

# An ensemble-type Kalman filter for neutral thermospheric composition during geomagnetic storms

M. V. Codrescu, T. J. Fuller-Rowell, and C. F. Minter

Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, Colorado, USA

Space Environment Center, National Oceanic and Atmospheric Administration, Boulder, Colorado, USA

Received 5 May 2004; revised 24 July 2004; accepted 10 September 2004; published 11 November 2004.

[1] Global circulation models (GCMs) for the thermosphere ionosphere system have been in use for more than 20 years. In the beginning the GCMs were run on supercomputers, were expensive to run, and were used mainly to provide insight into the physics of the region and to interpret measurements. Advances in computer technology have made it possible to run GCMs on desktops and to compare their results with real-time or near-real-time measurements. Today's models are capable of reproducing generic geomagnetic storm effects, but modeling specific storms is still a challenge because accurate descriptions of the energy input during storms are not easy to obtain. One way to compensate for the uncertainty in model inputs for a given period is to assimilate measurements into the model results. In this way, meteorologists have been improving their ability to model tropospheric weather for the last few decades. Data assimilation algorithms have seen an explosive growth in the last few years, and the time has come to apply such techniques to the thermospheric storm effects problem. We present results from an ensemble Kalman filter scheme that determines the best estimate of the global height-integrated O/N<sub>2</sub> ratio by combining GCM results and uncertainties with measurements and their errors. We describe the differences that result from the application of an ensemble Kalman filter to an externally forced system (neutral chemical composition) versus a system dominated by the initial condition and internal dynamics (tropospheric weather and ocean models). The results demonstrate that an ensemble of 10 members is able to characterize the state covariance matrix with sufficient fidelity to enable the Kalman filter to operate in a stable mode. Some information about the external forcing was extracted from the estimate of the state. The general trend of the forcing was followed by the filter, but departures were present over some periods.

**INDEX TERMS:** 0355 Atmospheric Composition and Structure: Thermosphere---composition and chemistry; 0358 Atmospheric Composition and Structure: Thermosphere---energy deposition; 3210 Mathematical Geophysics: Modeling; 3299 Mathematical Geophysics: General or miscellaneous; **KEYWORDS:** Kalman filter, data assimilation, thermosphere composition

**Citation:** Codrescu, M. V., T. J. Fuller-Rowell, and C. F. Minter (2004), An ensemble-type Kalman filter for neutral thermospheric composition during geomagnetic storms, *Space Weather*, 2, S11002, doi:10.1029/2004SW000088.

## 1. Introduction

[2] The purpose of this paper is to describe the application of a data assimilation scheme that combines measurements and model results on the basis of their respective uncertainties. In the absence of detailed knowledge of the high-latitude input, simulated measurements of the chemical composition in the thermosphere are used to guide the model results toward a solution that is consistent with the input without actually measuring it. In this section we describe the state of modeling and the need for data assimilation for the thermosphere ionosphere system.

[3] Important progress has been made in modeling the effects of geomagnetic storms in the thermosphere and ionosphere using global models with large grids (2°–5° latitude, 5°–18° longitude). Numerical simulations of

generic [e.g., Fuller-Rowell *et al.*, 1994, 1996] and specific [e.g., Codrescu *et al.*, 1997; Fuller-Rowell *et al.*, 2000] storms using global circulation models (GCMs) have provided a better understanding of the dynamics of the upper atmosphere and have permitted the identification of the processes responsible for global ionospheric storm effects at high latitudes and midlatitudes. Neutral composition changes driven by high-latitude energy inputs have been identified as one of the main mechanisms responsible for global storm effects [Rishbeth *et al.*, 1987; Prölss, 1987; Burns *et al.*, 1991]. However, one should keep in mind that the models do not include the physics of processes with horizontal spatial scales smaller than 400 km.

[4] Most GCM simulations use statistical patterns of electric field and particle precipitation to calculate the high-latitude energy input into the upper atmosphere.

Storm effects are very sensitive to the temporal and spatial distribution of the energy input into the high-latitude region. However, statistical patterns do not reproduce even the global-scale characteristics of specific storms. As a result, GCMs can predict typical storm effects (generic storms) rather well but have difficulties modeling specific storm periods because the precise temporal and spatial energy inputs cannot be specified from statistical patterns even for the large grid spacing ( $2^\circ$  latitude and  $18^\circ$  longitude) used in this study.

[5] The high-latitude energy input and its spatial distribution have been recognized as the single largest source of uncertainty in a GCM simulation of specific storm conditions in the upper atmosphere [Codrescu *et al.*, 1997]. The need for better inputs is being addressed by the community, and efforts are being made to produce forcing patterns using the SuperDARN radar network and techniques like the assimilative mapping of ionospheric electrodynamics [Richmond and Kamide, 1988]. However, as the possibility of entirely removing the forcing uncertainty through direct measurements of electric fields and conductivities is impractical at this time, the use of data assimilation techniques and measurements of the variable of interest, like the  $O/N_2$  ratio in this case, is an attractive alternative.

[6] There are many ways to implement “data assimilation” using techniques that range from a simple replacement of a model result by a “raw” measurement at one location (nudging) to sequential statistical methods like the extended Kalman filter (KF). The computational burden generally increases with the sophistication of the data assimilation scheme and quickly becomes prohibitive in the case of global GCMs for the thermosphere-ionosphere system (see Minter [2002] for full details).

[7] The thermosphere-ionosphere system is strongly forced, meaning that the initial condition can become irrelevant in a matter of hours or even tens of minutes if the external forcing changes rapidly. From a modeling point of view, during storms, if we have accurate inputs over tens of minutes, then one can start a simulation from any state (including all zeros) and still get a better answer than if we start from the perfect initial state but have no knowledge of the input. The strong forcing of the system creates new challenges and makes the direct application of data assimilation methods developed for tropospheric weather and ocean circulation difficult.

[8] The full implementation of the extended KF is not practical for large dimensional problems, and the approximate treatment of the state error representation may lead to unbounded error growth [Evensen, 1994]. The difficulties have been resolved in the oceanographic community by the introduction of Monte Carlo–based methods to forecast error statistics using the ensemble Kalman filter (enKF) [Evensen, 1994; Houtekamer and Mitchell, 1998]. More recently, new approaches that combine data assimilation and ensemble predictions for atmospheric and ocean science have proven to perform significantly better in ensemble adjustment Kalman filters (eaKF) [Anderson,

2001, 2003]. However, the enKF and eaKF techniques described in the literature need to be adapted for the case of a strongly forced system before they can be applied to the thermosphere-ionosphere system [Fuller-Rowell *et al.*, 2004; Minter *et al.*, 2004].

[9] In this paper we present preliminary results on the use of an ensemble-type Kalman Filter (entKF) technique on a strongly forced nonlinear system, using software [Minter, 2002] developed as part of the Utah State University Global Assimilation of Ionospheric Measurements effort [Schunk *et al.*, 2004]. The system is modeled by a GCM, the coupled thermosphere ionosphere model (CTIM) described in section 2. The CTIM-propagated neutral composition of the thermosphere is constrained using simulated measurements from a more sophisticated model, the Coupled Thermosphere Ionosphere Plasma-sphere Electrodynamics (CTIPE) model, to which 10% random variability was added. The use of simulated measurements is justified by the need to demonstrate the concept in the absence of true measurements with sufficient spatial and temporal coverage at this time. However, measurements from the Special Sensor Ultraviolet Limb Imager and Special Sensor Ultraviolet Spectrographic Imager on DMSP satellite series will be available in the near future.

[10] The parameter (state) optimized by the filter is the global distribution of the height-integrated  $O/N_2$  ratio. The Kalman state has 1820 elements corresponding to the spatial cells of height-integrated  $O/N_2$  ratio. The height-integrated  $O/N_2$  ratio is calculated as in the study by Strickland *et al.* [1995]. The global three-dimensional distribution of O,  $O_2$ , and  $N_2$  is calculated from height-integrated  $O/N_2$  on the basis of a set of tables derived by Fuller-Rowell using the Mass Spectrometer Incoherent Scatter (MSIS) model and CTIM results. The tables are available by request.

[11] The entKF runs a number of versions of the CTIM model, forced at different levels of geomagnetic activity. Statistics derived from the results are used to specify the uncertainties associated with the model prediction. One version of the model, forced at the most probable level, is used to propagate the Kalman state. The propagated state and “measurements” are then combined, taking into account their associated uncertainties using the Kalman filter equations described in section 3.

## 2. Propagation Model for entKF

[12] The KF technique requires a model to propagate the state in time. When the state is well sampled by measurements, the propagation model is not important, and persistence (the modeled state does not change in time) or a Gauss-Markov process in which the state tends, with a predefined time constant, to a nominal value (climatology) in the absence of measurements, is successfully used. When the state is undersampled, and especially when parts of the state are unobservable, which is likely the case in space sciences at this time,

the propagation model becomes very important. This is why we use CTIM to propagate the state in our data assimilation scheme.

[13] CTIM is a nonlinear, coupled thermosphere-ionosphere physically based numerical code. The model consists of three distinct components which run concurrently and are fully coupled. Included are a global thermosphere, a high-latitude ionosphere, and a midlatitude and low-latitude ionosphere. The thermospheric model was originally developed by Fuller-Rowell [Fuller-Rowell and Rees, 1980; Rees et al., 1980] and is fully described in the Ph.D. thesis of Fuller-Rowell [1981]. The high-latitude ionospheric model was developed by S. Quegan [Quegan, 1982; Quegan et al., 1982]. CTIM has a 2° latitude, 18° longitude, and one scale height vertical resolution, resulting in 27,300 grid points. The time step is 1 min.

[14] CTIM is global and solves numerically the nonlinear primitive equations of momentum, energy, continuity, and major species composition to simulate the time-dependent structure of neutral temperature and density, height of pressure surfaces, and atomic and molecular oxygen and molecular nitrogen concentrations. The ionospheric code includes horizontal transport, vertical diffusion, and ion-ion and ion-neutral chemical processes. The model has been presented in publications by Fuller-Rowell and Rees [1980, 1983] and Fuller-Rowell et al. [1987, 1996] and has been used extensively over the last 15 years to understand the upper atmosphere. More emphasis has been placed lately on operational applications.

[15] The magnetospheric input to the model can be specified in several ways. We normally use time sequences of statistical patterns of auroral precipitation and electric fields described by Fuller-Rowell and Evans [1987] and Foster et al. [1986], respectively, keyed to the hemispheric power index derived from TIROS/NOAA auroral particle measurements [Evans et al., 1988]. Recently, we have updated the magnetospheric input to include the effects of small-scale electric field fluctuations [Codrescu et al., 2000].

[16] Ten identical versions of the model, forced at different levels of geomagnetic activity, are run concurrently as members of the entKF and are used to compute the covariance matrix needed by the filter. One additional run is used to propagate the state in time.

### 3. Ensemble Kalman Filter

[17] The KF is a set of mathematical equations that minimizes the mean of the squared error of the estimate of a process in a computationally efficient way. The technique supports estimations of past, present, and even future states, and it can do so even when the precise nature of the modeled system is unknown. The KF has been used to estimate total electron content values from GPS dual frequency measurements [e.g., Wilson et al., 1995] and to infer the state of the ionosphere [e.g., Howe et al., 1998]. Here we present the basic KF equations and point out the main differences in the case of entKF.

[18] Following loosely Kalman [1960], the classic Kalman filter equations can be written as

$$x^F(t+1) = \Phi(t+1; t)x(t) + u(t) \quad (1)$$

$$P^F(t+1) = \Phi(t+1; t)P(t)\Phi^T(t+1; t) + Q(t) \quad (2)$$

$$y(t+1) = M(t)x^F(t+1) + v(t) \quad (3)$$

$$K(t+1) = P^F(t+1)M^T(t)[M(t)P^F(t+1)M^T(t) + R(t)]^{-1} \quad (4)$$

$$x(t+1) = x^F(t+1) + K(t+1)[y(t+1) - M(t)x^F(t+1)] \quad (5)$$

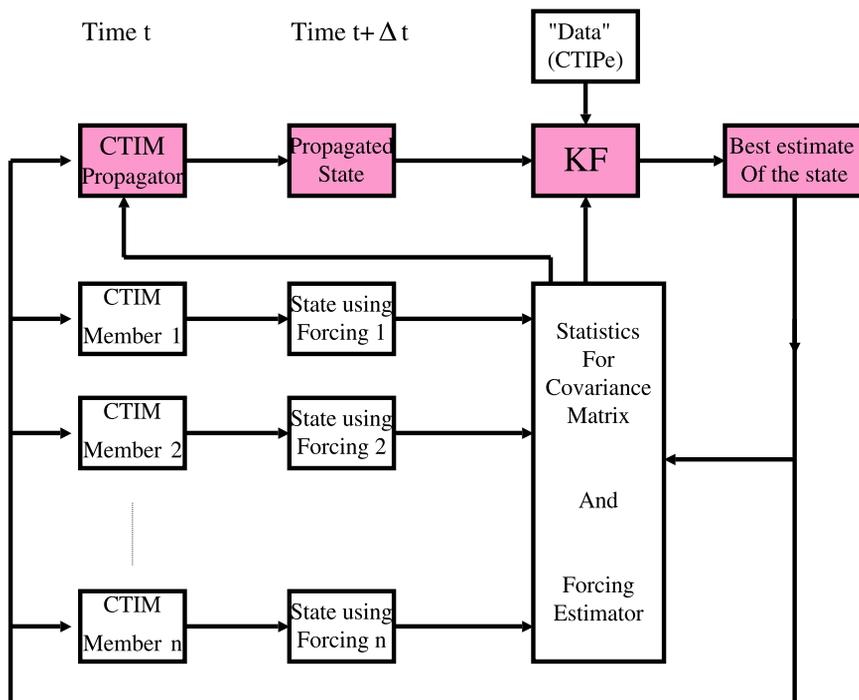
$$P(t+1) = (I - K(t+1)M(t))P^F(t+1), \quad (6)$$

where

- $x(t)$  best estimate of the state vector (of length  $n$ ) at time  $t$ ;
- $x^F(t+1)$  forecast value of the state vector at  $t+1$ , before new data are assimilated;
- $\Phi(t+1; t)$   $n$  by  $n$  transition matrix describing the evolution of the system from time  $t$  to time  $t+1$ ;
- $u(t)$  vector of length  $n$  realization of the process noise, assumed Gaussian random with zero mean;
- $P(t)$   $n$  by  $n$  covariance matrix at time  $t$ ;
- $P^F(t+1)$  forecast value of the covariance matrix at time  $t+1$ , before data is assimilated;
- $Q(t)$  diagonal  $n$  by  $n$  covariance matrix of the process noise, that is,  $E(uu^T)$ ;
- $y(t+1)$  measurement vector (of length  $m$ ) at time  $t+1$ ;
- $M(t)$   $m$  by  $n$  measurement matrix, relating the measured values to the state vector  $x^F(t+1)$ ;
- $v(t)$  vector realization (of length  $m$ ) of the measurement noise, assumed Gaussian random with zero mean;
- $K(t+1)$   $n$  by  $m$  Kalman gain matrix; and
- $R(t)$   $m$  by  $m$  observation error covariance matrix, that is,  $E(vv^T)$ .

[19] In present-day data assimilation language, equations (1) and (2) are the forecast equations for the state and error, respectively, equation (3) is the observation equation, equation (4) is the Kalman gain equation, and equations (5) and (6) are the analysis equations for the state and error, respectively. It is assumed that  $\Phi(t+1; t)$  corresponds to a linear local approximation of the nonlinear dynamical system.

[20] The implementation of the full Kalman filter, although conceptually straightforward, is difficult for real systems with large states. The computational burden quickly becomes impractical, especially for large nonlinear systems. The calculation of the  $\Phi$  matrix for the locally linearized system, the propagation of errors (equation (2)), and the computation of the updated error covariance matrix (equation (6)) are the most expensive computational tasks.



**Figure 1.** Flowchart of our implementation of the entKF. The state is propagated forward by a nonlinear model (CTIM). The error covariance matrix  $P^F(t + 1)$  is evaluated from a number of versions of the CTIM model, called members, that run in parallel and form the ensemble. Simple statistics calculations using the states of the members of the ensemble are used to calculate the covariance matrix. The data used in the assimilation scheme are generated by sampling results from a more sophisticated model (CTIPE) that includes the plasmasphere and electrodynamics that are not modeled in CTIM. Gaussian distributed random variations of 10% are added as measurement noise.

These difficulties have prompted a search for new methods to estimate and propagate the error covariance matrix. Evensen [1994] proposed the use of Monte Carlo methods for error covariance evolution and estimation in an algorithm now called the ensemble Kalman filter (entKF).

[21] Figure 1 illustrates the flowchart of our implementation of the entKF. The state can be propagated forward by a nonlinear model, making the calculation of  $\Phi(t + 1; t)$  unnecessary. The error covariance matrix  $P^F(t + 1)$  is evaluated from a number of versions of the forward model, called members, that run in parallel and form the ensemble. Simple statistics calculations using the states of the members of the ensemble are used to calculate  $P^F(t + 1)$ , as described below. Equations (3), (4), (5), and (6) are implemented as in the classic Kalman filter.

[22] In the case of systems where the initial condition is important (i.e., troposphere and ocean), the members' initial conditions must be distributed according to the uncertainty in the state at the start of every assimilation step. By contrast, in strongly forced systems (i.e., thermospheric neutral composition) the members must be forced at a variety of levels reflecting the probability distribution of the forcing during the next data assimilation step. We start all the ensemble members, at each assimilation step, from the best estimate of the state at that time so that the

only difference between members is the forcing over the current data assimilation time step. The forcing is kept constant for each member over the assimilation time step.

[23] We calculate the forecast covariance matrix using equation (7),

$$P_{ij}^F(t + 1) = P_{ij}(t) + \frac{\sum_{m=1}^M (x_i^m - \bar{x}_i)(x_j^m - \bar{x}_j)}{n - 1}, \quad (7)$$

where  $i$  and  $j$  define the position in the matrix ( $i, j = 1, 1820$  in this case) and  $M$  is the number of ensemble members ( $M = 10$  in this study).  $P_{ij}(t)$  is the covariance matrix at the end of the previous assimilation step. Equation (7) replaces equation (2). Note that  $P_{ij}(t)$  is not properly propagated forward; that is, the existing uncertainties are assumed to apply unchanged after the state is propagated. The use of the unpropagated  $P_{ij}(t)$  is justified by the coarse resolution of the model and the short (12 min) data assimilation step. This is a good approximation, as the uncertainties in the state do not change significantly over 12 min because of system dynamics. Note that in the case of unforced systems, there is no need to add  $P_{ij}(t)$  in equation (7) because the second term in equation (7) already accounts for the uncertainty in the state at the beginning of the assimilation

step due to the random perturbations of the initial condition of the ensemble members.

[24] The number of members  $M$  used in this study was arbitrarily chosen to be equal to the number of activity levels in the high-latitude convection and particle precipitation models used in CTIM. The optimum number of members depends on the system. The addition of members to the ensemble improves the statistical representation of the system, but the improvement becomes smaller as the total number of members increases. The results illustrate, in the case of our system, that a small number of members (10) may be sufficient.

[25] The Kalman state used in this study is the global distribution of the height-integrated  $O/N_2$  ratio, which is tightly related to neutral chemical composition. The  $O/N_2$  ratio is controlled by the integral heating over periods of hours and therefore is sensitive only to large-scale (hundreds of kilometers) effects. The state contains 1820 elements, corresponding to the grid points in CTIM. The grid points are distributed uniformly over the globe at  $18^\circ$  geographic longitude and  $2^\circ$  geographic latitude ( $20 \times 91$ ). The dynamics of the nonlinear system are not included directly in the state but are captured indirectly through the proper forcing of the model used to propagate the state forward in time. The forcing could easily be included in the state, for example, as coefficients of some empirical orthogonal functions or as a perturbation of the statistical pattern at all the high-latitude grid points (about 400 values for each hemisphere). The coefficients or the perturbations would be estimated by the filter at every assimilation time step, but we have not done so in this study. Including the forcing in the state is a natural next step.

[26] The Kalman filter does not use any direct information of the forcing (i.e., hemispheric power; activity level; and  $Kp$ ,  $Ap$ ,  $AE$ , or any other index) in this study. This is deliberately done to test the limits of the assimilation scheme. The propagation of the state for the current assimilation step is performed at a level of forcing that is estimated on the basis of the comparison of the individual members' final state with the best estimate of the state after the assimilation of measurements. At the end of each assimilation step we choose the forcing of the member that has the lowest root-mean-square (RMS) difference from the assimilated state for the next state propagation. The result is that the assimilation scheme is one assimilation step behind in propagation of the state, and the performance of the scheme is limited. The limitation can be considerably reduced by making the assimilation time step small compared with the time constant of the modeled system or by including additional information in the decision process for the level of forcing of the state propagation step, for example, appropriately delayed solar wind data from the Sun-Earth libration point (L1).

[27] The key to the success of the data assimilation scheme is to have a good estimate of the global forcing (the activity level, one number from 1 to 10, in this case) which assures proper global dynamics and to compensate

for its uncertain spatial distribution using data assimilation. Better results are to be expected in the case of an increased number of forcing parameters as long as there are enough measurements to properly constrain them. Smaller temporal and spatial scales can easily be addressed if appropriate measurements become available in the future and the simulations are carried out on finer grids.

#### 4. Results

[28] The truth file was created by running the Coupled Thermosphere Ionosphere Plasmasphere Electrodynamics (CTIPE) model [Millward *et al.*, 1996, 2001] for 17 April 2002. CTIPE includes, in addition to all the modules contained in CTIM, a physical numerical model of the plasmasphere and a self-consistent global electrodynamic calculation for dynamo electric fields. The high-latitude forcing used for CTIPE is illustrated in Figure 2 (thick line). Activity level 5 corresponds to  $Kp = 3-$ , and level 10 corresponds to  $Kp = 6-$  and above.

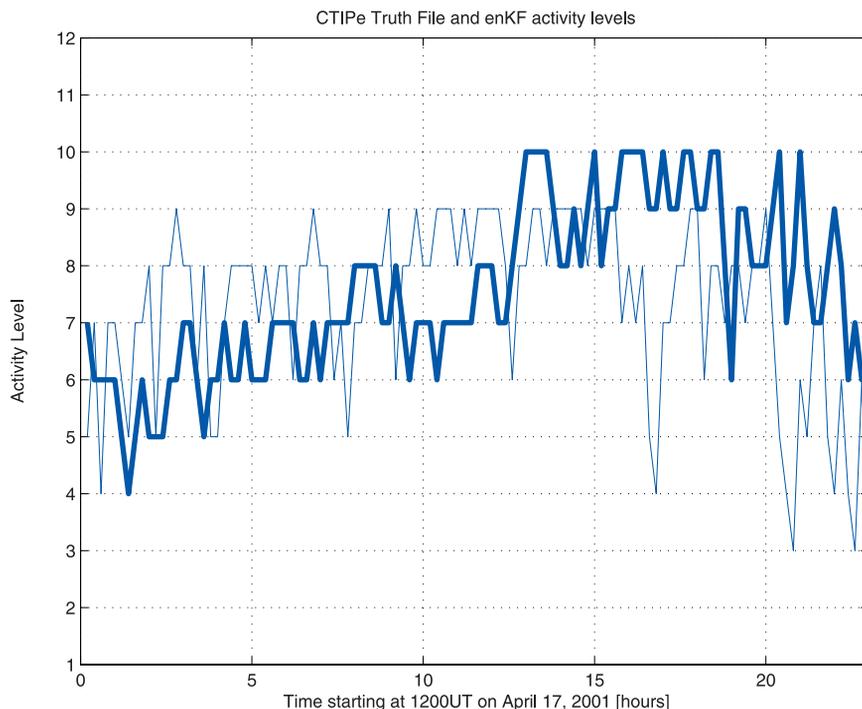
[29] CTIPE forcing for this day is based on 1-min ACE measurements obtained from the Space Environment Center in Boulder, Colorado, averaged over 12 min and used to drive the Weimer electric potential model [Weimer, 1995]. Note that CTIM uses the Foster model [Foster *et al.*, 1986] for high-latitude convection electric fields. The difference between the models and convection patterns used in the generation of the truth file and the members of the ensemble Kalman filter provides for a more realistic test of the assimilation scheme. Using the same model for the truth file and assimilation can reduce the RMS error of the assimilation scheme to 5–6% even when 10% errors are added to simulated measurements.

[30] The truth file is sampled in an ideal way to provide measurements of 20% of the state elements ( $4 \times 91 = 364$  state values) every assimilation step (12 min). The observations are in four constant longitude (local time) sectors and cover all latitudes. An entirely different set of state elements is observed every assimilation step (except for the poles, which get four different measurements every step) until all state elements are observed (60 min), when the pattern repeats. This is not a physical observing system and is used only for convenience.

[31] A white noise component obtained from a random number generator with 0 mean and 0.1 standard deviation (about 10% of the mean value of the state variable) is added to each truth file value before it is used in the data assimilation procedure. One of the results of adding noise to the truth values is that the state of the pole values in the entKF is always the result of assimilating four measurements that can be widely different at times.

[32] The initial state of the filter is generated using a constant value. The large RMS error of over 25% of the initial state is intentional as we were interested in testing the convergence speed of the filter for different values of the filter parameters.

[33] Figure 2 (thin line) shows the forcing used in the propagation of the state by the Kalman filter. The filter



**Figure 2.** Forcing for the CTIPE model used to generate the “truth” file used in the ensemble Kalman filter data assimilation scheme (thick line). The forcing is based on 1-min ACE measurements obtained from the Space Environment Center in Boulder, Colorado, averaged over 12 min and used to drive the Weimer electric potential model [Weimer, 1995]. Activity level 4 corresponds to a cross-polar cap potential (CPCP) of 41 kV, while activity level 10 corresponds to CPCP of over 100 kV [Foster *et al.*, 1986]. Forcing is estimated from the results of the entKF (thin line). The estimating procedure is external to the entKF and has not been tuned for this example.

forcing is not part of the state and was inferred outside of the Kalman filter on the basis of the assimilated state as described in section 3. The limitations of our simple procedure to derive the forcing used in the state propagation are obvious in Figure 2 since the derived activity level is not following the truth activity level very well. We plan to discuss the optimization of the forcing for the state propagation in a future paper.

[34] Figure 3 shows the evolution of the global RMS difference between the assimilated state and the truth file for 17 April 2001. The large drop in RMS in the first few assimilation steps is due to the large error associated with the initial condition and to a larger representation error assigned to the model for the first 10 assimilation steps. The RMS difference increases slightly with the external forcing but stays below 14% for the entire simulation of this geomagnetically disturbed period.

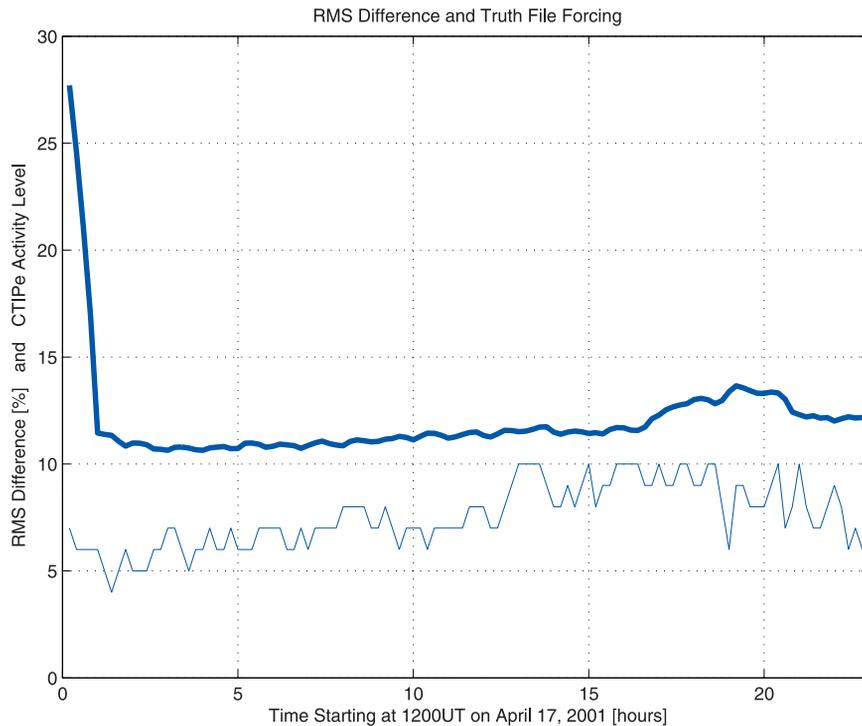
[35] Figure 4 illustrates the truth generated by the CTIPE model run (top), the assimilated state produced by the entKF (middle), and the difference (error) between the two (bottom), immediately following the first data assimilation step. The changes produced by the assimilation procedure are easy to identify. The error of the initial state (around 25%) is reduced to about 10%.

[36] Figure 5 is similar to Figure 4 but for 0200 UT, corresponding to 17 hours of data assimilation. Figure 6

corresponds to 1200 UT, after 24 hours of data assimilation. The presence of a persistent region of large errors in the southern midlatitudes (Figures 5 and 6) is due to the inability of the CTIM model used in the propagation of the state in the entKF to capture some dynamical features produced by the more sophisticated CTIPE model used to generate the truth file. The propagation model (CTIM) does not propagate the corrected state and, in fact, wipes out the result of the assimilation because it cannot reproduce the dynamics of the model that produced the truth file. This is, to some extent, always the case for real systems, as the models are simplifications of reality.

## 5. Conclusions

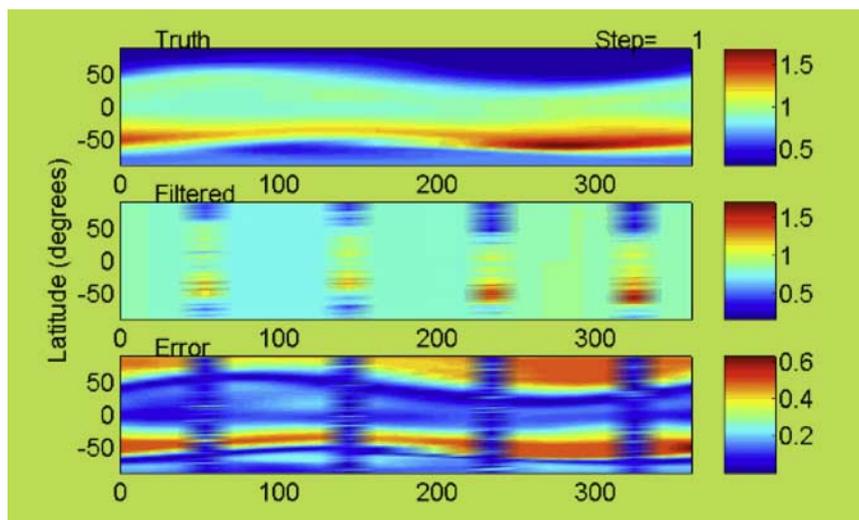
[37] We described results from the application of an ensemble Kalman filter data assimilation technique to a strongly forced system, using a 10-member ensemble. The estimated state is the global  $O$  to  $N_2$  ratio evaluated on a  $2^\circ$  by  $18^\circ$  grid of the externally forced thermosphere ionosphere system. In the case of strong external forcing, the necessity of randomly perturbing the initial condition of the members is replaced by the need to force the members in a way that reflects the possible future evolution of the system.



**Figure 3.** Global RMS difference between the entKF estimated state and the truth file versus time (thick line) and truth file run forcing (thin line).

[38] Our results demonstrate that an ensemble of 10 members is able to characterize the state covariance matrix with sufficient fidelity to enable the Kalman filter to operate in a stable mode. The RMS difference of the estimated  $O$  to  $N_2$  ratio is less than 14% over the simulation period that includes a geomagnetic storm.

[39] Using the results of the entKF, we were able to extract information about the external, large-scale forcing of the thermosphere ionosphere system. Our algorithm to estimate the forcing is external to the entKF, very simple, and far from optimal. The general trend of the forcing was followed by the filter, but departures were also present



**Figure 4.** Global snapshot of height-integrated  $O$  to  $N_2$  ratio calculated as in the study by Strickland *et al.* [1995], immediately following the first data assimilation step (after 12 min of data assimilation): (top) truth, (middle) the assimilated state, and (bottom) the RMS difference between the two.

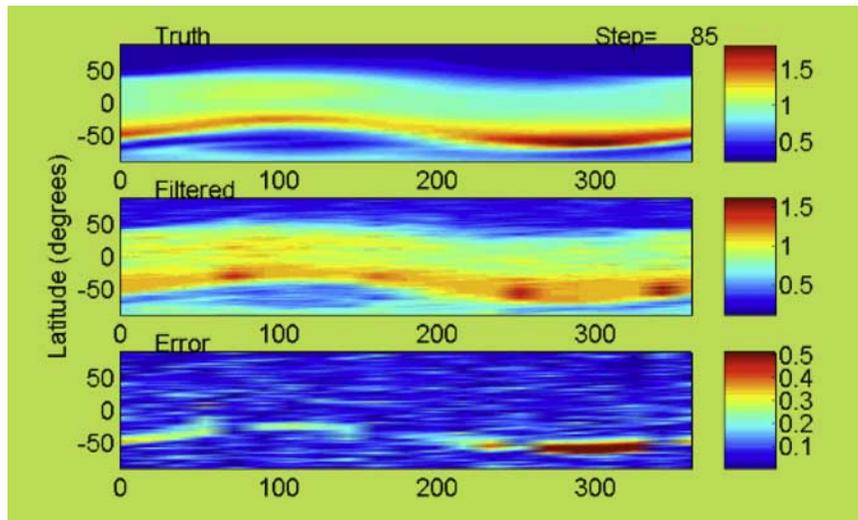


Figure 5. Same as Figure 4 but after 17 hours of data assimilation.

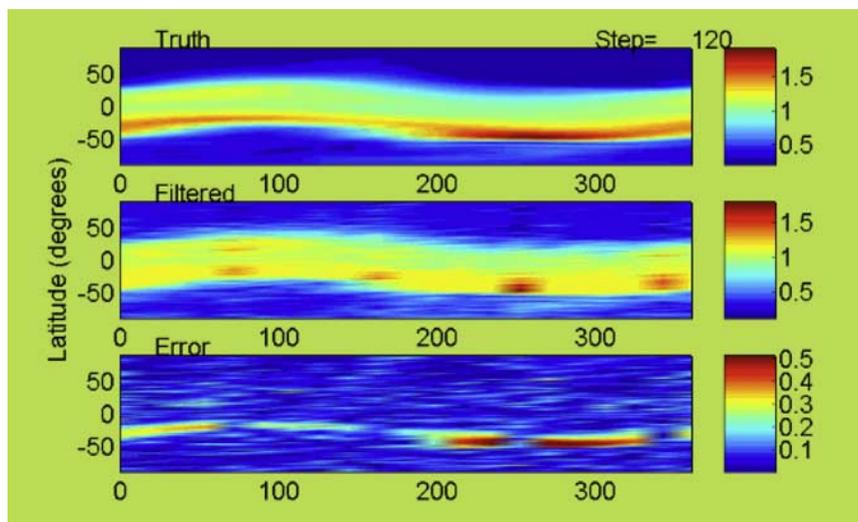


Figure 6. Same as Figure 4 but after 24 hours of data assimilation.

over some periods. There are several ways to dramatically improve the estimate of the external forcing. One such way would be to include properly delayed solar wind data from L1 in the decision making for the forcing of the propagation step.

[40] **Acknowledgments.** The funding for this project was provided by a DOD Multidisciplinary University Research Initiative, contract N00014-99-1-0712. This work was done as part of the USU GAIM program.

## References

- Anderson, J. L. (2001), An ensemble adjustment Kalman filter for data assimilation, *Mon. Weather Rev.*, *129*, 2284–2903.
- Anderson, J. L. (2003), A local least squares framework for ensemble filtering, *Mon. Weather Rev.*, *131*, 634–642.
- Burns, A. G., T. L. Killeen, and R. G. Roble (1991), A theoretical study of thermospheric composition perturbations during an impulsive geomagnetic storm, *J. Geophys. Res.*, *96*, 14,153–14,167.
- Codrescu, M. V., T. J. Fuller-Rowell, and I. S. Kutiev (1997), Modeling the F-layer during specific geomagnetic storms, *J. Geophys. Res.*, *102*, 14,315–14,329.
- Codrescu, M. V., T. J. Fuller-Rowell, J. C. Foster, J. M. Holt, and S. J. Cariglia (2000), Electric field variability associated with the Millstone Hill electric field model, *J. Geophys. Res.*, *105*, 5265–5273.
- Evans, D. S., T. J. Fuller-Rowell, S. Maeda, and J. Foster (1988), Specification of the heat input to the thermosphere from magnetospheric processes using TIROS/NOAA auroral particle observations, *Adv. Astronaut. Sci.*, *65*, 1649–1667.
- Evensen, G. (1994), Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *J. Geophys. Res.*, *99*, 10,143–10,162.
- Foster, J. C., J. M. Holt, R. G. Musgrove, and D. S. Evans (1986), Ionospheric convection associated with discrete levels of particle precipitation, *Geophys. Res. Lett.*, *13*, 656–659.
- Fuller-Rowell, T. J. (1981), A three dimensional time dependent global model of the thermosphere, Ph.D. thesis, Univ. College London, London.
- Fuller-Rowell, T. J., and D. S. Evans (1987), Height integrated Pedersen and Hall conductivity patterns inferred from the TIROS-NOAA satellite data, *J. Geophys. Res.*, *92*, 7606–7618.
- Fuller-Rowell, T. J., and D. Rees (1980), A three-dimensional, time-dependent, global model of the thermosphere, *J. Atmos. Sci.*, *37*, 2545–2567.
- Fuller-Rowell, T. J., and D. Rees (1983), Derivation of a conservative equation for mean molecular weight for a two constituent gas within a three-dimensional, time-dependent model of the thermosphere, *Planet. Space Sci.*, *31*, 1209–1222.
- Fuller-Rowell, T. J., D. Rees, S. Quegan, R. J. Moffett, and G. J. Bailey (1987), Interactions between neutral thermospheric composition and the polar thermosphere using a coupled global model, *J. Geophys. Res.*, *92*, 7744–7748.
- Fuller-Rowell, T. J., M. V. Codrescu, R. J. Moffett, and S. Quegan (1994), Response of the thermosphere and ionosphere to geomagnetic storms, *J. Geophys. Res.*, *99*, 3893–3914.
- Fuller-Rowell, T. J., M. V. Codrescu, H. Rishbeth, R. J. Moffett, and S. Quegan (1996), Response of the thermosphere and ionosphere to geomagnetic storms, *J. Geophys. Res.*, *101*, 2343–2353.
- Fuller-Rowell, T. J., M. V. Codrescu, and P. Wilkinson (2000), Quantitative modeling of the ionospheric response to geomagnetic activity, *Ann. Geophys.*, *18*, 766–781.
- Fuller-Rowell, T. J., C. F. Minter, and M. V. Codrescu (2004), Data assimilation for neutral thermospheric species during geomagnetic storms, *Radio Sci.*, *39*, RS1S03, doi:10.1029/2002RS002835.
- Houtekamer, P. L., and H. L. Mitchell (1998), Data assimilation using an ensemble Kalman filter technique, *Mon. Weather Rev.*, *126*, 796–811.
- Howe, B. M., K. Runciman, and J. A. Secan (1998), Tomography of the ionosphere: Four-dimensional simulations, *Radio Sci.*, *33*, 109–128.
- Kalman, R. E. (1960), A new approach to linear filtering and prediction problems, *Trans. ASME J. Basic Eng.*, *82*, 35–45.
- Millward, G. H., R. J. Moffett, S. Quegan, and T. J. Fuller-Rowell (1996), A coupled thermosphere ionosphere plasmasphere model (CTIP), in *STEP Handbook on Ionospheric Models*, edited by R. W. Schunk, pp. 239–279, Utah State Univ., Logan.
- Millward, G. H., I. C. F. Müller-Wodrag, A. D. Aylward, and T. J. Fuller-Rowell (2001), An investigation into the influence of tidal forcing on F region equatorial vertical ion drift using a global ionosphere-thermosphere model with coupled electrodynamics, *J. Geophys. Res.*, *106*, 24,733–24,744.
- Minter, C. F. (2002), Thermospheric composition forecasting using Kalman filtering techniques, Ph.D. dissertation, Univ. of Colorado, Boulder.
- Minter, C. F., T. J. Fuller-Rowell, and M. V. Codrescu (2004), Estimating the state of the thermospheric composition using Kalman filtering, *Space Weather*, *2*, S04002, doi:10.1029/2003SW000006.
- Pröls, G. W. (1987), Storm-induced changes in the thermospheric composition at middle latitudes, *Planet. Space Sci.*, *35*, 807–811.
- Quegan, S. (1982), Theoretical studies of large-scale features of the Earth's mid-to-high-latitude ionosphere, Ph.D. thesis, Univ. of Sheffield, Sheffield, U. K.
- Quegan, S., G. J. Bailey, R. J. Moffett, R. A. Heelis, T. J. Fuller-Rowell, D. Rees, and R. W. Spiro (1982), A theoretical study of the distribution of ionisation in the high-latitude ionosphere and the plasmasphere: First results on the mid-latitude trough and the light ion trough, *J. Atmos. Terr. Phys.*, *44*, 619–640.
- Rees, D., T. J. Fuller-Rowell, and R. W. Smith (1980), Measurements of high latitude thermospheric winds by rocket and ground-based techniques and their interpretation using a three-dimensional, time-dependent dynamical model, *Planet. Space Sci.*, *28*, 919–932.
- Richmond, A. D., and Y. Kamide (1988), Mapping electrodynamic features of the high-latitude ionosphere from localized observations: Technique, *J. Geophys. Res.*, *93*, 5741–5759.
- Rishbeth, H., T. J. Fuller-Rowell, and A. D. Roger (1987), F-layer storms and thermospheric composition, *Phys. Scr.*, *36*, 327–336.
- Schunk, R. W., et al. (2004), Global Assimilation of Ionospheric Measurements (GAIM), *Radio Sci.*, *39*, RS1S02, doi:10.1029/2002RS002794.
- Strickland, D. J., J. S. Evans, and L. J. Paxton (1995), Satellite remote sensing of the thermospheric O/N<sub>2</sub> and solar EUV: 1. Theory, *J. Geophys. Res.*, *100*, 12,217–12,226.
- Weimer, D. R. (1995), Models of high-latitude electric potentials derived with a least error fit of spherical harmonic coefficients, *J. Geophys. Res.*, *100*, 19,595–19,608.
- Wilson, B. D., A. J. Mannucci, and C. D. Edwards (1995), Subdaily Northern Hemisphere ionospheric maps using an extensive network of GPS receivers, *Radio Sci.*, *30*, 639–648.

M. V. Codrescu, T. J. Fuller-Rowell, and C. F. Minter, Space Environment Center, National Oceanic and Atmospheric Administration, 325 Broadway, Boulder, CO 80303-3328, USA. (mihail.codrescu@noaa.gov)