SPACE WEATHER, VOL. 5, S11001, doi:10.1029/2006SW000236, 2007

Branch prediction and speculative execution: A magnetospheric data assimilation scheme for space weather forecasting

I. Doxas,[1,5] W. Horton,[2] J. Lyon,[3] M. Wiltberger,[4] and R. S. Weigel[1,6]

Received 24 March 2006; revised 21 May 2007; accepted 1 June 2007; published 15 November 2007.

[1] Although space weather is modeled after tropospheric weather, both in its conception as a weather system and in our efforts to forecast it, no capability exists today for assimilating magnetospheric data into space weather simulations. In this paper a scheme is proposed for assimilating magnetospheric data into a global MHD code. The scheme is similar to ensemble Kalman filters, but it is less reliant on dense data coverage and allows numerical models easier adherence to conservation laws. Three different estimates of the computational cost of the proposed scheme indicate that it is easily achievable with current computational resources.

Citation: Doxas, I., W. Horton, J. Lyon, M. Wiltberger, and R. S. Weigel (2007), Branch prediction and speculative execution: A magnetospheric data assimilation scheme for space weather forecasting, *Space Weather*, *5*, S11001, doi:10.1029/2006SW000236.

1. Introduction

[2] Recent advances in the development of integrated models of the Sun-Earth environment are placing increasing emphasis on data assimilation schemes that can maximize the intelligence extracted from our sparse sampling of upwind conditions. Standard Kalman filter techniques, widely used in tropospheric weather modeling, require significantly better coverage than is available upwind. To maximize the input of sparse upwind and magnetospheric data and to reduce the forecast lead time computational penalty, we proposed to use branch prediction and speculative execution (BPSE) for data assimilation [*Doxas and Horton*, 2002]. Branch prediction and speculative execution consists of making probabilistic estimates of current upstream conditions and distributing among available machines a large number of simulations that assume each of the probabilistically estimated states as initial conditions. As the near-Earth space evolves and near-Earth satellite data are compared with the models, some of the speculatively executed simulations will be seen to diverge from the observations more than others. At that point the machines that were executing them will be reassigned to new lines of speculative simulation, resulting in a continuous ensemble of runs that are in the neighborhood of the measured values. The scheme is particularly suited to space weather since our upwind early warning sentries can provide only sparse sampling of the incoming solar wind, while the bulk of our monitors, which can provide significantly better coverage, are located close to Earth and provide much shorter lead times. By the time the data come in from the near-Earth monitors, the forecasts of the speculative simulations are already in hand, reducing the lead time computational penalty (the portion of the lead time devoted to advancing the model) to almost zero. The scheme is similar to ensemble Kalman filters but is less reliant on dense data coverage, does not introduce any violation of conservation laws that were not present in the code, and can be used with empirical models without modification.

[1]Department of Physics, University of Colorado, Boulder, Colorado, USA.
[2]Institute for Fusion Studies, University of Texas at Austin, Austin, Texas, USA.
[3]Dartmouth College, Hanover, New Hampshire, USA.
[4]High Altitude Observatory, Boulder, Colorado, USA.
[5]Currently at Tech-X Corporation, Boulder, Colorado, USA.
[6]Currently at Computational and Data Sciences, George Mason University, Fairfax, Virginia, USA.

2. Tropospheric Data Assimilation Schemes

[3] In this section we will review both the theoretical underpinnings and some of the technical schemes used in tropospheric weather data assimilation, and we will situate BPSE in that context.

2.1. Bayes' Theorem

[4] At the heart of all probabilistic forecasting methods lies Bayes' theorem, which correlates the posterior and anterior probabilities of two events. If $P(A)$ is the probability of A occurring, $P(A \cap B)$ is the probability of both A

and B occurring, and $P(A|B)$ the conditional probability of A occurring given that B has occurred, then we can write

$$P\left(A \bigcap B\right) = P(B)\,P(A|B) \tag{1}$$

$$= P(A)\,P(B|A) \tag{2}$$

which leads to Bayes' Theorem:

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)} \tag{3}$$

In other words, the posterior probability of A occurring given that B has already occurred is given as a function of the independent probabilities of A and B, and the conditional probability of B occurring given A. If we think of B as the "observed quantity" and replace events by continuous variables (so that the probabilities are now distributions), then Bayes' theorem becomes

$$p(x|x^o) = \frac{p(x^o|x)\,p(x)}{p(x^o)} \tag{4}$$

which forms the basis of the standard sequential data assimilation methods.

2.2. Cost Function

[5] To develop a data assimilation scheme, Bayes' theorem is combined with an (arbitrary) function of the errors. The minimization of that function (called the cost, or penalty function) will then determine how the predicted values of the observables, x^p, and their error estimate, V^p will combine with the observed values and error estimates, x^o and V^o, to give the new values and error estimates, x^a and V^a.

[6] It can be shown, for instance, that for normal distributions, the prediction that x is near the predicted value x^p with variance V^p combines with the observed value x^o with variance V^o to give the new estimate that x is near x^a with variance V^a by the usual Gaussian error propagation formulas:

$$\frac{x^a}{V^a} = \frac{x^o}{V^o} + \frac{x^p}{V^p} \tag{5}$$

$$\frac{1}{V^a} = \frac{1}{V^o} + \frac{1}{V^p} \tag{6}$$

so that for normal error distributions the standard data assimilation methods are equivalent to a least squares fit with errors propagating in the usual Gaussian manner (equations (5)–(6)).

[7] It is important to note that the initial error estimate, V^p, determines not only the error estimate of the predicted value, V^a, but also the value of the new estimate itself, x^a (cf. equation (5)). So the choice of the cost function does in general affect the forecast. Although distance squared is by far the most common function chosen to be minimized (leading to least squares fits), other functions can also be used [e.g., *Houtekamer*, 1995].

2.3. Kalman Filters

[8] In the context of a numerical model that is used in combination with observations for specification and forecasting (as in tropospheric weather, or in the geospace environment) the observable x becomes the vector of all the grid points in the simulation. The predicted value, x^p, is therefore the vector of all grid points in the simulation, while the predicted variance, V^p, is the vector of the variances at each grid point. This vector of predicted variances is calculated from the results of the previous time step by a matrix (called the covariance matrix) that connects the error at each observation point in the vector x_i to every other point in the vector $x_{j\neq i}$. This calculation is usually done analytically, but for ensemble Kalman filters the covariance matrix is calculated using the statistical spread of an ensemble of runs.

[9] The standard method for data assimilation in tropospheric weather is the Kalman filter [*Kalman*, 1960], which is a fast algorithm for performing the least squares fit and error propagation. For large simulations, $\mathbf{x}^{\mathbf{p}}$, and large observation vectors, $\mathbf{x}^{\mathbf{o}}$, performing the least squares fit and error propagation can become very expensive. For tropospheric weather the vector of observation points, $\mathbf{x}^{\mathbf{o}}$, can be of the order of 10^5 (cf. Figure 1), while the simulations can have 10^6-10^7 points, necessitating calculations with very large matrices. In general, the errors cannot be assumed to be distributed normally, which makes computations even more expensive. Kalman filters make a number of simplifying assumptions (e.g., locality, linearity) that simplify the covariance matrix and make the computations faster, but all Kalman filter techniques are essentially a least squares fit with mostly Gaussian error propagation.

2.4. Relation of BPSE to Tropospheric Schemes

[10] A data assimilation method that relies on a least squares fit has to have adequate observational coverage over all relevant length scales, or risk loosing touch with the "true" solution. Tropospheric weather models for instance are well known to deviate significantly from the truth over the South Pacific, where sampling is sparse. Figure 1 shows a map of observation points for tropospheric weather. In total, there are of the order of 10^5 observation points that are used for data assimilation (cf. the Met Office Web site at http://www.met-office.gov.uk/research/nwp/observations/data_coverage/). A cursory examination of the map shows that a typical weather front spanning half the north-south extent of the US, would be sampled with hundreds of observations. In contrast, in the

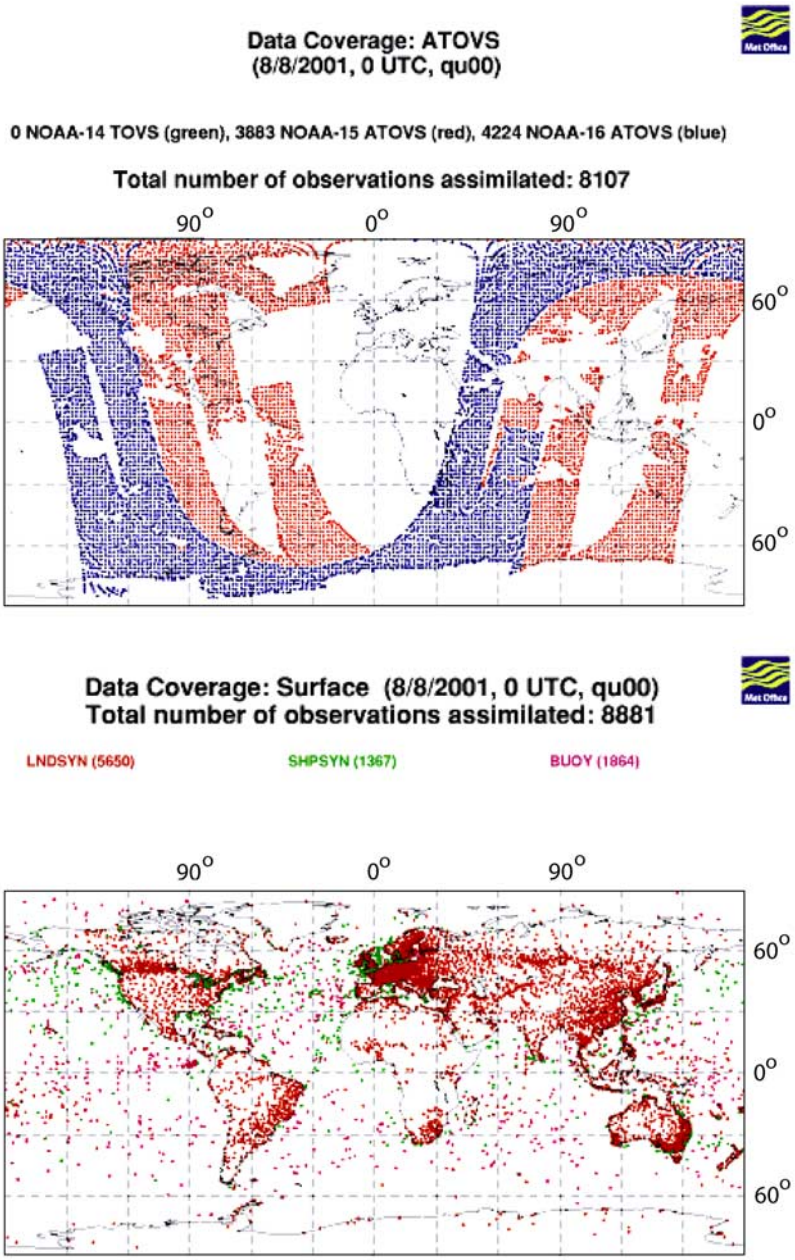

Figure 1. Some tropospheric weather data assimilation points. For more data categories see the Met Web site at http://www.met-office.gov.uk/research/nwp/observations/data_coverage. In total, there are of order 10^5 data assimilation points covering the globe. ATOVS is NOAA's Advanced TIROS Operational Vertical Sounder satellite.

geospace environment we have only a very small number of spacecraft that can warn us of incoming solar wind structures. Global MHD simulations, for instance, will typically assume that the entire sunward front of the simulation (typically $60 \times 60\ R_E$) is uniform, and is given by the measured values at the L1 point. Since typical length scales in the direction perpendicular to the Sun-Earth line are of the order of $20-40\ R_E$ [e.g., *Collier et al.*, 1998, 2000], the incoming front used as input to the global MHD models would typically consist of four uncorrelated regions. There are indications from global MHD simulations that varying the values received from the L1 point can result in better agreement with near-Earth observations (cf. section 3.2 below).

[11] Since it is unlikely in the foreseeable future that we will be able to sample upstream conditions with a density

comparable to that of tropospheric weather, relying on a least squares fit data assimilation scheme is not the best option. We therefore propose that magnetospheric data assimilation schemes use branch prediction and speculative execution (BPSE) to overcome the need for densely sampled upstream conditions. BPSE overcomes that difficulty by augmenting the available data with ''guessing'' of the upstream conditions (on the basis of climatology or other criteria) and postponing the decision of which guesses should form the forecast until more dense near earth data is available.

[12] The method is illustrated schematically in Figure 2. Let us consider the trajectory of the magnetosphere in phase space and let the blue dots at time T_0 represent the initial guesses we make in the vicinity of the actual magnetospheric state. These guesses are then advanced all the way to some time T_2, at which we wish to generate a forecast. During the integration, the trajectories of the initial guesses have spread to cover the volume denoted by blue points at time T_2. At some time $T_1 < T_2$ (typically $\sim$1 h after T_0) the ensemble of speculative runs (blue) is compared to magnetospheric data, and only the runs that are in good agreement with the observations are kept (red points at time T_1). The simulation results which correspond to these points (red points at time T_2) are then used to assemble the forecast at time T_2. By speculatively executing the simulations ahead of the arrival of the near-Earth data, the scheme does not pay a lead time penalty; by the time the near-Earth data is available at time T_1, the simulation results at T_2 are already in hand. The scheme combines features of both variational methods [e.g., *Lorenc*, 1997] and ensemble methods [e.g., *Evensen*, 1994] and proposes to use cost functions that are weighted for relevance to the forecast [e.g., *Houtekamer*, 1995].

[13] The proposed scheme also ensures that the simulations which are finally used to assemble the forecast have maintained conservation laws (like energy conservation) continuously from the point of the solar wind driver to the forecast. Traditional data assimilation schemes that use equations (5)–(6), obviously create a discontinuity in conservation laws at the time of data assimilation, since equations (5)–(6) introduce what, from the point of view of the code, are random ''corrections'' to physical quantities (like plasma velocity and density) which should be obeying conservation laws. This means that in general, the part of the simulation that includes the solar wind driver does not strictly constrain, for example, the energy or momentum in the part of the simulation used in the forecast. By using simulations that have run continuously without arbitrary intervention from T_0 to T_2, BPSE ensures that the results used to assemble the forecast are indeed constrained by the driving conditions through the relevant conservation laws.

[14] It should be emphasized here that the purpose of following an ensemble of runs with different initial conditions is not meant to correct for the variance between any one simulation and the near earth data but rather to account in a more or less systematic way for the errors introduced in the forecast by our imperfect knowledge of the driver. The following scenario might illustrate this point more clearly. Suppose that as envisioned in our search for coupled models, a heliospheric code is coupled to a magnetospheric code by passing values of solar wind parameters on the grid points (y_j, z_k) of an interface plane. In order to institute a tropospheric type data assimilation scheme to this system of coupled codes, the heliospheric code will have to pass not only the values of the parameters, but also estimates of their error, so at position (y_j, z_k) it should be passing (for example) B_{jk} and δB_{jk}. These values would be uniquely defined by the heliospheric code over the entire front, except at the L1 point. At that point the forecast of the heliospheric code, $(B_1, \delta B_1)$, will have to be combined with the (generally different) observation values $(B_2, \delta B_2)$, to give $(B_0, \delta B_0)$ where

$$\frac{B_0}{\delta B_0} = \frac{B_1}{\delta B_1} + \frac{B_2}{\delta B_2} \tag{7}$$

$$\frac{1}{\delta B_0} = \frac{1}{\delta B_1} + \frac{1}{\delta B_2} \tag{8}$$

assuming the standard Gaussian error propagation (equation (5)–(6)).

[15] After a few time steps of the magnetospheric code, this uncertainty in the value of the magnetic field at L1 will have contributed to the uncertainties at nearby grid points, (x_l, y_m, z_n) in ways determined by the differential equations used to advance the system. After a long time, this error at L1 will have contributed a small amount to the error of (in principle) all other grid points in the simulation. In most tropospheric data assimilation schemes the contribution of δB_0 on the errors of all the other grid points in the simulation is calculated analytically and is codified in the covariance matrix. The difference between Kalman filters and BPSE, is that BPSE calculates this contribution in a Monte Carlo way, by following an ensemble of runs through the simulation.

3. Computational Cost Estimates

[16] In order to estimate the number of speculative runs that will be needed for the scheme, we performed a series of tests using WINDMI, a computationally inexpensive dynamical model, and LFM, a computationally more demanding but more realistic MHD model. An estimate of the nonlinear wraparound time (the time it takes the magnetospheric attractor to explore the longest scales [cf. *Goode et al.*, 2001]) also gives similar results.

3.1. WINDMI Tests

[17] Since MHD simulations are computationally intensive we have used WINDMI, a physics-based low-dimensional model for the coupled solar wind–

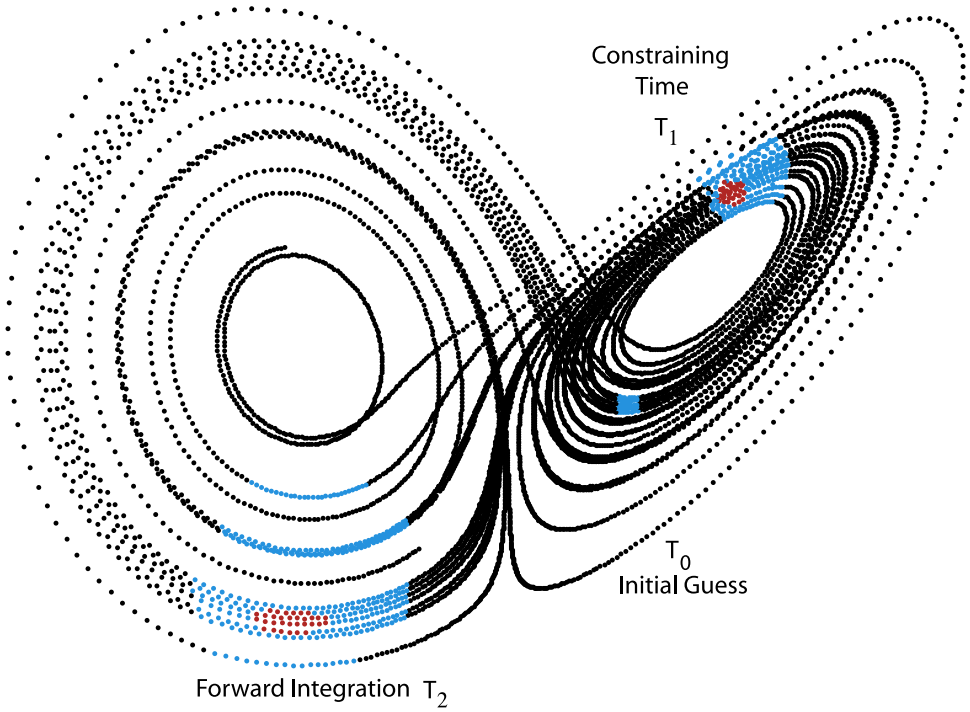

Figure 2. A schematic representation of the divergence of an ensemble of runs. The blue points represent all the runs that were initiated at T_0, while the red points represent only this subset of the runs that was in good agreement with observational data at the data assimilation time T_1. At time T_0 a guess is made as to the variability of the current conditions (blue) owing to the uncertainty in the specification of the sparsely sampled solar wind. All guesses are then advanced to the forecast time (blue points at T_2). At time T_1 incoming magnetospheric data constrain the possible solutions to pass close to the observed conditions (red). In assembling the forecast for T_2 (red), only trajectories that pass near the constraining conditions at T_1 are kept; the others are discarded. The number of speculative runs needed to implement the scheme is determined by the spread of the initial guess between T_0 and T_1, since we need to have a sufficient number of runs covering the constraining area (red points at T_1) to produce a functional forecast (red points at T_2). If the interval $T_1 - T_0$ (typically about an hour) is comparable to the time it takes for the initial ensemble to spread over the entire attractor, then the initial guesses might have to cover the entire attractor in order to assure coverage over the target area at T_1. An integration time short compared to the wrap-around time would mean that the initial conditions (blue points at T_0) would only have to cover a small volume to assure coverage of the target state (red points at T_1). The nonlinear time estimates in section 3.2 indicate that this is the case. The scheme insures that a data-validated forecast is available as soon as the constraining magnetospheric data becomes available at T_1, reducing the computational advance time to zero.

magnetosphere-ionosphere system [*Horton and Doxas*, 1996, 1998] for extensive testing to obtain an estimate of the number of simultaneous speculative runs that would be required in an implementation of the proposed scheme. The parameters used in WINDMI can be calculated from inexpensive particle simulations, which allows us to execute large random searches in parameter space. The results can then guide us in our estimates of the computational resources that are likely to be needed to implement the proposed scheme with realistic global MHD simulations.

[18] WINDMI couples the four basic energy components of the nightside magnetosphere to the ionosphere via the region-1 currents. The model is a six-dimensional 13-parameter system, given by

$$L\frac{dI}{dt} = V_{sw}(t) - V + M\frac{dI_1}{dt} \tag{9}$$

$$C\frac{dV}{dt} = I - I_1 - I_{ps} - \Sigma V \tag{10}$$

$$\frac{3}{2}\frac{dP}{dt} = \Sigma\frac{V^2}{\Omega} - u_0 K_{\parallel}^{1/2} H(I - I_c)P \tag{11}$$

$$\frac{dK_{\parallel}}{dt} = I_{\mathrm{ps}}V - \frac{K_{\parallel}}{\tau} \tag{12}$$

$$L_I\frac{dI_1}{dt} = V - V_I + M\frac{dI}{dt} \tag{13}$$

$$C_I\frac{dV}{dt} = I_1 - \Sigma_I V_I \tag{14}$$

The quantities L, C, Σ, L_I, C_I, and Σ_I are the magnetospheric and ionospheric inductance, capacitance, and conductance, respectively. M is the mutual inductance. The pressure gradient driven current is given by $I_{\mathrm{ps}} = \alpha P^{1/2}(t)$ as derived from force balance and Ampere's law. The parameter α is an average over the pressure profile in the current sheet. The solar wind driving voltage V_{sw} in equation (7) is the input time series for this nonlinear driven dissipative system. The term $u_0\ K_{\parallel}^{1/2}H(I\ -\ I_c)P$ in equation (9), where H is the Heavyside step function, represents the rapid unloading of the stored energy when the current exceeds a critical value, $I > I_c$, and comes from the heat flux limit that is neglected in the MHD closure [*Horton and Doxas*, 1996]. In the absence of driving ($V_{\mathrm{sw}} = 0$) and damping ($\Sigma_I = 0$) and below the unloading limit ($I < I_c$), the total energy is conserved. All parameters of the model can be calculated explicitly from their integral definitions given the plasma conditions in the tail.

[19] For WINDMI to be used as a forecasting tool, the parameters of the model need to be calculated. This can be done explicitly from the plasma parameters in the tail (density, temperature, etc.) since all WINDMI parameters are defined as integrals of physically identifiable quantities (the capacitance, C, for instance is the volume integral of the $E \times B$ kinetic energy in the current sheet; the unloading parameter, u_0, is the surface integral of the heat flux over the boundary, etc.). Absent in situ data, simulations can also be used to make estimates of the values of these parameters, and this is the avenue that has been pursued so far. Particle simulations are used to provide the moments of the distribution function, which in turn can be used to calculate the parameters of the model using their integral definitions. The code advances test particles in model fields that consist of a constant northward field and a time-dependent tearing mode perturbation superimposed on a standard Harris sheet [*Horton and Doxas*, 1996]. The particles are advanced for approximately 20 min real time equivalent, and the moments of the particle distribution function are used to calculate the model parameters. The output of the model is then compared to two widely used data sets of the *AL* activity index which are complementary in nature, *Blanchard and McPherron* [1993] (a data set of 121 events that are carefully selected to include only well-isolated substorms) and *Bargatze et al.* [1985] (a set of 34 intervals each of which includes multiple substorm events). The *AL* magnetospheric activity index is a measure of the strength of the westward electrojet and has been extensively used in dynamical studies of the magnetospheric system. We consistently find that when using this procedure, WINDMI produces good agreement with observations both for the isolated and for the multiple events [*Doxas et al.*, 1999, 2002].

[20] Figure 3 shows the results for two of the events in the *Blanchard and McPherron* [1993] database of isolated events that were calculated using parameters obtained from particle simulations (the bottom frames show the worst fits for the corresponding event, for comparison). Similarly good agreement is obtained for all the events in the *Blanchard and McPherron* [1993] data that we have simulated [*Doxas and Horton*, 1999; *Doxas et al.*, 2002] as well as most of the intervals in the *Bargatze et al.* [1985] data [*Horton and Doxas*, 1998; *Doxas and Horton*, 1999]. It is worth noting that including the heat flux term gives good agreement with the measured *AL* index, while the predictions of the model without the heat flux limit (MHD closure) give an unphysical large spike (top left frame of Figure 3). This is due to the fact that without a heat loss term the system (7)–(12) has no mechanism for cooling the plasma sheet plasma nonadiabatically, resulting in an unphysical build-up of energy.

[21] A typical simulation used to obtain the above results has approximately 1,000,000 particles, which is sufficient to provide reliable values of higher-order moments like the heat flux. The event shown in the left frames of Figure 3 required a total of 27 randomly initialized simulations to arrive at good agreement, the event on the right required a total of 38. Although the simulations used in this case are less demanding computationally than global MHD simulations, running an ensemble of 30–40 MHD simulations at low resolution is within current capabilities. This result is, of course, no guarantee that good agreement can be achieved with a tractable number of MHD simulations, but it is encouraging.

3.2. Global MHD Simulations

[22] In addition to the extensive testing that we performed with the computationally inexpensive WINDMI model, we have also performed a smaller number of test runs with LFM in order to obtain a more realistic estimate of the number of speculative LFM runs that would need to be executed. We performed tests for three different scenarios; a steady solar wind, a pulse, and a slow rotation of the magnetic field by 90°. In all three scenarios we executed multiple runs by changing the magnitude of the incoming solar wind velocity from 5% (of maximum variability) below to 5% above the reference value in steps of 1%. In all three cases we measure variability of any parameter in terms of the maximum observed variability

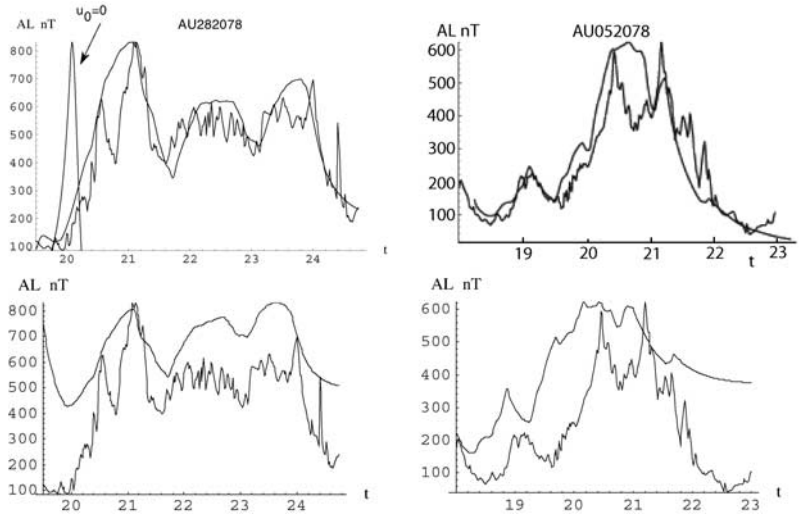

Figure 3. Measured *AL* index and the output of the WINDMI model for two events in the *Blanchard and MacPherron* [1993] data set. The top frames give the best fit and the bottom frames give the worst fit for comparison. When the thermal flux term is neglected (top left frame; $q = 0 \Rightarrow u_0 = 0$) the model gives an unphysical short spike followed by large-amplitude ringing (the ringing part has been removed from the figure for clarity). Using particle simulations to obtain the particle distribution function, and then using the integral definitions of the model parameters [*Horton and Doxas*, 1996, 1998] to calculate the parameters from the particle distribution, results in good agreement between the WINDMI output and observations [*Doxas and Horton*, 1999; *Horton et al.*, 1999; *Doxas et al.*, 2002]. The event shown on the left required 27 randomly initialized simulations to obtain good agreement, the event on the right required 38.

of that quantity, in order to have an absolute measure of the fraction of physically relevant parameter volume that we need to explore. For instance, we estimate the maximum variability of the solar wind speed to be of the order of 1000 km/s so the incoming speed in our numerical experiments was varied over a range of 100 km/s in steps of 10 km/s. The value of 1000 km/s is obviously approximate (solar wind speeds as high as 2000 km/s have been observed) but gives a reasonable estimate of the range of values usually encountered. Similarly, the other maximum variabilities we use (40 nT for the maximum variability of the solar wind B_z for example, and 20 nT for the maximum variability of the current sheet B_z) are also approximate and are only meant to give a useful estimate of the volume of physically relevant parameter space that we are exploring.

[23] We then compared the output of the 10 perturbed runs to the output of the reference run after 1 h of real time. One hour is the typical lead time from our upwind sentries, so the divergence of the code after 1 h would be typical of the expected divergence in our speculative runs. Figure 4 shows the value of δB_z on the $x-y$ plane for the impulse scenario and for a 2% increase of the solar wind speed of the reference run. The color scale gives $\delta B_z/B_{z-var}$, where B_{z-var} is the maximum variability of the northward component in the current sheet, estimated at $\sim$20 nT. We see that for the most part, in the critical area around the center of the current sheet, a 2% variance of the incoming solar wind results in approximately a 10% variance in the result after 1 h real time. This result is similar in nature to the results obtained with the WINDMI model and indicates that we can cover an extensive part of the expected variability at T_2 with only a modest number of speculative runs.

[24] The sort of effects that can be expected by using an ensemble of solar wind inputs for the global MHD codes can be seen in the global MHD simulations performed for the GEM Substorm Challenge [*Slinker et al.*, 2001]. The LFM code was driven by solar wind data from Wind which had a sequence of relatively constant IMF: first northward, then southward, then northward again. During this period, Polar magnetometer data was available for comparison with the model results. The simulation results showed a consistent deviation from the observations in B_x (cf. Figure 7 of *Slinker et al.* [2001], reproduced in Figure 5). Given Polar's orbit high in the northern hemisphere, this indicated that the reconnected magnetic field in the simulation was too strongly draped where Polar was located. For unrelated reasons (to try to change the time of substorm onset in the simulation), an experiment was performed where the IMF B_z from Wind was reduced by approximately 2 nT (about 5% of maximum solar wind B_z variability) after the southward turning. The effect at Polar was to bring the simulation B_x almost totally in agreement

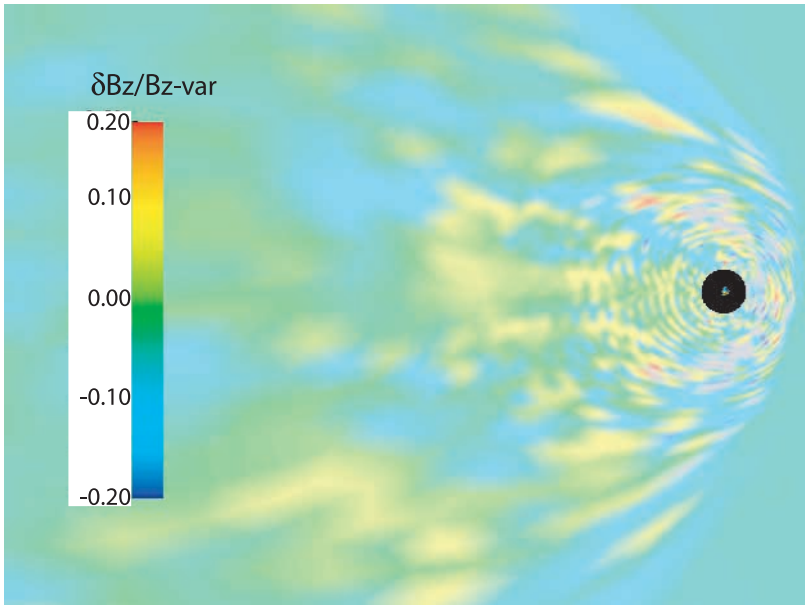

Figure 4. The B_z variability as a fraction of maximum observed variability, in the output of LFM for a 20 km/s (2% of maximum nominal variability of 1000 km/s) change in the solar wind input velocity. The figure shows the $z = 0$ plane, and the nominal maximum variability of B_z on the plane was taken to be 20 nT. We see that a 2% change in the driver results in a change of approximately 10% (yellow) in the critical region around the $y = 0$ line where an X-line might appear. The concentric rings in the near-Earth region are for the most part a numerical effect caused by the nonlinear switches used in the LFM to control high-frequency noise in the simulation. The switches miss high-frequency (in space) noise that only causes ripples on the general decline of the magnetic field outward. These ripples are different for different runs, giving the concentric pattern. We should note that in this inner region the unperturbed magnetic field is generally much larger than the estimated Bz-var, and so the actual relative errors are much smaller than the figure would indicate at first glance.

with observation for approximately an hour and a half (although the B_z value is slightly worse than the original run, cf. Figure 5). This result is, of course, no guarantee that the proposed scheme can be implemented with similar success in other cases, and it indicates that changing a single parameter (the solar wind velocity) will not necessarily improve the results in every respect; however, the experiment does show that changing the solar wind inputs in ways that are within the uncertainty associated with an L1 monitor can materially affect the modeled results and bring model and observations closer together.

[25] The results described above imply that a relatively small number or speculative runs can adequately cover a meaningful volume of the relevant parameter space at the data assimilation time, T_1, and that mangnetospheric data can make clear distinctions between runs that differ by a few percent in one of the driving parameters. An initial ensemble covering the solar wind velocity, density, and B_z with three to four different values each, over a range of $\sim\pm5\%$ of their respective maximum variability (in steps of approximately 2.5%−3.5%) can therefore ensure that one or more of the runs will be in the vicinity of the actual magnetospheric state and allow the data to clearly distinguish between the different runs. Depending on the exact resolution in parameter space (e.g., $3 \times 3 \times 4$) this can easily fit in the range of 30−40 speculative runs, which is comfortably within current capabilities. Furthermore, it is not unreasonable to believe that given an operational system, we will be able, with experience, to make guesses that are somewhat more educated than a systematic coverage of the parameter space.

3.3. Nonlinear Time Estimates

[26] Once a number of speculative runs have been initiated, they will diverge from each other at a rate defined by the nonlinear dynamical properties of the magnetotail (see Figure 2 for a schematic representation) and they will eventually spread over a significant percentage of the entire attractor (e.g., blue points in Figure 2). The number of speculative runs needed to implement the scheme is therefore determined by the spread of the initial guess between T_0 and T_1, since we need to have a sufficient number of runs covering the constraining volume

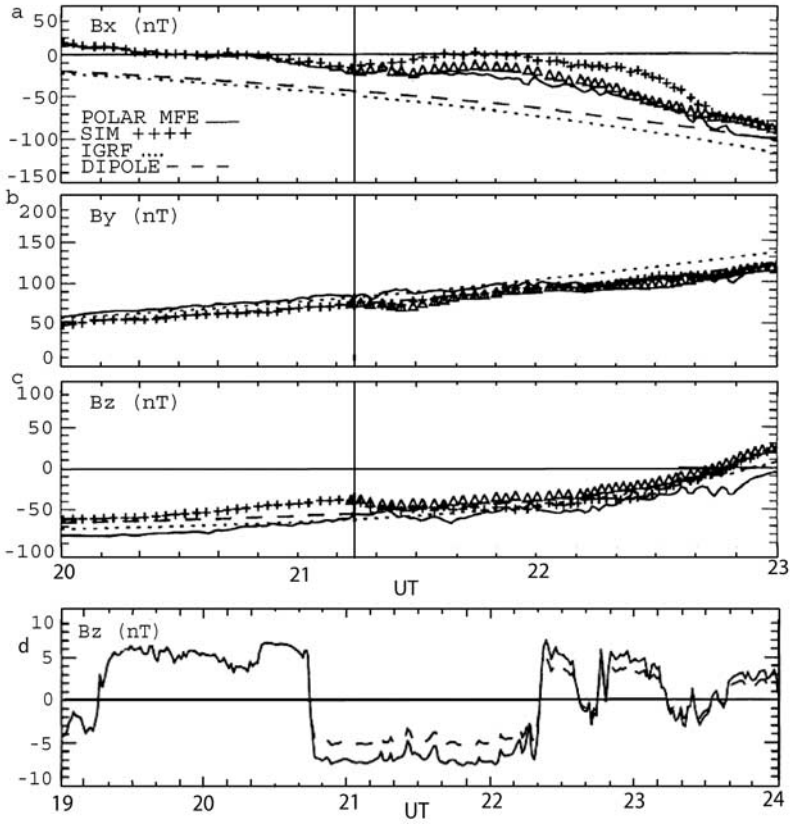

Figure 5. (a)–(c) Comparison of the LFM code results with Polar Magnetic Field Experiment (MFE) measurements. The MFE value is the solid line. The dotted and dashed lines are the International Geomagnetic Reference Field (IGMF) and dipole field, shown for comparison. The crosses show the simulation values, while the triangles show the values of the second simulation with the incoming IMF B_z reduced by 2 nT from the L1 data. (d) Also shown is the value of the incoming solar wind B_z, as well as the value used for the alternative run (dashed). This represents a change of 5% of the nominal maximum variability of 40 nT for the solar wind B_z. The new run shows better agreement with the MFE B_x. From *Slinker et al.* [2001]. The vertical line corresponds to the time at which the alternative simulation was initiated (the time difference is due to the time lag between the L1 point and the magnetospheric response). In a BPSE context the results to the right of the line would be the forecast, which is in better agreement with observations than the reference run for about an hour and a half.

(red points at T_1) to produce a functional forecast (red points at T_2).

[27] The volume over which an ensemble of orbits will spread after a certain time depends on the average rate of divergence of nearby orbits. This rate of divergence is measured by the Lyapunov time. The Lyapunov time is the inverse of the Lyapunov exponent which is given by [*Lyapunov*, 1907; *Benettin et al.*, 1976, 1980]

$$\gamma_L = \frac{1}{\tau_L} = \lim_{\substack{t\to\infty \\ d(0)\to 0}} \left(\frac{1}{t}\right) \ln \frac{d(x_0,t)}{d(x_0,0)} \tag{15}$$

so that orbits which were a distance d_0 apart at time t_0 will be separated at later times, t, by a distance that goes as $e^{-\gamma_L(t-t_0)}$. A Lyapunov exponent is defined for each orthogonal direction at the point x_0, but the largest exponent will, of course, quickly dominate. Lyapunov times have recently been estimated for magnetospheric time series by *Athanasiou et al.* [2003], but for the present problem we are more interested in the fill-in time, the time over which an ensemble of runs will spread to cover a macroscopic part of the attractor, thus requiring the initial guesses to cover a large part of the attractor and rendering the proposed scheme prohibitively expensive. A measure of this time is the wrap-around time, the time it takes the magnetospheric attractor to explore the longest scales, which is obviously a lower bound for the timescale of interest (the distance between two nearby trajectories cannot become comparable to the largest scale of the

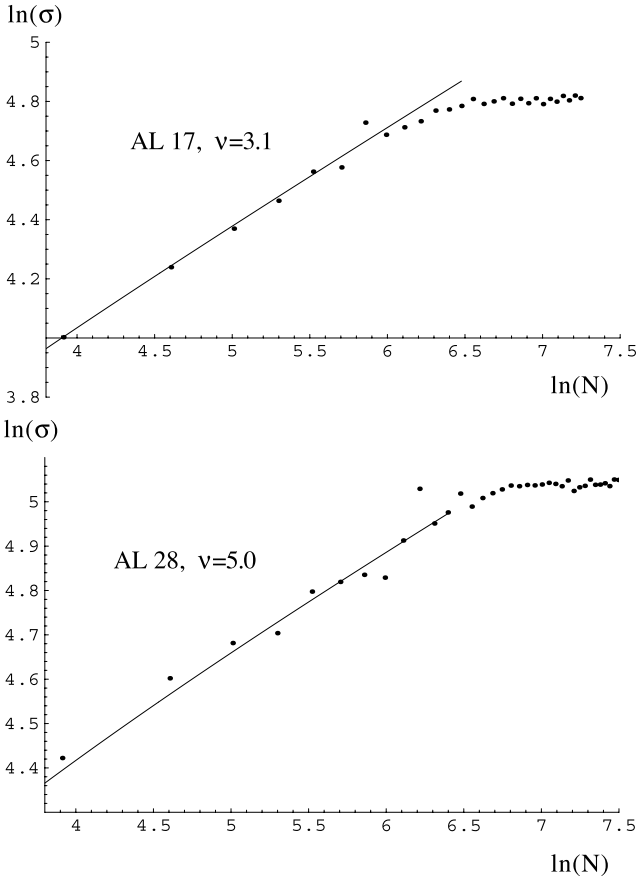

Figure 6. Variance growth test for auroral electrojet (*AL*) index for two of the intervals (interval 17 and interval 28) in the *Bargatze et al.* [1985] data set. Of the 11 intervals that are long enough for the test, nine exhibit this behavior, saturating after approximately 10–17 h (the cadence of the Bargatze data set is 2.5 min, so $\ln(N) \sim 6$ corresponds to approximately 17 h). See *Goode et al.* [2001]. The *AL* index is a measure of the westward electrojet and is often used in the study of the dynamics of the magnetosphere.

system faster than the time it takes a typical trajectory to reach those scales).

[28] An estimate of the wraparound time was recently obtained by *Goode et al.* [2001] on the basis of the variance growth of *AL* time series. For a dynamical system, a time series of an observable that is longer than the wraparound time is stationary, that is, the variance of a subset of length N of the time series

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2} \tag{16}$$

is constant as a function of N. For times shorter than the wraparound time, the variance grows linearly with record length N until the record length becomes comparable to the wraparound time, at which point the variance saturates to the long-time limit. Adding noise to the dynamical system can change the growth rate of the variance in the linear regime, but the variance will still saturate to the long-time limit for record lengths comparable to the wraparound time. Figure 6 shows the variance test for two intervals in the *Bargatze et al.* [1985] data set. We see that the variance saturates for record lengths longer than approximately $\ln(N) \sim 5.5-6$, which corresponds to a wraparound time of 10–17 h (the cadence of the data is 2.5 min).

[29] Nine of the 11 intervals in the Bargatze data set exhibit that behavior [*Goode et al.*, 2001], suggesting that the time series represent a dynamical process (albeit with significant noise) with a typical wraparound time of 10–17 h. Since the lead time from the L1 point is of the order of 1 h, the integration time for the speculative runs is at least an order of magnitude shorter than the wraparound time for the magnetospheric dynamics as exhibited by the data set, which suggests that the initial guesses need only cover a small volume of phase space. This is consistent with the estimates obtained with WINDMI and with LFM, which suggests that the scheme needs only a modest number of speculative runs to cover the target area in phase space.

4. Conclusions

[30] Branch prediction and speculative execution consists of making probabilistic estimates of the true state of the solar wind and following an ensemble of simulations with the different solar wind states as input conditions. As data come in from the much more densely sampled near-Earth space, some simulations will be in better agreement with the data than others, so that at each time a forecast needs to be produced, it can be produced using the runs that best agree with the data.

[31] The critical point for the usability of the scheme as a forecasting tool is the number of speculative runs that will need to be simultaneously followed for the ensemble to have sufficient coverage in phase space to cover the "true" state as it is described by the near-Earth data. We have presented three separate estimates of the number of such speculative runs that an operational system would require, one theoretical and two empirical, and all three estimates indicate that the target phase space can be covered with a moderate (30–40) number of speculative runs. This is a tractable number even for a modest present-day Linux cluster, which argues that the scheme would be eminently feasible even today. More detailed studies will have to focus on the manner in which the ensemble runs are weighted to produce the forecast and on the algorithms that guess the initial conditions in an optimal way as to minimize the numbers of speculative runs required.

References

Athanasiou, M. A., G. P. Pavlos, D. V. Sarafopoulos, and E. T. Sarris (2003), Dynamical characteristics of magnetospheric energetic ion time series: Evidence of low dimensional chaos, *Ann. Geophys.*, *21*, 1995.

Bargatze, L. F., D. N. Baker, R. L. McPherron, and E. W. Hones Jr. (1985), Magnetospheric impulse response for many levels of geomagnetic activity, *J. Geophys. Res.*, *90*, 6387.

Benettin, G., L. Galgani, and J. M. Strelcyn (1976), Kolmogorov entropy and numerical experiments, *Phys. Rev. A*, *14*, 2338.

Benettin, G., L. Galgani, A. Giorgilli, and J. M. Strelcyn (1980), Lyapunov characteristic exponent for smooth dynamical systems and hamiltonian systems: A method for computing all of them, *Mechanica*, *3*, 21.

Blanchard, G. T., and R. L. McPherron (1993), A bimodal representation of theresponse function relating the solar wind electric field to the AL index, *Adv. Space Res.*, *13*, 71.

Collier, M. R., J. A. Slavin, R. P. Lepping, A. Szabo, and K. Ogilvie (1998), Timing accuracy for the simple planar propagation of magnetic field structures in the solar wind, *Geophys. Res. Lett.*, *25*, 2509.

Collier, M. R., A. Szabo, J. A. Slavin, R. P. Lepping, and S. Kokubun (2000), IMF length scales and predictability: The two length scale medium, *Int. J. Geomagn. Aeron.*, *2*, 3.

Doxas, I., and W. Horton (1999), Magnetospheric dynamics from a low dimensional nonlinear dynamics model, *Phys. Plasmas*, *6*, 2198.

Doxas, I., and W. Horton (2002), Using branch prediction and speculative execution to forecast space weather, paper presented at Geomagnetic Environment Modeling Conference, Natl. Sci. Found., Telluride, Colo.

Doxas, I., W. Horton, and J. P. Smith (1999), A physics based nonlinear dynamical model for the solar wind driven magnetosphere–ionosphere system, *Phys. Chem. Earth*, *94*, 67.

Doxas, I., W. Horton, and R. Weigel (2002), Using particle simulations for parameter tuning of dynamical models of the magnetotail, *J. Astrophys. Sol. Terr. Phys.*, *64*, 633.

Evensen, G. (1994), Sequential data assimilation with a nonlinear quasigeostrophic model using Monte-Carlo methods to forecast error statistics, *J. Geophys. Res.*, *99*, 10,143.

Goode, B., J. R. Cary, I. Doxas, and W. Horton (2001), Differentiating between colored random noise and deterministic chaos, *J. Geophys. Res.*, *106*, 21,277.

Horton, W., and I. Doxas (1996), A low-dimensional energy-conserving state space model for substorm dynamics, *J. Geophys. Res.*, *101*, 27,223.

Horton, W., and I. Doxas (1998), A low-dimensional dynamical model for the solar wind driven geotail-ionosphere system, *J. Geophys. Res.*, *103*, 4561.

Horton, W., J. P. Smith, R. Weigel, C. Crabtree, I. Doxas, B. Goode, and J. Cary (1999), The solar-wind driven magnetosphere–ionosphere as a complex dynamical system, *Phys. Plasmas*, *6*, 4178.

Houtekamer, P. L. (1995), Methods for ensemble prediction, *Mon. Weather Rev.*, *123*, 2181.

Kalman, R. E. (1960), A new approach to linear filtering and prediction problems, *J. Basic Eng.*, *82*, 35.

Lorenc, A. C. (1997), Development of an operational variational assimilation scheme, *J. Meteorol. Soc. Jpn.*, *75*, 339.

Lyapunov, A. M. (1907), Probléme généerale de la stabilité du mouvement, *Ann. Math. Stud.*, *17*, 1947.

Slinker, S. P., J. A. Fedder, J. M. Ruohoniemi, and J. G. Lyon (2001), Global MHD simulation of the magnetosphere for November 24, 1996, *J. Geophys. Res.*, *106*, 361.

I. Doxas and R. S. Weigel, Department of Physics, University of Colorado, Boulder, CO 80309, USA. (doxas@colorado.edu)

W. Horton, Institute for Fusion Studies, University of Texas at Austin, 1 University Station, Austin, TX 78712, USA.

J. Lyon, Department of Physics and Astronomy, Dartmouth College, Hanover, NH 03755, USA.

M. Wiltberger, High Altitude Observatory, 3450 Mitchell Lane, Boulder, CO 80301, USA.