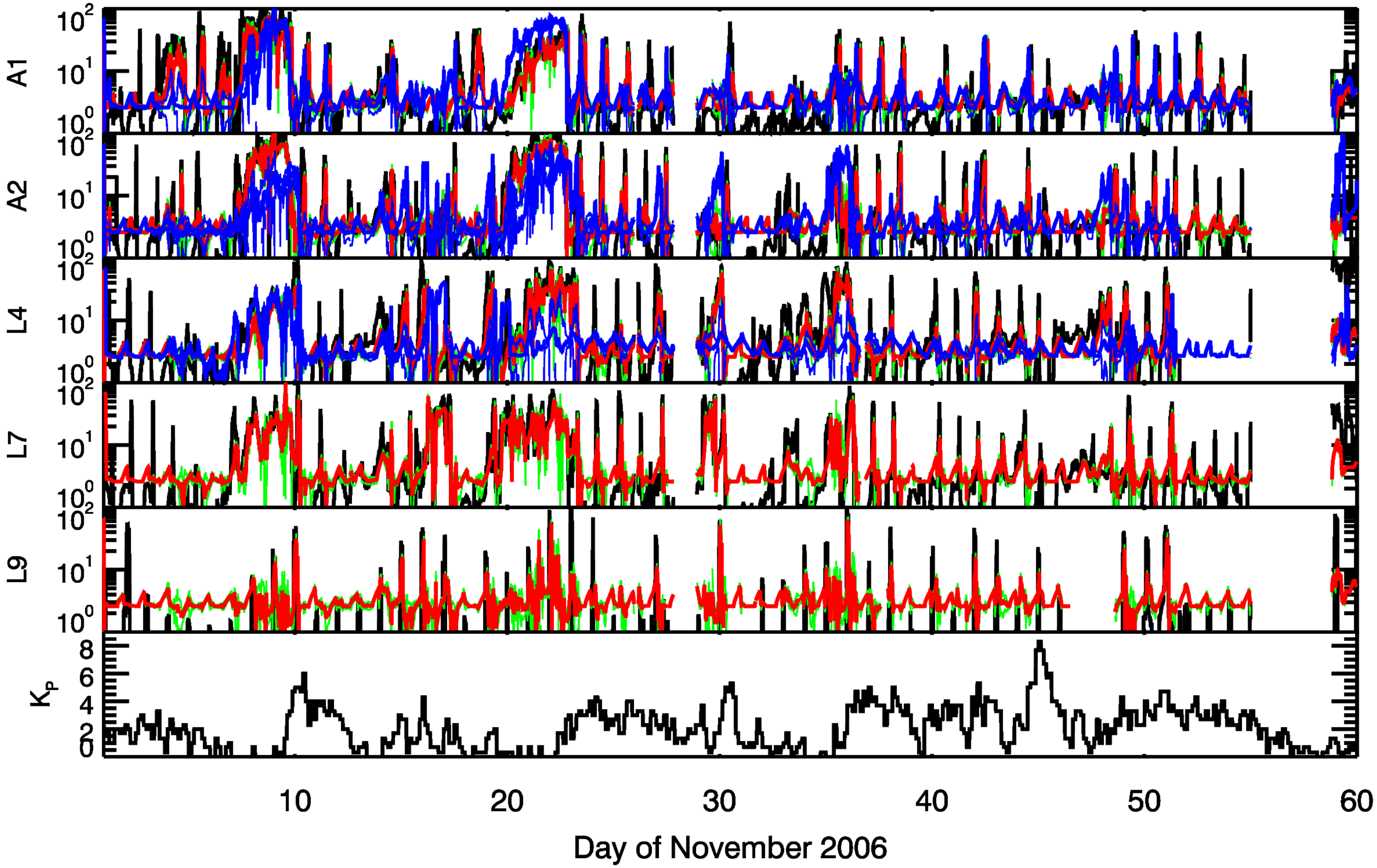
MS 7: LANL Geo Data Assimilation

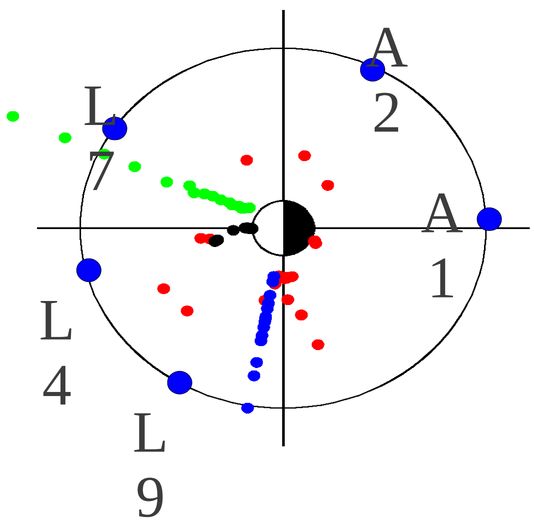
- Observations:
  - Agreement between input data and reproduced input data are quite good
- Other conclusions:
  - Similar results for different number of parameters (and different parameterizations), different time-scales, and different grid sizes
- Remaining questions:
  - Reproducing input data is the easiest. But how about agreement with out-of-sample observations?
  - This concerns the extent to which the model can be trusted to properly interpolate or extrapolate to locations or times where there are no observations.

MS 7: Out-of-Sample Verification

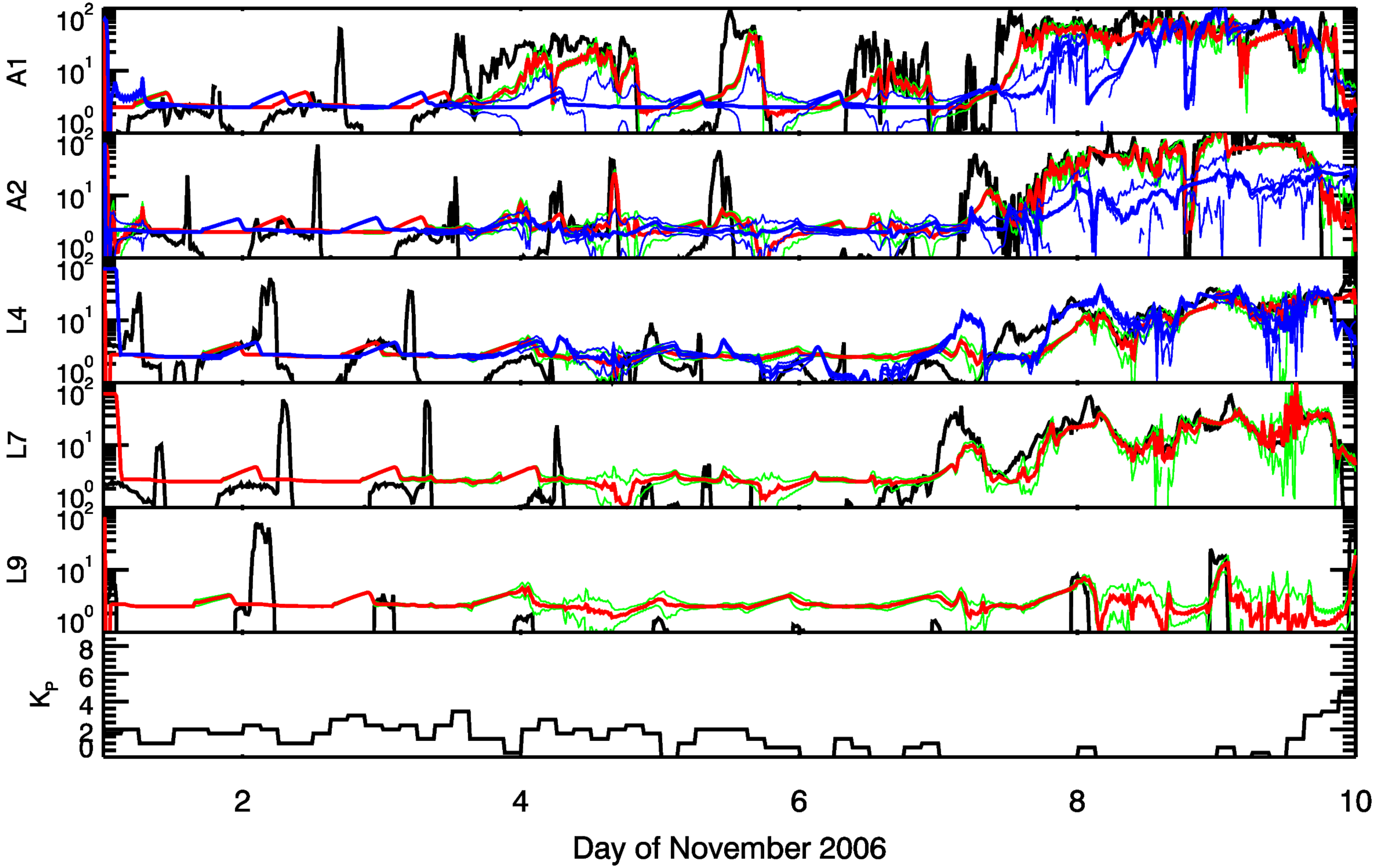
- Repeat assimilation but leaving out one of A1, A2, and L4 in turn.
- Compare assimilation with and without A1, A2, and L4.
- If assimilation can correctly interpolate/extrapolate the data at the satellite then the two should be identical
- Original data in black
- Assimilation with all data in red, green
- Assimilation with omitted satellite in blue, blue

MS 7: Out-of-Sample Verification

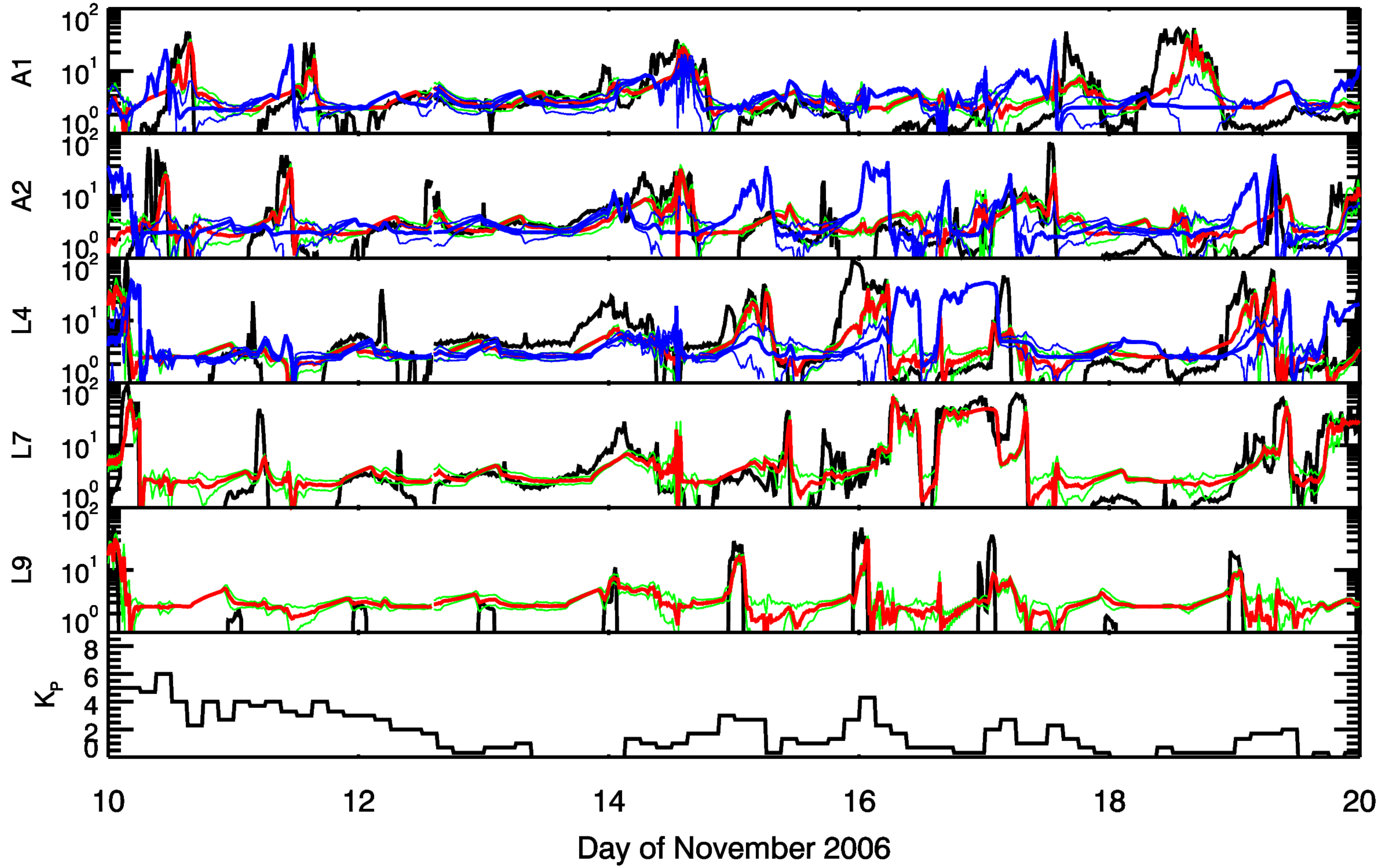




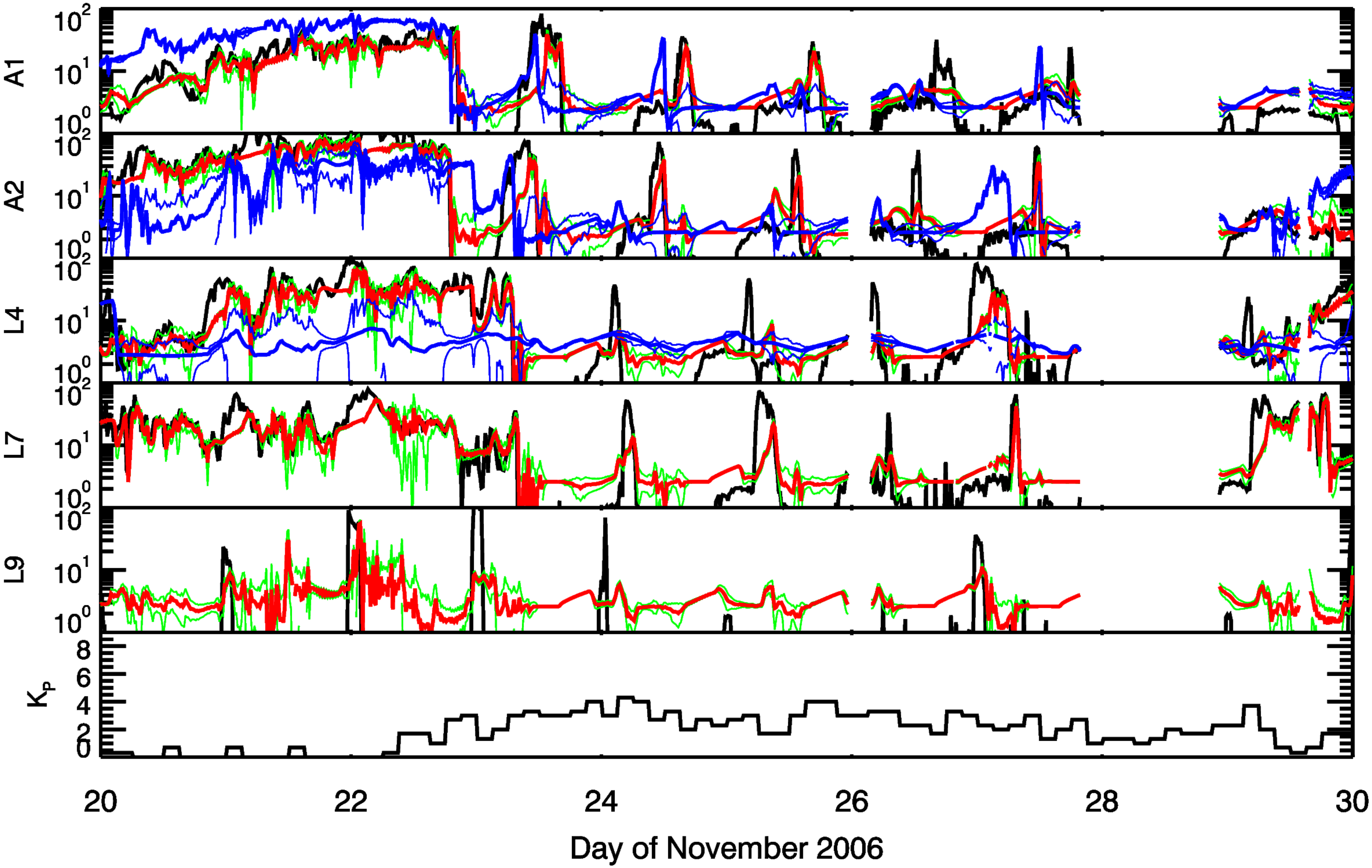
MS 7: Out-of-Sample Verification



MS 7: Out-of-Sample Verification



MS 7: Out-of-Sample Verification



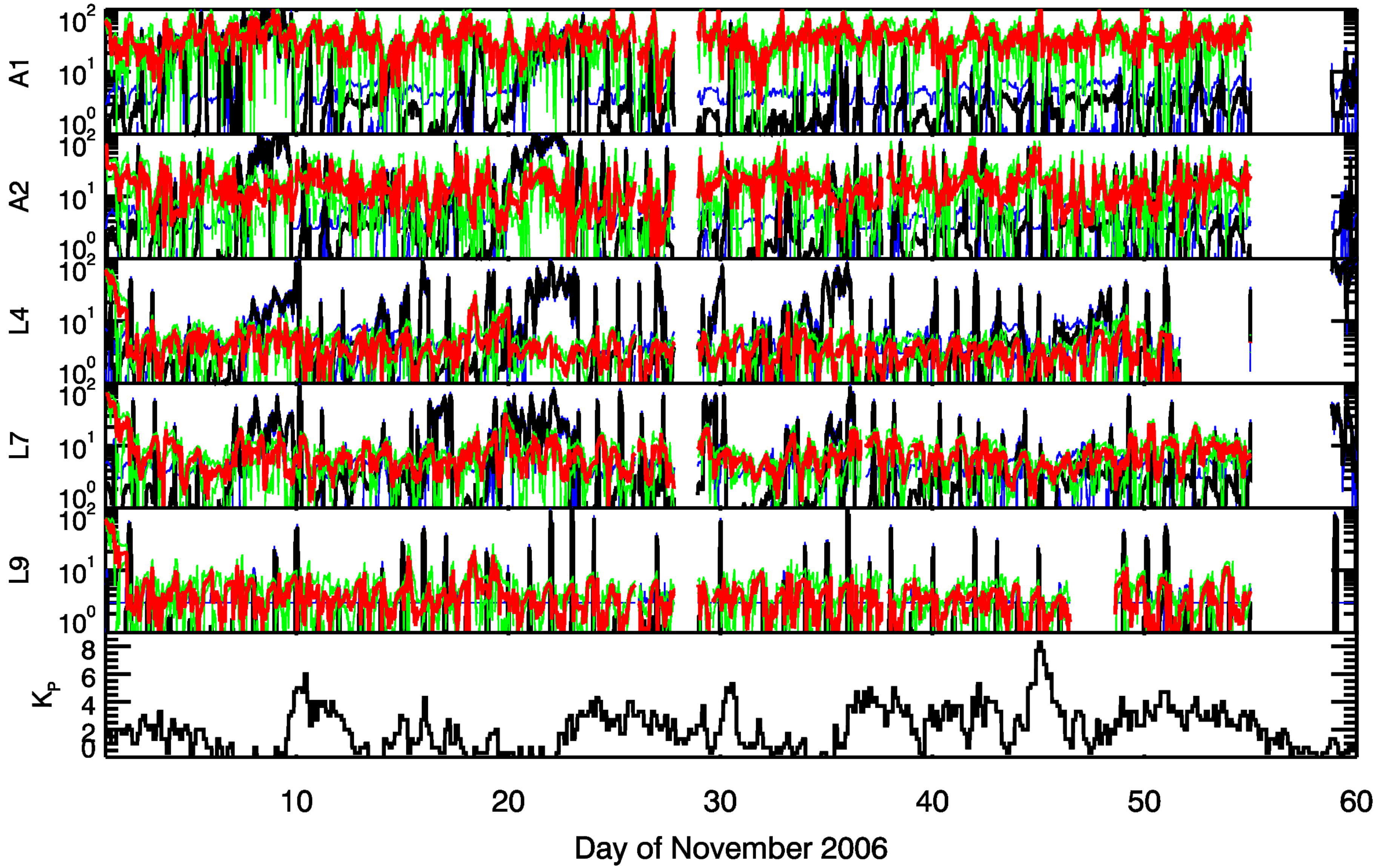
MS 7: Out-of-Sample Verification

- The agreement between out-of-sample and data is not as good as between in-sample and data – this is not surprising
- Some out-of-sample agreement is better – for example A1 agrees the best with its in-sample equivalent.
- Good out-of-sample agreement actually suggests an observation is not as important a contribution.
- Out-of-sample agreement for L4 is the worst. This may be due to the large gap to A1, since L9 is poor quality data and may not contribute much additional information.
- Conclusion: Do not expect miracles – leaving out data degrades the accuracy. The more observations the better. However gross features are preserved in most cases.

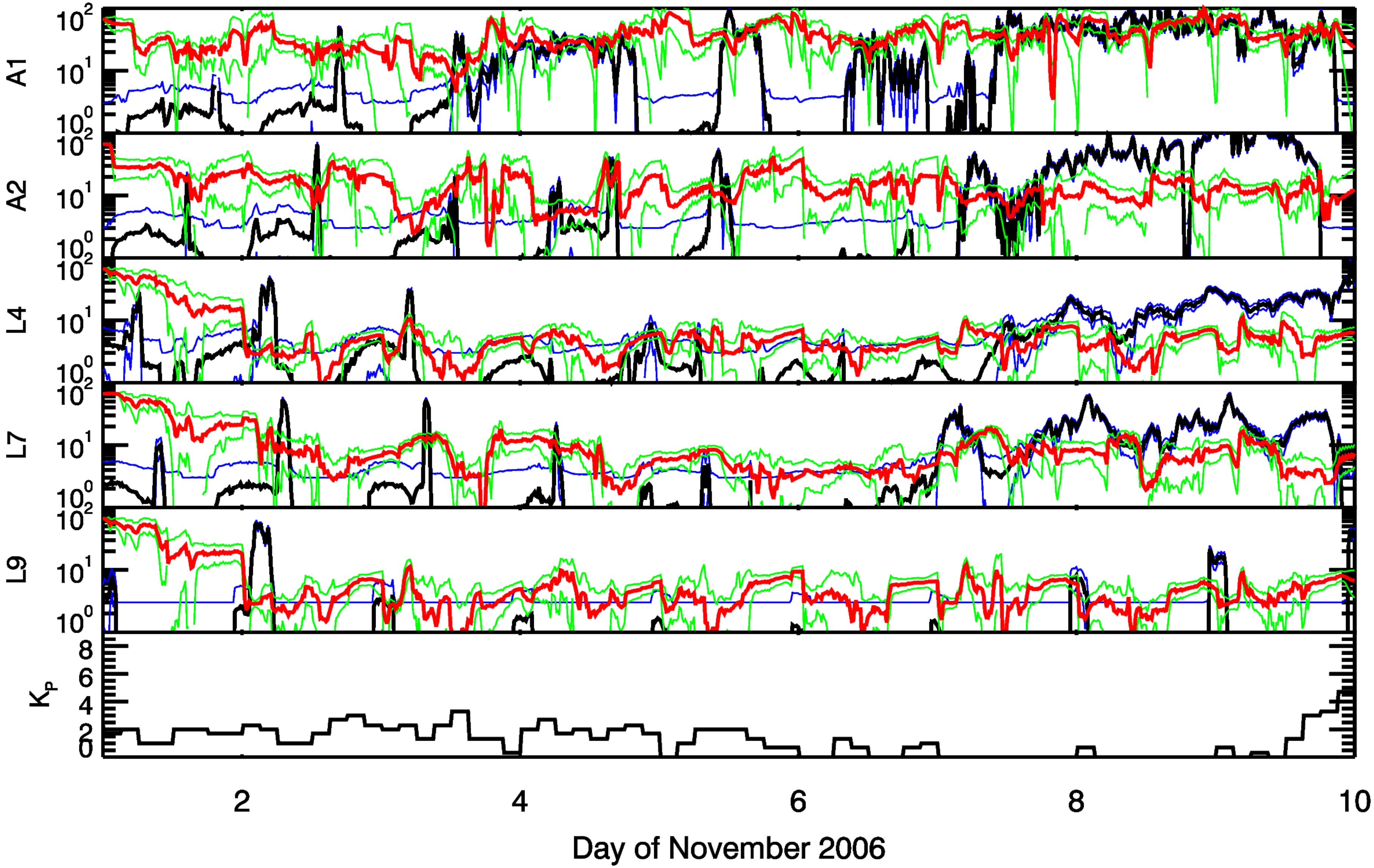
MS 7: Simulated PLASMON Data

- Question: is there enough information in the PLASMON network to model the plasmasphere?
- Answer: TBD
- One concern is the lack of high-latitude stations
- Simulated data at all PLASMON stations with following parameters:
  - Used LANL data assimilation output from 10 parameter, tau=1 hour run
  - 30% uncertainty
  - Only 10 hour coverage for VLF (night) and FLR (day)
  - Only data 25% of the time during coverage period
  - 4 parameters, tau=3 hours
- Try to reproduce original simulation input (LANL observations)

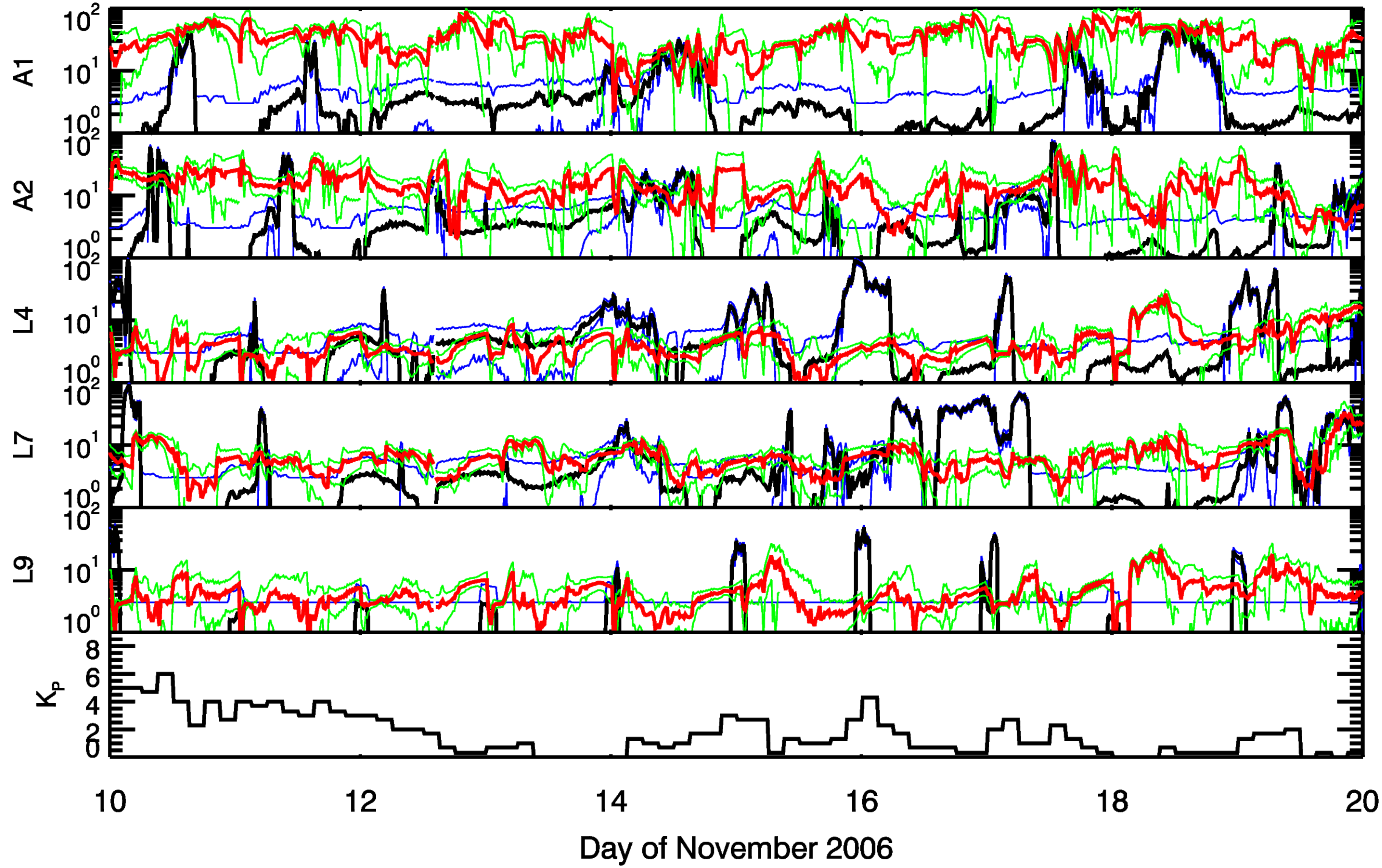
MS 7: Simulated PLASMON data



MS 7: Simulated PLASMON data



MS 7: Simulated PLASMON data



MS 7: Simulated PLASMON data

- Observations:
- Agreement is better at L4, L7, and L9
- That is also the longitude of the European and American sector magnetometer chains which reach higher latitude
- Conclusion:
- To obtain good high-latitude determination of plasma density we need good high-latitude data coverage
- This is not surprising
- This was a challenge: correctly model high-latitude densities from low-latitude observations.
- No miracles occurred

MS 7: Discussion

- The model can reproduce assimilated observations with good accuracy
- The further we move away from observations the worse the model performs – this is not surprising
- We probably cannot predicting high-latitude plasma densities from low-latitude observations and vice-versa
- No free lunch!

Details of Milestones and Deliverables

- MS 6 and D3.1: Basic Ensemble Kalman Filter
- MS 7: Demonstrate Ensemble Kalman Filter with Simulated and In-Situ data
- MS 8: Demonstrate Ensemble Kalman Filter with VLF Whistler and Magnetometer FLR data
- MS 9/D3.2: Begin delivering plasma density maps to WP4

MS 8: PLASMON first assimilation

- Data assimilation for moderate storm in August 2010.

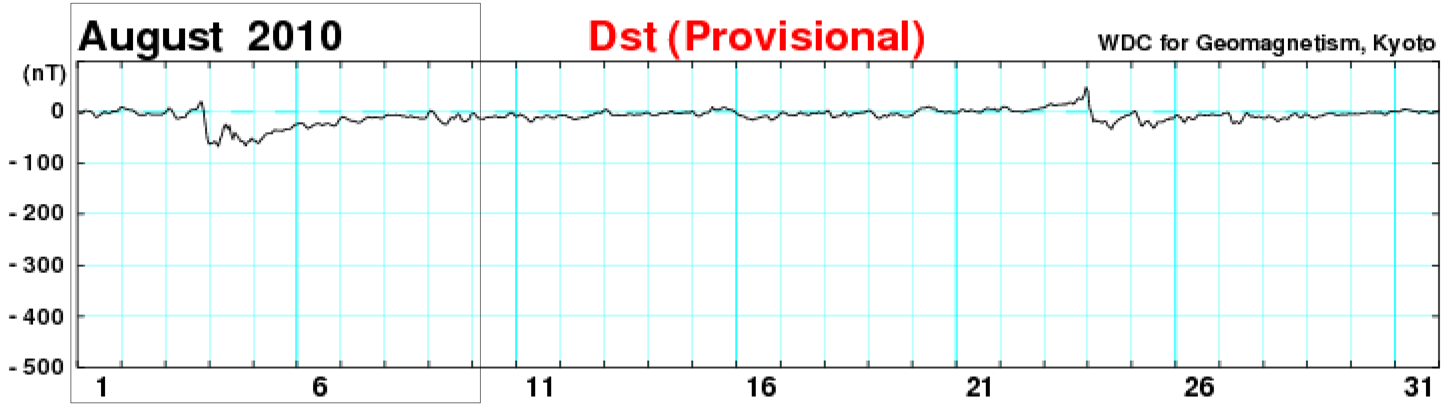
- 8 days of PLASMON data:

  3 FLR station pairs

  - SUW/BEL L=2.18
  - NUR/TAR L=2.24
  - MEK/NUR L=2.27

  1 VLF station

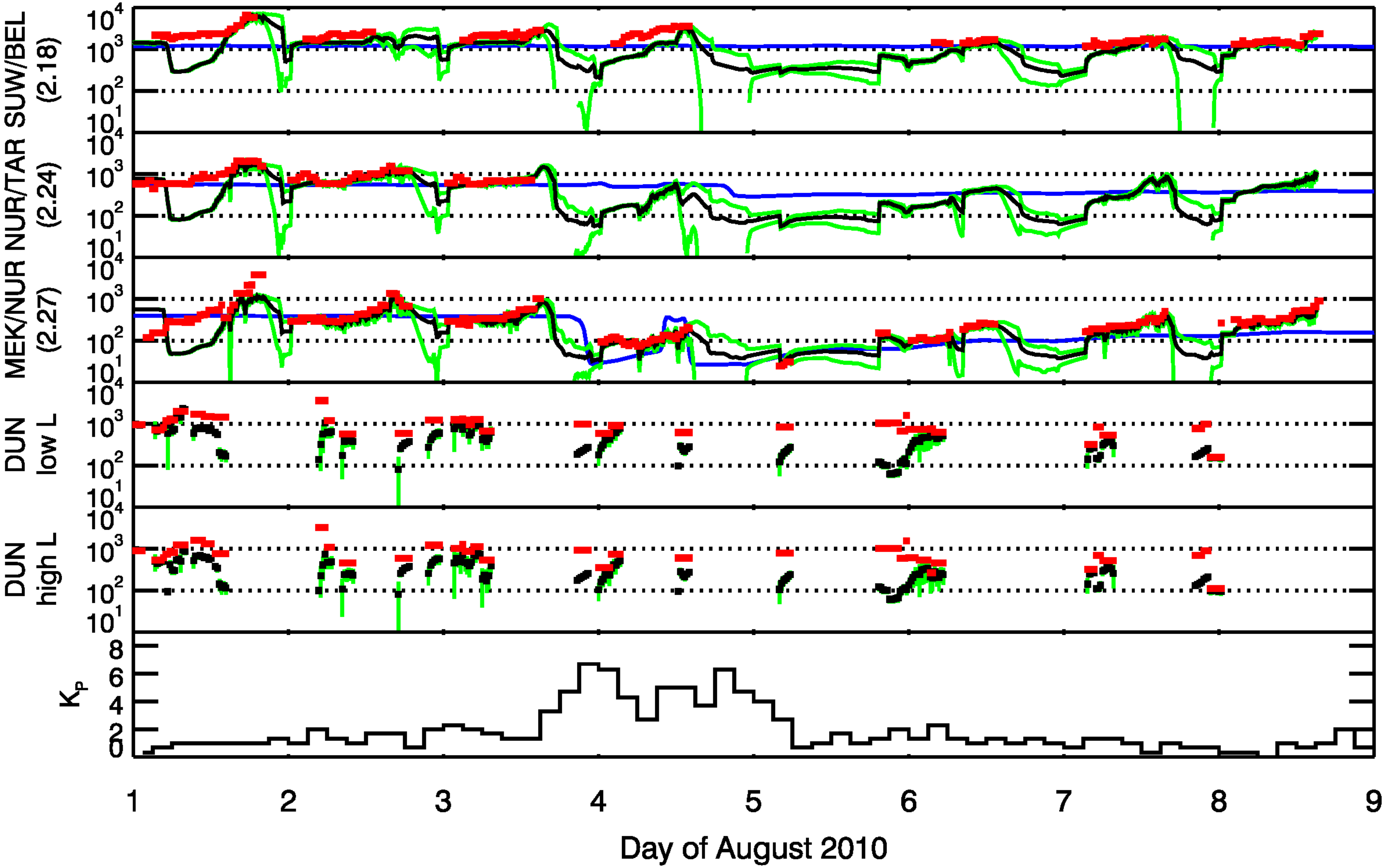
  - Dunedin (New Zealand)



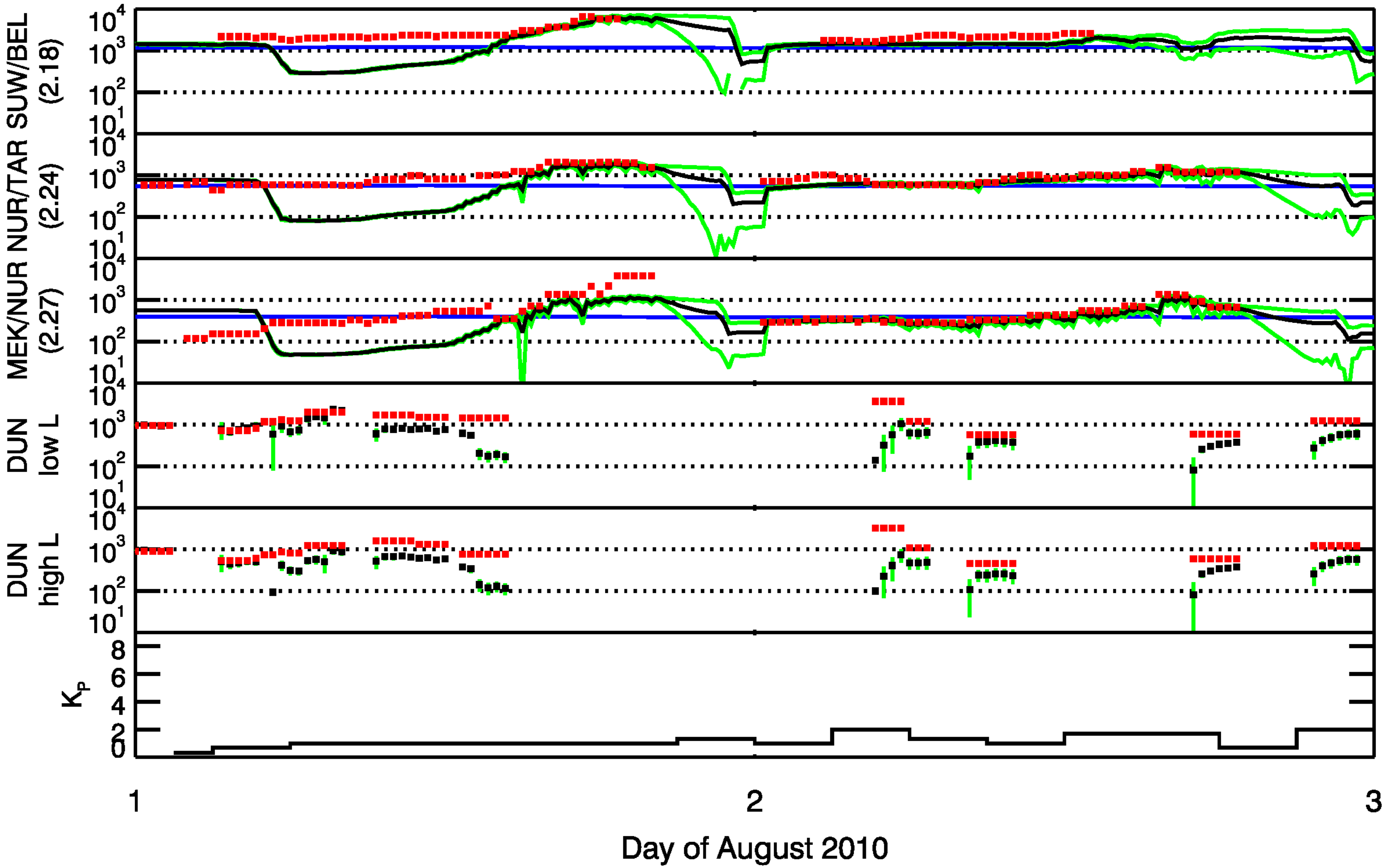
MS 8: PLASMON first assimilation

- Model parameters:
- 4-parameter electric field model
- 3-hour time-constant for each parameter
- Default refilling, decay, and saturation levels

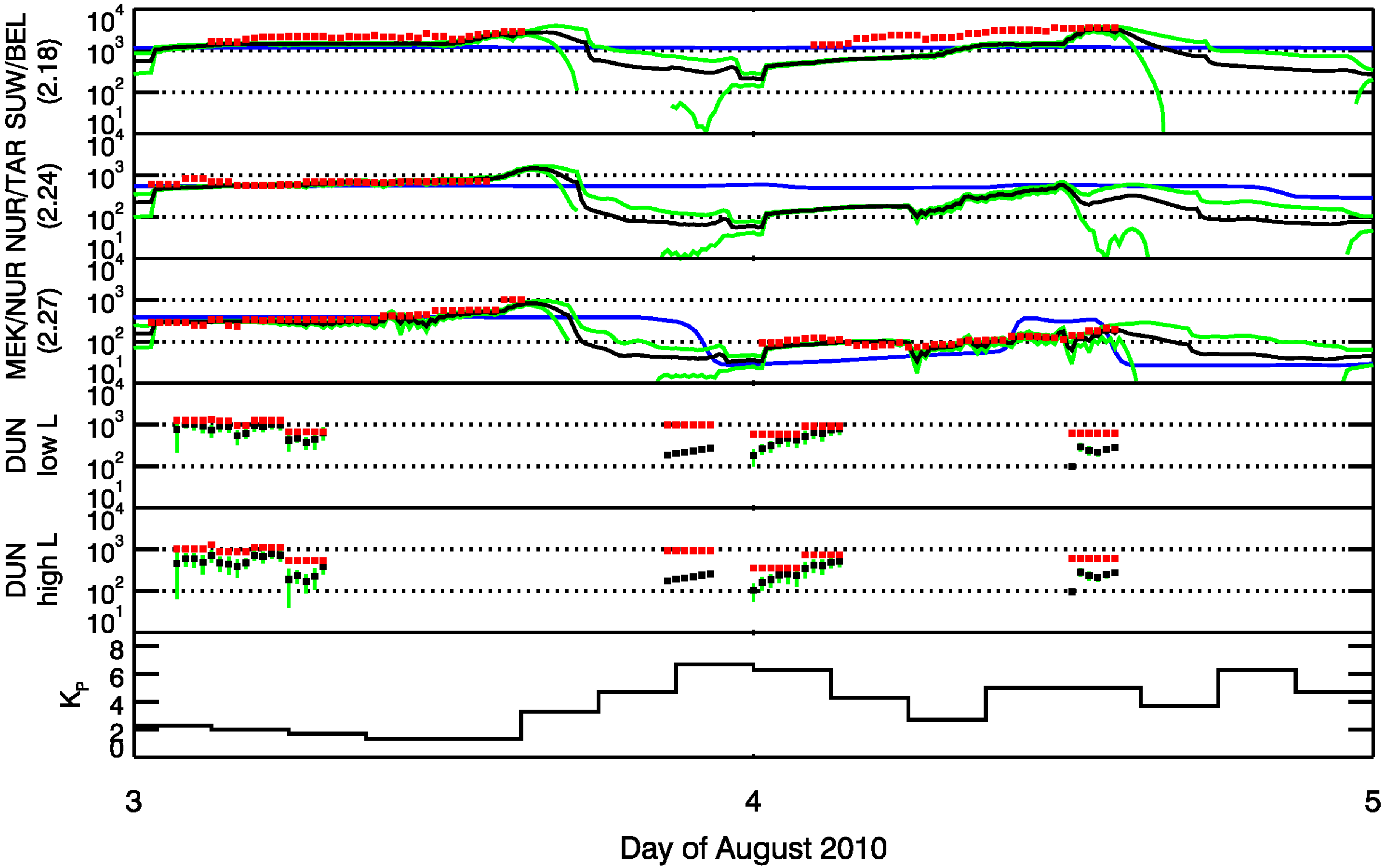
MS 8: PLASMON first assimilation



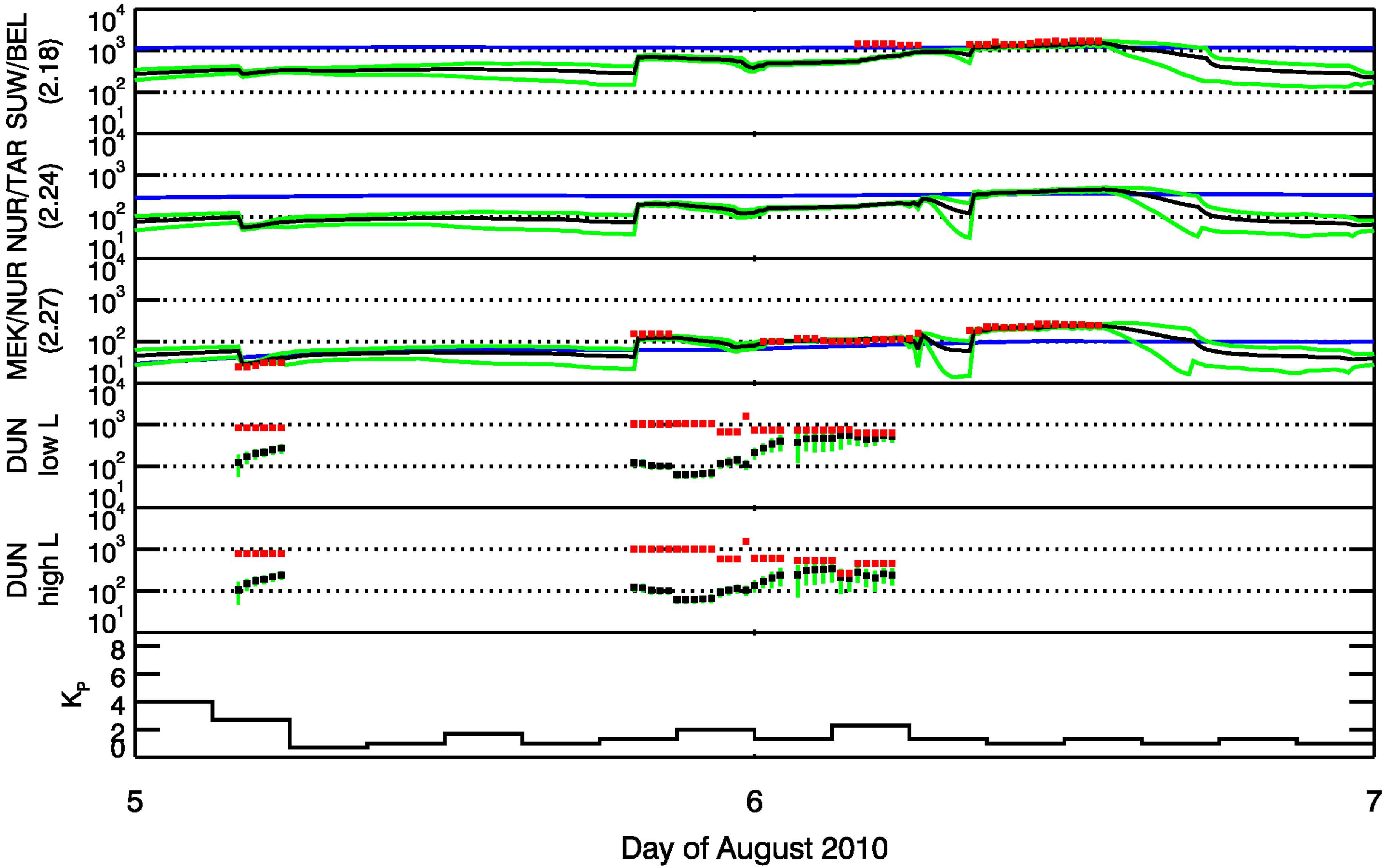
MS 8: PLASMON first assimilation



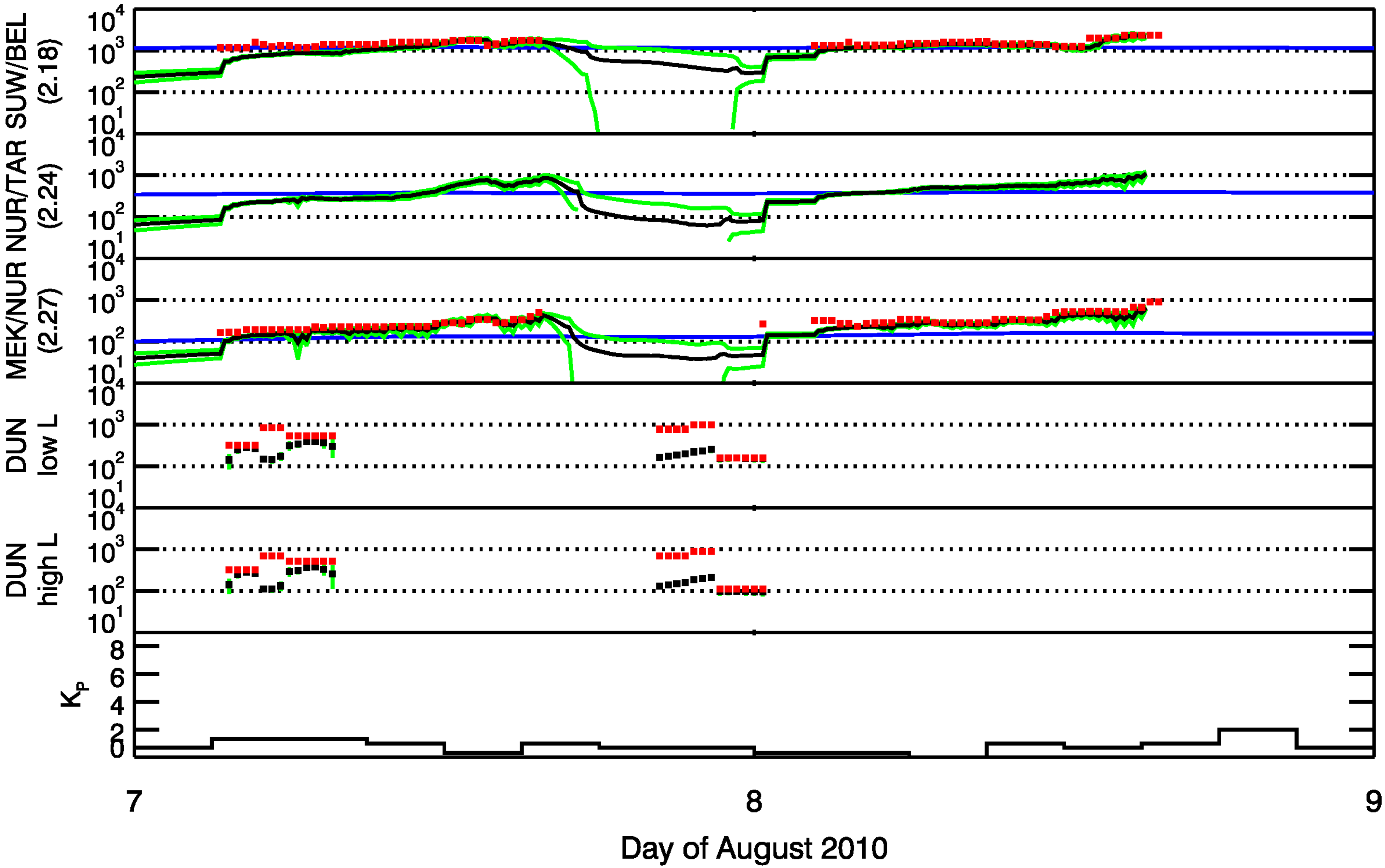
MS 8: PLASMON first assimilation



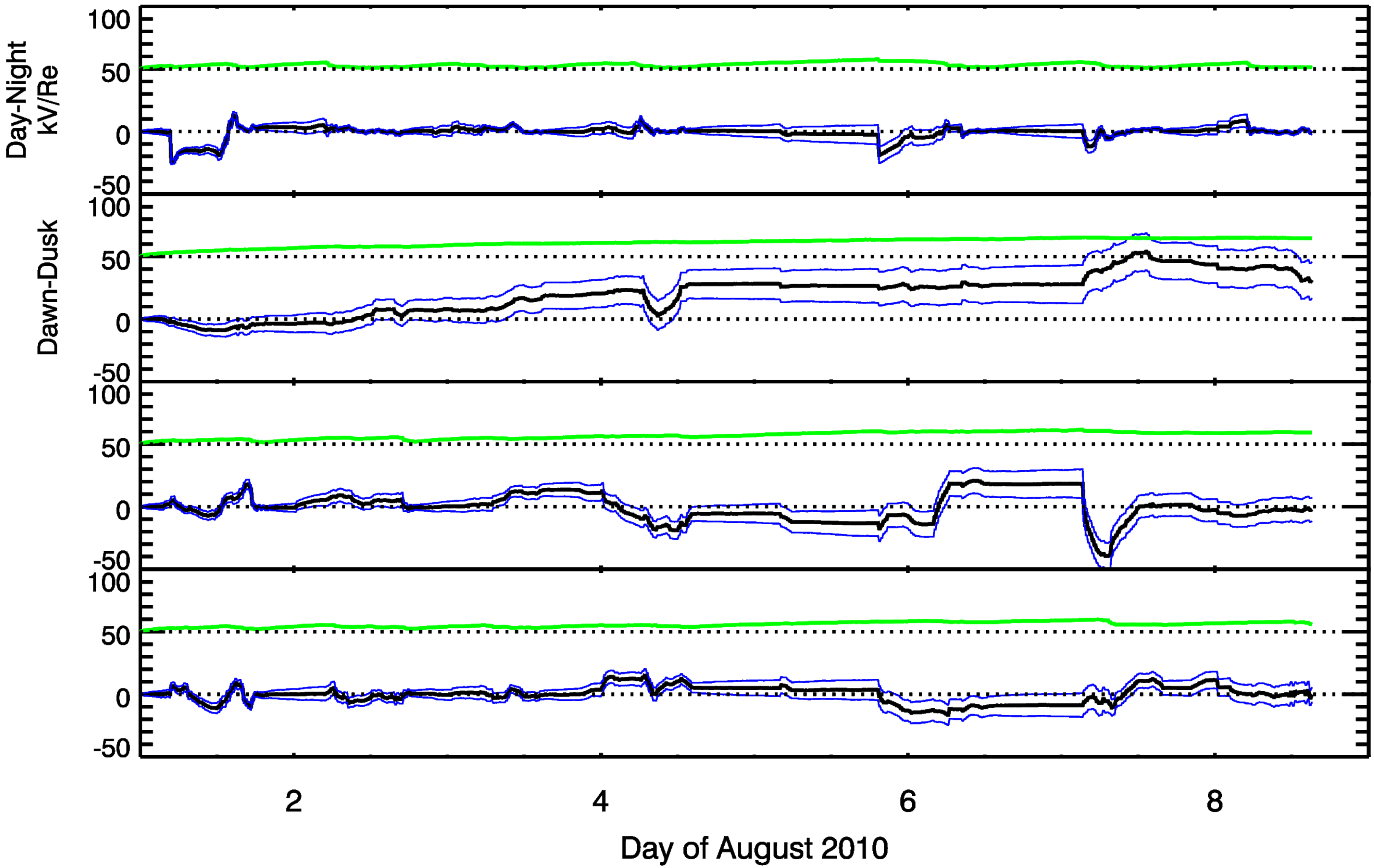
MS 8: PLASMON first assimilation



MS 8: PLASMON first assimilation



MS 8: PLASMON first assimilation

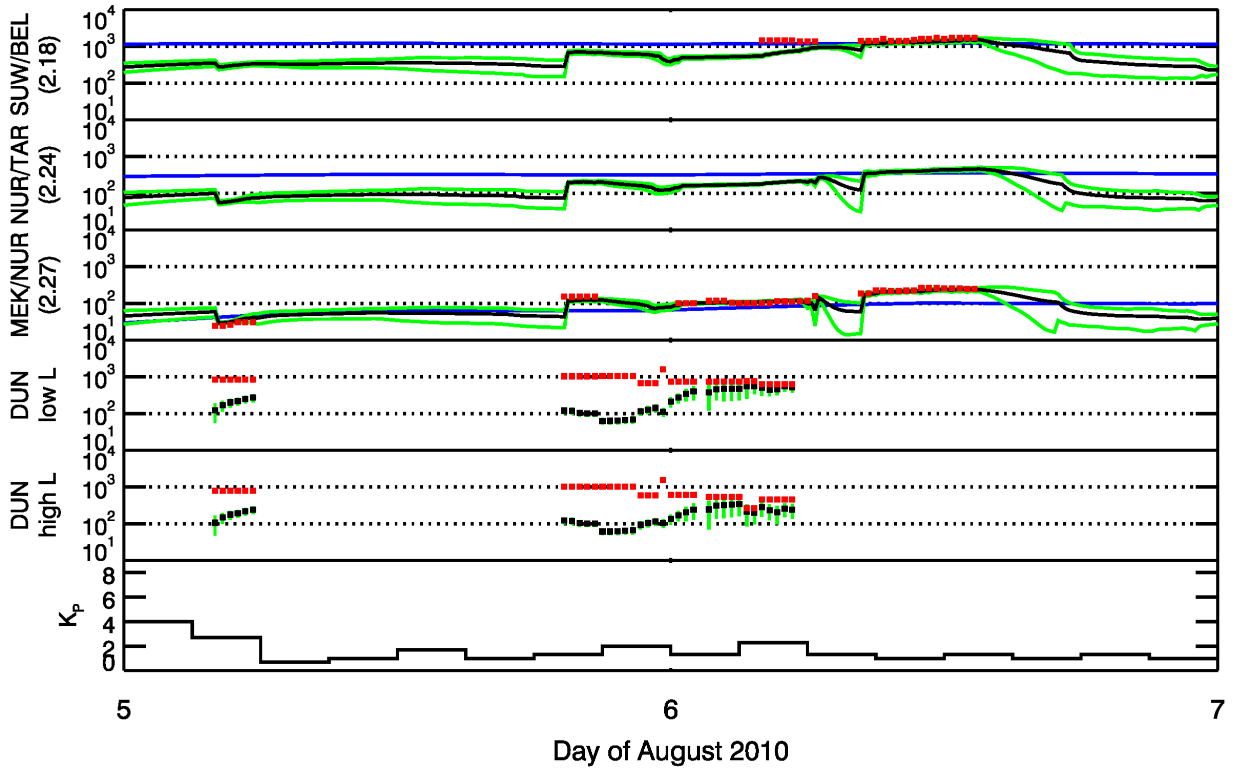


MS 8: Discussion

- When observations present it is relatively easy to make the model agree with the observations, although there are some significant exceptions.
- At the end of observation gaps there are sometimes assimilation model discontinuities
- Perhaps a tendency for FLR to pull model up and VLF to pull the model down. That would be consistent with heavy ion contribution:
  - I assume one electron per AMU conversion between FLR and VLF (H+)
  - With heavy ion composition I need to adjust FLR down or VLF up or both
  - Is there a tentative signature of composition?

MS 8: Discussion (2)

- In this assimilation there were large gaps. These are seen as divergence in the model density and the drivers



- Two ways to improve the model....

MS 8: Improving the model

- Convert to a Kalman Smoother filter
- Use external electric field information

- Both to be discussed in a moment ...

Details of Milestones and Deliverables

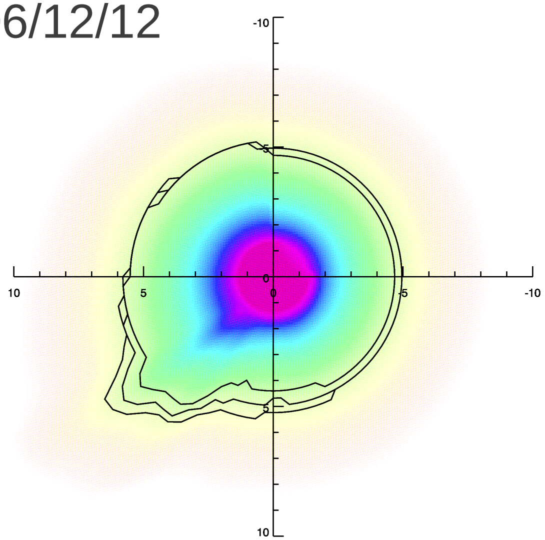
- MS 6 and D3.1: Basic Ensemble Kalman Filter
- MS 7: Demonstrate Ensemble Kalman Filter with Simulated and In-Situ data
- MS 8: Demonstrate Ensemble Kalman Filter with VLF Whistler and Magnetometer FLR data
- MS 9/D3.2: Begin delivering plasma density maps to WP4

MS 9: Delivering Plasma Density Maps

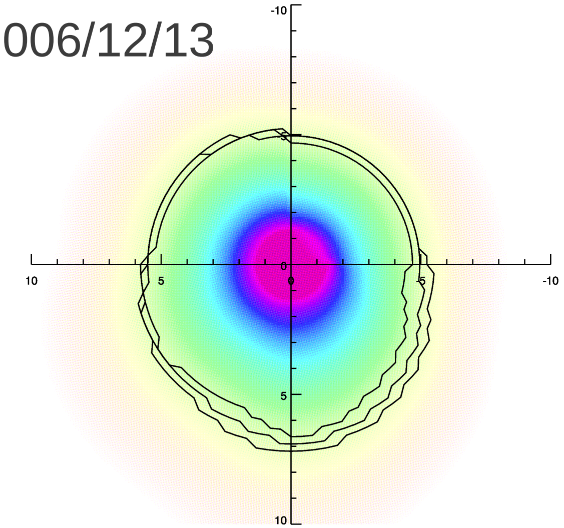
- Assimilation output is plasma density on a grid as well as variance of the plasma density, across the ensemble, on that same grid.
- We can provide these and also use them to determine the location of the plasmapause.

MS 9: Plasma Density Maps from LANL assimilation

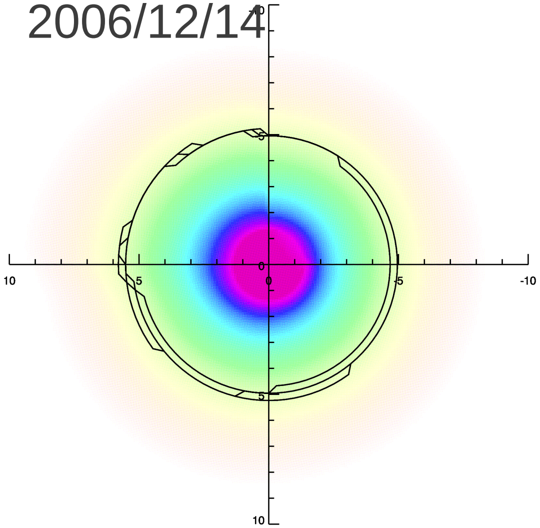
2006/12/12



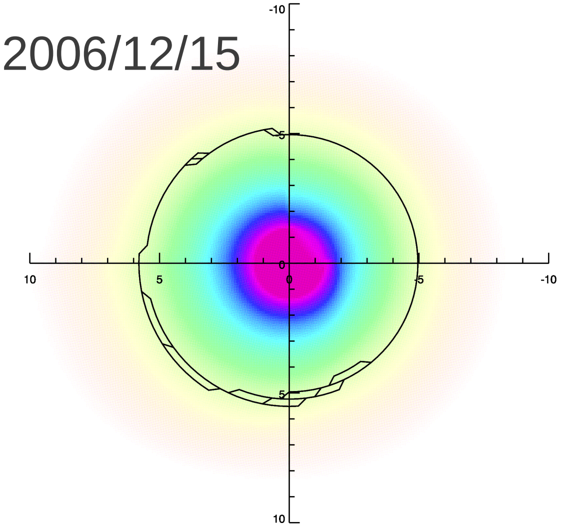
2006/12/13



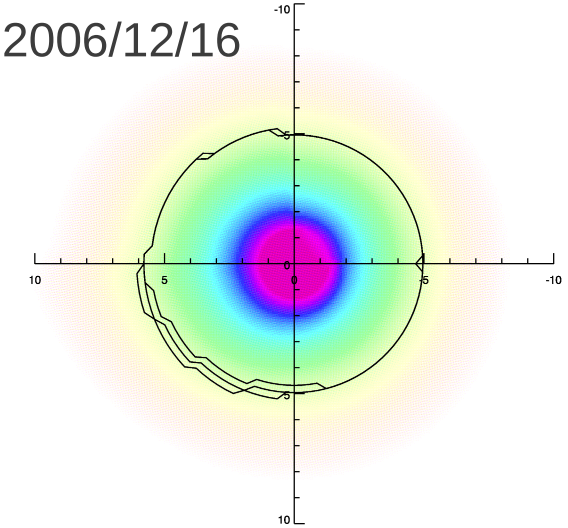
2006/12/14



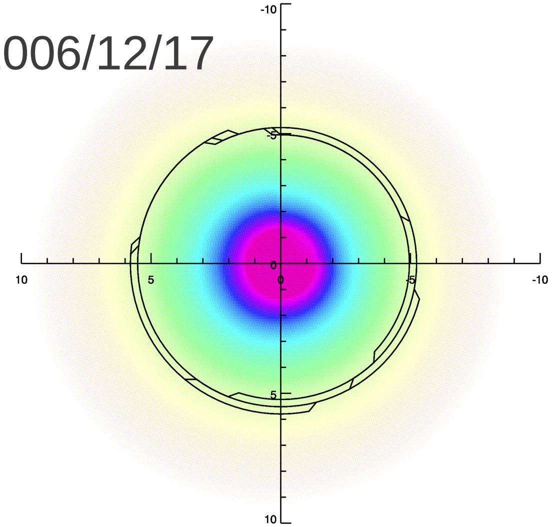
2006/12/15



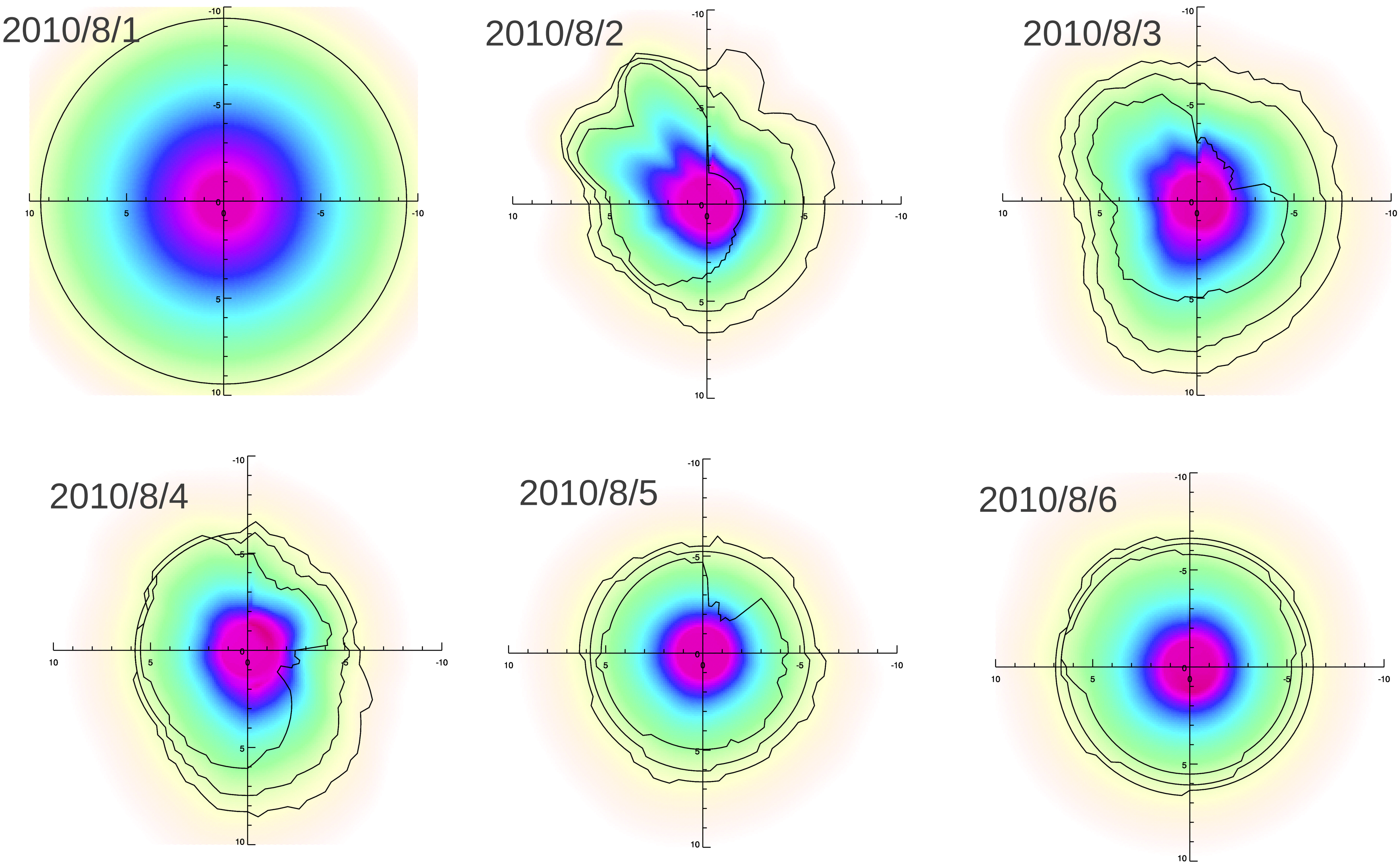
2006/12/16



2006/12/17



MS 9: Plasma Density Maps from LANL assimilation



D3.3 manual

- Description of parameters and settings
- Description of data input format
- Description of model output format
- Not started yet

Outline

- Overview of Milestones and Status
- Overview of Deliverables and Status
- Details of Milestones and Deliverables
- Work Plan for Year 3

Science Results and Lessons Interspersed...

Year 3 Plans for WP 3

- Select PLASMON events and process them
- Enhance assimilation model:
- Incorporate an external electric field model additively. For example Weimer electric field. This will make the external electric field model the default state instead of the zero electric field. Requires solar wind input data stream.
- Switch to a Kalman Smoother filter. This is already incorporated in the code as “history” and associated linear transformations. Just needs to be turned on with a long enough time-history.
- Develop an automated run for near-real time data assimilation with incoming data.

Year 3 Plans for WP 3 (2)

- Automated real-time run methods:
  - Checkpointing and restart when new data are available
    - Preferred appoach
  - Code waits for data
  - Re-run segment of XX days when data become available.
  - Run-time:
    - compact version 1 min per day on 16-thread machine
    - Full version estimated 60 min per day on 16-thread machine.
    - GPGPU still under development
- GPGPU code – new Tesla K20 GPGPU after flood destroyed Tesla C2070. Helpful but not necessary to achieve results.

Year 3 Plans for WP 3 (2)

- Investigate possible difference between VLF and FLR
  - Is there a difference?
  - If so, is it consistent with composition?

Data Format for Ingestion

- Any format that contains the follow information
- Necessary information: Time, density, L, MLT
- Desired information: Density uncertainty
- If I don't get uncertainty I will estimate it. For example 10% of value but no smaller than a minimum value.
- Examples:
  - For FLR I have used 10% uncertainty
  - For VLF I have used 30% uncertainty
  - For LANL In-Situ I have used 10% uncertainty but minimum $3/cm^3$ (to prevent small densities from forcing model)

Outline

- Overview of Milestones and Status
- Overview of Deliverables and Status
- Details of Milestones and Deliverables
- Work Plan for Year 3

Science Results and Lessons Interspersed...