Final Report for NASA grant NAGS-1093
Project Number: 1534134


This report has two parts. Part I summarizes the University of Colorado component, and Part II summarizes the component from the subcontract to NCAR.

Part I

Abstract

A data assimilation system for specifying the thermospheric density has been developed over the last several years. This system ingests GRACE/CHAMP-type in situ as well as SSULI/SSUSI remote sensing observations while making use of a physical model, the Coupled Thermosphere-Ionosphere Model (CTIM) (Fuller-Rowell et al., 1996). The Kalman filter was implemented as the backbone to the data assimilation system, which provides a statistically ‘best’ estimate as well as an estimate of the error in its state. The system was tested using a simulated thermosphere and observations. CHAMP data were then used to provide the system with a real data source. The results of this study are herein.

INTRODUCTION

The particular advantage of this research is the creation of a data assimilation system that makes use of a physical model, the Coupled Thermosphere-Ionosphere Model (CTIM) (Fuller-Rowell et al., 1996). To date, no system makes use of a physical model for state propagation. The advantage of the physical model, CTIM, over empirical models comes from its ability to represent unclimatological features that are often present during geomagnetic storm conditions. A physical model, like CTIM, in a predictor/corrector-type data assimilation technique like the Kalman filter, is ideal since the physical model has the ability to provide the ‘best’ state of the thermosphere based on conditions at a previous time. Correcting the physical model with observations, in a statistically rigorous manner, provides an optimal method for minimizing the errors while representing the time-dependent conditions of the thermospheric density. Because the thermosphere is strongly driven by external processes, the thrust of the research in the most recent years has focused on improving the specification of the drivers as well as the main parameters for density. It has been shown that improved driver specification can greatly improve accuracy during geomagnetic storms. Improved driver specification has proven itself to be essential since satellite coverage is not globally available at a single epoch and since the upper atmosphere can change much more rapidly in comparison to the satellite revisit rate during geomagnetic storms.

The research examines two main type of satellite platforms: remote sensing and in situ. The remote sensing instruments examined here are the Special Sensor Ultraviolet Imager (SSUSI) developed by the Applied Physics Laboratory (Paxton et al., 1992) and the Special Sensor Ultraviolet Limb Imager, SSULI (McCoy and Thonnard, 1997). The
in situ satellites under examination are the Challenging Minisatellite Payload (CHAMP) from GFZ Potsdam and NASA’s Gravity Recovery and Climate Experiment (GRACE).

**Kalman Filter Overview**

Because of its specific design to handle nonlinear systems, an extended form (Costa and Moore, 1991) of the Kalman filter is used. The Kalman filter (Kalman, 1960; Kalman and Bucy, 1961) is an alternative way of implementing the least squares method by sequentially solving the problem with each new observation, where

\[ y = Hx + \varepsilon \]

\( y \) is the observation, \( x \) is the state, and \( H \) is the linear relationship between the state and observation. \( \varepsilon \) represents the error in the observation that is to be eliminated. The primary data assimilation system, based on the extended Kalman filter, can also be written in ensemble form, as will be shown in a moment. The elements of the state vector represent the driver (ap), the density/temperature at each grid point over the globe.

To propagate the estimated state and its error variance-covariance matrix forward in time, CTIM is used.

\[ \bar{x} = \Phi_{CTIM} \hat{x} \]

and

\[ \bar{P} = \Phi_{CTIM} P \Phi_{CTIM}^T \]

where \( \hat{x} \) is the previous best estimate, \( \bar{x} \) is the model propagated state, \( P \) is the error covariance matrix of the previous best estimate of the state, and \( \bar{P} \) is the model propagated error covariance matrix. The physical model, CTIM, is used to calculate the state transition matrix, \( \Phi_{CTIM} \), where the superscript, \( T \), is the transpose.

The weighting, or Kalman gain, \( K \), is determined from the propagated state error covariance matrix, \( \bar{P} \), the observation error variance-covariance matrix, \( R \), and the mapping matrix, \( H \), as

\[ K = \bar{P} H^T \left( H \bar{P} H^T + R \right)^{-1} \]

The calculated Kalman gain then is used to map the observation vector correction to the model propagated state as

\[ \hat{x} = \bar{x} + Ky \]

where the new associated error covariance matrix is calculated as

\[ P = (I - KH) \bar{P} (I - KH)^\top + KRK^\top. \]
The extended Kalman filter equations are then repeated, using the corrected state estimate and its corrected error variance-covariance matrix, to obtain a state and state error variance-covariance matrix for the next observation time.

The advantage of writing the equations of the Kalman filter in extended form comes from the continual correction to the nominal state. Typically, estimation methods specify the deviation from a nominal state. This deviation is assumed linear, which is an adequate assumption if the deviation is sufficiently small enough to lie within the linear regime. If, however, the state is not well known, as can happen during storm conditions, the current state estimate may not lie within this linear regime of the nominal state making the linear assumption invalid. The extended version of the Kalman filter allows for the continual correction to the nominal state, which in turn helps minimize the effects of nonlinearities by continually decreasing the difference between the current and nominal states.

Additional improvements include the ensemble version of the extended Kalman filter (Evensen et al., 2000; Keppenne, 2000; Houtekamer and Mitchell, 1998) which avoids the numerical difficulties of calculating the transition and covariance matrices, allowing one to correlate the drivers with the thermospheric response. The correlation length between adjacent points within the thermosphere is very short, and therefore, adjacent points are poorly correlated in the transition and covariance matrices of the Kalman filter. However, the relationships between driver, density/temperature, and composition change are strongly correlated at a given point and must be properly represented in the transition and covariance matrices. Including these correlated points in the transition matrix makes traditional calculations of the transition and covariance matrices too tedious, even for a computer, and therefore impractical, and an ensemble version of the extended Kalman filter must be applied.

Although 20 members were used in the ensemble Kalman filter in this research, a simplified schematic, using 3 members, is shown in figure 1 to illustrate the mechanics of the ensemble Kalman filter.

![Diagram of an ensemble Kalman filter](image)
The ensemble Kalman filter uses direct calculation of the covariances based on the Monte Carlo statistics of an ensemble of separate filters. In the ensemble Kalman filter, each member of the ensemble is a separate filter estimating a separate state. The ensemble Kalman filter uses the variance in these separate states to estimate the covariance of the entire ensemble.

A set of 20 filter processes are run in parallel where the initial states are randomized with a Gaussian number generator. The standard deviation of the random number generator is the same as that for the desired a priori state error covariance matrix. As each of these processes of the ensemble is propagated independently forward in time, the resultant distribution of the propagated states of the ensemble will have a new associated error variance-covariance matrix. Thus, the propagated state error covariance matrix, $\bar{P}$, may be calculated directly from the ensemble of propagated states using the standard statistical equations below:

\[
\sigma_n^2 = \frac{\sum_{i=1}^{m} (\bar{x}_{i,k} - \mu_1)^2}{m - 1} \tag{1}
\]

which provides the diagonal, or variance, elements of $\bar{P}$. The ensemble is made of $m$ members. $\sigma_n^2$ is the variance of all of the member states in the ensemble of the $i$th diagonal element, and $\mu_1$ is the mean of the ensemble for the $i$th state element. $\bar{x}_{i,k}$ is the $i$th element of the $k$th propagated member state element. The off-diagonal, or covariance, terms are found by the similar equation

\[
\sigma_y^2 = \frac{\sum_{i=1}^{m} (\bar{x}_{i,k} - \mu_1)(\bar{x}_{j,k} - \mu_2)}{m - 1} \tag{2}
\]

where $\sigma_y^2$ is the $i,j$th element of $\bar{P}$, $\mu_1$ is the mean of the ensemble for the $i$th state element, and $\mu_2$ is the mean of the $j$th state element. $\bar{x}_{i,k}$ is the $i$th propagated state element of the $k$th member, and $\bar{x}_{j,k}$ is the $j$th propagated state element of the $k$th member.

Each member of the initial estimated state, $\bar{x}_i$, is randomized initially and is propagated using the physical model. Once the state is propagated forward to the next time step, the state error variance-covariance matrix is calculated statistically, as described above in equations (1) and (2). As a result, the transition matrix need not be calculated. Since the ensemble carries the state error information, the current error variance-covariance matrix, $P$, need not be stored in memory. Since the current error covariance matrix $P$ is stored in the state ensemble.

The ensemble Kalman filter has an advantage over the traditional Kalman filters in that the state error covariance matrix is calculated directly using equations (1) and (2).
Also, by calculating the state error covariance matrix directly from the ensemble distribution, the state error covariance matrix avoids the problem of asymmetry.

**PHYSICAL MODEL OVERVIEW**

The Coupled Thermospheric-Ionospheric Model, CTIM (Fuller-Rowell et al., 1996 and 2000), provides a complex and well-tested physical model for the propagation of the state. CTIM is a combination of two independently developed physical models. The first part of CTIM contains a global, non-linear, time-dependent neutral atmospheric model developed at the University College London (Fuller-Rowell and Rees, 1980 and 1983). The second part contains a mid- and high-latitude ionospheric convection model that originated at Sheffield University (Quegan et al., 1982). The amount of Joule heating, which depends on the electric field strength, is one of the contributors to the ionospheric-thermospheric coupling. The neutral atmospheric portion of CTIM, however, is the primary part of the algorithm used for the state propagation when assimilating the neutral species.

The thermospheric portion of the code numerically solves the non-linear equations of momentum, energy, and continuity to provide a time-dependent structure of the wind vector, temperature, and density in the neutral atmosphere. The altitude scale of the thermospheric model is pressure dependent. The range begins at 1 Pascal, 80 km altitude, and extends from 300 km to 700 km altitude depending on the amount of expansion from heating.

The non-linear equations of the thermospheric portion are solved self-consistently with a high- and mid-latitude ionospheric convection model poleward of 23 degrees latitude. The ionospheric portion numerically solves the non-linear equations for ion continuity, diffusion, and temperature. Additionally, odd nitrogen species are taken into account in calculating the molecular ion concentrations as well as other diffusion parameters.

The advantage of CTIM becomes more obvious during geomagnetic storm times when localized Joule heating creates regions of sudden density change in the thermosphere. The dynamical equations in CTIM reproduce these sudden local changes whereas models based on statistical climatology cannot. Additionally, these changes are generally too sudden for persistence to accurately represent, and thus, a physical model may provide the only means for accurately forecasting the neutral composition during geomagnetic storm times.

One may also be faced with loss of data for extended periods of time. Delays in downloading the data from the satellite to the user will inevitably occur. Also, large areas may not be covered due to satellite and instrument malfunctions or daily and seasonal effects. Therefore, a physical model is necessary to provide a forecast to cover any limitations in the availability of data.

Lastly, it may be necessary to provide a forecast of the state in the future. In this case, a physical model, like CTIM, provides the most accurate means for propagating the state. However, the accuracy of this forecast will degrade with time, and relaxation of the state to climatology over a specific time period will be necessary.
DRIVER ESTIMATION

In the most recent years of this project, the main thrust has primarily focused on specifying the thermospheric drivers. In meteorological and ocean data assimilation, the fluid atmosphere or ocean is also described by a constantly changing state. In comparison to the troposphere, however, estimating the upper atmosphere brings about even greater challenges as the winds typically have higher speeds in the range of hundreds of meters per second. Additionally, the coverage is poorer for the upper atmosphere in comparison to the more numerous observation sources available to the meteorological and oceanographic communities. As a result, the state describing the neutral atmospheric density can change significantly between observations – particularly during geomagnetic storms.

Because the upper atmosphere is strongly forced, driver specification can also be used to one’s advantage. Unlike the troposphere, external processes drive most of the variability in the upper atmospheric dynamics as described in figure 2. During quiet times, the Sun directly heats the upper atmosphere by solar radiation in the extreme ultraviolet (EUV) frequencies. EUV heating occurs on the sunlit side of Earth with the maximum heating occurring at the region nearest to the sub-solar point. In the troposphere, for example, much of the variability arises from internal stochastic processes and is not strongly localized. Unlike the troposphere, most of the minute-to-minute variability in the thermosphere arises from the magnetospheric source imposed at high latitudes. During geomagnetic storm times, the heating process becomes more complicated and even more spatially and temporally variable. Geomagnetic storms occur when material, ejected from the sun by a coronal mass ejection, hits the Earth’s magnetosphere. If the solar wind plasma has a southward magnetic field, it creates a coupling with the magnetosphere. Initially, plasma convection increases and auroral particle precipitation expands to lower latitudes. Besides the increased heating rate from particle precipitation and from Joule dissipation, the expanded convective electric field also redistributes the plasma.

Figure 2: The Sun to Neutral Composition Connection.
As often occurs during geomagnetic storms, localized heating can result from the particle precipitation and Joule heating. Localized heating creates changes in very specific regions causing greater variations in the neutral density structure. Empirical models have greater difficulty representing this more varied structure due to their statistical representation of the neutral atmosphere structure, and errors typically increase to up to 50% (Bowman and Storz, 2003). During these times, the data assimilation system must rely more on improved driver specification to correctly ‘drive’ the physical model representation.

The most recent work has concentrated on incorporating the high latitude forcing into the data assimilation system. This forcing includes knowledge of the spatial and temporal variations of the convection electric field and the auroral precipitation. Inaccurate knowledge of these drivers in data assimilation systems for space weather applications will lead to limitations in improving the specification further. Compared with the solar wind parameters that force the magnetosphere, the drivers of the upper atmosphere, although well understood, have been to date poorly quantified. This research has sought to better quantify the balance between solar and magnetospheric forcing over the range of solar and geomagnetic activity through the implementation of an ensemble Kalman filter to specify the drivers for the troposphere. Since the solar heating at low latitudes and the magnetospheric sources at high latitudes control the magnitude and spatial distribution of the global circulation, these drivers strongly affect the neutral composition and density structure, and as a result, the ensemble Kalman filter approach for specifying the drivers is suited for the thermosphere. Further, the improved driver specification, in turn, improves the density specification in the data assimilation system.

DRIVER ESTIMATION THROUGH THE PHYSICAL MODEL

Compared with the solar wind parameters that force the magnetosphere, the magnitude and spatial distribution of the upper atmospheric drivers, although well understood, are difficult to quantify. To truly have any effect on reducing the errors during storm conditions, specification of the spatial and temporal variations of the convection electric field and the auroral precipitation is required. Without data assimilation, the globally averaged Joule heating rate at a given time is probably only known within a factor of two. At a given location, this uncertainty can rise to a factor of ten. Therefore, research has concentrated primarily on quantifying the high altitude forcing during geomagnetic storm by using the approach of also observing the response, rather than just the conventional input parameters alone. Since a given response defines the magnitude and spatial distribution of the source, vastly more information is available for specifying the drivers as well as the current upper atmospheric conditions. From observing the thermospheric response, research has focused on determining if it is possible to specify the real-time driver and heating distribution from observations of the upper atmospheric conditions alone. Since the dynamics of the composition structure is strongly directed by the distribution of the Joule heating, observing the change in composition should act as an additional source for the upper atmospheric drivers and subsequently should provide a better specification for the change in density through the physical model.
The energy input, primarily through the Joule heating, is typically the most significant driver of the changes in neutral composition during geomagnetic storms. As energy is added to the thermosphere, the temperature increases, and the pressure at specific altitudes changes hydrostatically. In a pressure coordinate system, the mechanism can be illustrated as in Figure 3.

Figure 3. The creation of horizontal winds by increasing the pressure level height.

Localized Joule heating raises the temperature within a given area, and the heights of the pressure levels, say $P_1$ and $P_2$, increase at the location where the heating is occurring. In Figure 2, a constant reference altitude, $h^*$, is chosen so that it passes through the two pressure levels. Because of the localized increase in height of the pressure levels, the pressure is no longer uniform at all locations along $h^*$, and a pressure gradient results. Since $P_1$ is at a lower altitude than $P_2$, by the hydrostatic equation, the pressure at $P_1$ will be higher, or in other words, the pressure at point $B$ is higher than at point $A$. As a result of this gradient, a flow occurs from point $B$ toward point $A$, and thus, a diverging horizontal wind at $h^*$ moves away in the radial direction from the region of Joule heating.

Since, in pressure coordinates, the atmosphere can be treated as incompressible, the divergence of the flow on the global scale is conserved, a larger circulation pattern is formed and a convergence of the horizontal winds occur at lower altitudes as shown in Figure 4 below.

Figure 4: Circulation pattern resulting from the divergence of the horizontal winds at higher altitudes and the balancing convergence at lower altitudes.

The size of the circulation pattern depends largely on the temperature gradient, i.e. the amount of Joule heating and the previous state of the thermosphere. The divergence of the horizontal winds at the higher altitudes is the source of the vertical wind that is responsible for neutral composition changes. An example of the resultant horizontal
winds due to heating is shown in figure 5. The true vertical velocity is a combination of
the change in height due to temperature change from the Joule heating and the upward
flow resulting from the divergence pattern.

Figure 5: The Resultant Horizontal Winds Due to Joule Heating.

These horizontal and vertical neutral winds are responsible for the transport of the
neutral species. The horizontal winds, resulting from the localized Joule heating,
transport changes in composition. By observing the neutral composition and density
response, the magnitude of the heat sources can be inferred. An example of how this is
possible is shown in figure 6. The top panel shows simulated thermospheric density
conditions under storm conditions. The middle panel shows typical quiet conditions.
Quiet conditions can be estimated with a high degree of accuracy, and any deviation from
these conditions are very noticeable and indicate the storm effects. The bottom panel
shows the difference between the storm and quiet conditions where the red regions show
where the thermosphere is being heated and the density in these regions at a given
altitude (450 km) is increasing. Also, in the bottom panel, the electric field patterns are
illustrated by the blue outlines. Because the heating (red regions) overlap with the
electric field location (blue outlines), a strong correlation between electric field strength
and distribution and heating magnitude and distribution can be statistically drawn.
Figure 6. How deviations from the quiet state can indicate the regions of heating due to storm conditions.

Estimating the magnitude and distribution of the drivers enables one to ‘drive’ the physical model toward the observations. Two examples of controlling the physical model are shown in figures 7 and 8. Figure 7 shows the measurements (indicated with the black dots) taken by Challenging Minisatellite Payload (CHAMP). The model output is shown by the blue circles in each panel, when no driver is included in the state estimate. Since the model has no driver correction, the data assimilation system cannot properly represent the thermospheric density and an offset can be seen.

Figure 7, on the other hand, shows a much different story. Here, the ‘innovation’ is shown, which is the difference between the actual observation and the physical model prediction. In this example, an attempt is made to estimate the amount of heating in the upper atmosphere. The driver is included in the Kalman state, and the physical model is ‘driven’ based on this estimate. Ideally, one wishes to obtain an innovation that is as near zero as possible, i.e. the observation matches the model prediction. The innovation is often a good indicator of the accuracy in the Kalman filter if no ‘truth’ thermosphere or other data are available for comparison. A non-zero innovation indicates that the physical model is being pushed away from the ‘truth’ due to inaccurate or non-existant driver estimation.
Figure 6: A comparison of CHAMP data and CTIM (without driver specification).

Figure 7: Obtaining a Near-Zero Innovation through Driver Estimation.

Specifying the driver greatly improves the physical model prediction, as also indicated by the blue circles in figure 8. Here, a closer match to the CHAMP
measurements (black dots) can be seen as compared to the previous figure 6. Again, the estimates of the density are obtained from including the drivers in the state.

![Graph](image)

**Figure 8:** A comparison of CHAMP data and CTIM (with driver specification).

As can be seen in figures 7 and 8, some inaccuracies still exist in the Kalman filter estimate as indicated by the non-zero oscillations around the near-zero mean in figure 7 and various miss-matches seen in figure 8. The errors are an effect of the inability to globally observe all features at any given time using a single satellite. The use of a single instrument may miss storm-induced features in the density structure if the feature occurs in a region not covered by the satellite orbit.

An analysis is performed to see what effect satellite coverage has on estimate accuracy. Here a simulated thermosphere is created so that root mean squared (RMS) errors can be calculated for a given satellite arrangement.

Various combinations of SSULI/SSUSI-type (remote sensing) and GRACE/CHAMP (in situ) are examined, and the RMS errors are shown in figure 9. Errors that exist in the density specification can be attributed to errors in the heating distribution knowledge in poorly observed regions. Figure 9 indicates that estimating the magnitude and the distribution of the heating by increasing the satellite coverage can reduce errors, particularly during geomagnetic storm conditions. The errors that still exist come from the nonlinearities between the drivers and the thermospheric response.
A data assimilation system for specifying the thermospheric density has been built and tested using thermospheric simulation and also data from CHAMP. Results show that quiet geomagnetic conditions can be specified to a high degree of accuracy, but storm conditions pose special problems that have required a closer examination of the thermospheric physics as well as the operation of the filtering techniques. The focus over the most recent years of this project has moved toward specifying the thermospheric forcing since this system is so strongly externally driven. Results have shown that errors during geomagnetic storms can be reduced if the drivers can also be specified. Due to the nonlinear relationship between the drivers and the thermospheric response, special attention has been devoted to filtering techniques that are specifically designed for nonlinear systems, which have greatly helped in reducing numerical errors associated with these nonlinearities.
The following publications and conference papers have resulted from this project:


References


Part II
This project resulted in three publications, summarized below.

Matsuo et al. (2002) showed that Empirical Orthogonal Functions (EOFs) can be obtained from innovative analysis of satellite data, which capture a sizable fraction of the total variability of the high-latitude electric fields, and which offer the potential of simplifying the AMIE basis functions and covariance matrices. Matsuo et al. (2003) quantified the seasonal and interplanetary-magnetic-field (IMF) dependence of electric-field variability over the polar regions, based on statistical analysis of satellite data. This work has led to an improved ability to quantify Joule heating in the high-latitude thermosphere.

Matsuo et al. (2004) developed a version of the AMIE procedure that dynamically estimates the AMIE background covariance matrix using the data at hand, using EOFs as basis functions in order to make the matrix diagonal. Applying this to observations from the 1997 January 10 magnetic storm, we found that the dominant mode of electric-field variability changed with time during the storm, as the direction of the IMF rotated due to its flux-rope character. Autocorrelation functions calculated from the time series of fitted basis-function coefficients showed correlation times considerably shorter than the correlation times of variations in the IMF or solar-wind parameters. This suggests a substantial amount of variability associated with internal magnetospheric processes. The AMIE electric fields were used as input to a thermosphere/ionosphere general circulation model and compared with a companion simulation that used climatological electric fields derived from an empirical model, using observed IMF inputs. During certain storm phases these two simulations showed significantly different Joule heating, related partly to different wind systems that are produced.

PUBLICATIONS:

