Local Ensemble Transform Kalman Filter for Earth-System Models: An application to Extreme Events

by

Juan Durazo

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved June 2018 by the
Graduate Supervisory Committee:

Eric Kostelich, Co-Chair
Alex Mahalov, Co-Chair
Wenbo Tang
Mohamed Moustaoui
Rodrigo Platte

ARIZONA STATE UNIVERSITY

August 2018

ABSTRACT

Earth-system models describe the interacting components of the climate system and technological systems that affect society, such as communication infrastructures. Data assimilation addresses the challenge of state specification by incorporating system observations into the model estimates. In this research, a particular data assimilation technique called the Local Ensemble Transform Kalman Filter (LETKF) is applied to the ionosphere, which is a domain of practical interest due to its effects on infrastructures that depend on satellite communication and remote sensing. This dissertation consists of three main studies that propose strategies to improve space-weather specification during ionospheric extreme events, but are generally applicable to Earth-system models:

Topic I applies the LETKF to estimate ion density with an idealized model of the ionosphere, given noisy synthetic observations of varying sparsity. Results show that the LETKF yields accurate estimates of the ion density field and unobserved components of neutral winds even when the observation density is spatially sparse (2% of grid points) and there is large levels (40%) of Gaussian observation noise.

Topic II proposes a targeted observing strategy for data assimilation, which uses the influence matrix diagnostic to target errors in chosen state variables. This strategy is applied in observing system experiments, in which synthetic electron density observations are assimilated with the LETKF into the Thermosphere-Ionosphere-Electrodynamics Global Circulation Model (TIEGCM) during a geomagnetic storm. Results show that assimilating targeted electron density observations yields on average about 60%–80% reduction in electron density error within a 600 km radius of the observed location, compared to 15% reduction obtained with randomly placed vertical profiles.

Topic III proposes a methodology to account for systematic model bias arising

from errors in parametrized solar and magnetospheric inputs. This strategy is applied with the TIEGCM during a geomagnetic storm, and is used to estimate the spatiotemporal variations of bias in electron density predictions during the transitionary phases of the geomagnetic storm. Results show that this strategy reduces error in 1-hour predictions of electron density by about 35% and 30% in polar regions during the main and relaxation phases of the geomagnetic storm, respectively.

ACKNOWLEDGMENTS

The material in this dissertation is based upon work supported by the Air Force Office of Scientific Research under the award number FA9550-15-1-0096. The author was also supported by the National Science Foundation grant DMS-1419593.

TABLE OF CONTENTS

Page

CHAPTER Page

CHAPTER Page

CHAPTER Page

LIST OF FIGURES

Figure Page

Figure Page

Chapter 1

INTRODUCTION

Earth system models refer to the various interacting components that comprise the physical climate system, and include the atmosphere, oceans, land surface, sea ice and their related biogeochemical processes. The definition of Earth-system models may be expanded to include other processes that are important for society's well being, such as water, food and energy supply, as well as communication infrastructures. By its very nature, the development of Earth-system models requires highly interdisciplinary modeling efforts in vastly different domains. A key challenge in modeling of Earth systems includes not only the representation of dynamics in the individual components but also an accounting of the dynamics governing the interaction among them.

Modeling of the Earth system and its individual components can generally be posed as a three-dimensional initial/boundary value problem whose solution estimates the spatiotemporal evolution of the various physical quantities that make up the system of interest. The dynamical system describing this evolution may be used to predict the system's future state given an initial condition. Although the dynamics governing Earth-system components and their interaction with each other ideally may be represented through coupled models based on first principles, it is often necessary to represent some processes as boundary conditions or parameterizations, either because these processes are not understood well enough or because it is impractical to represent them explicitly. For example, sub-grid processes are often accounted for through parameterizations, or the influence of external conditions may be represented through changing boundary conditions.

Typical difficulties in spatiotemporal forecasting of Earth-system models include an incomplete understanding of the dynamics and physics, as well as uncertainties in the specification of initial conditions. Data assimilation addresses the latter problem by incorporating system measurements and observations to help constrain estimates and predictions made with a given dynamical system. The focus of this dissertation is on ensemble-based approaches to data assimilation, which offer a computationally feasible methodology for the estimation of state variables and model parameterizations, and uncertainty quantification of high-order non-linear dynamical models. Additionally, ensemble-based data assimilation techniques are highly portable, since the data assimilation procedure is essentially decoupled from the modeling component.

As observations of the various components of the Earth-system are becoming more abundant, efficient and innovative methodologies for the assimilation of these data are in high demand. This dissertation focuses on the application of techniques developed within the framework of a specific ensemble-based data assimilation scheme, called the local ensemble transform Kalman filter (LETKF). These techniques are applied and presented for the ionosphere, but it is emphasized that they can be generalized to other ensemble-based approaches and also be applied various components of Earth-system models.

1.1 Brief History of Data Assimilation

Much of the recent progress in numerical weather prediction can be attributed to advancements made in data assimilation. Early data assimilation approaches involved directly fitting observations to a grid on which the model equations were propagated. Some examples include empirical schemes such as successive correction methods (SCM) and Newtonian relaxation (or nudging), and statistical techniques such as optimal interpolation. Due to the relative scarcity and irregular distribu-

tion of the observations, these approaches were limited in providing reliable initial conditions needed to integrate atmospheric models of full complexity.

A more effective approach is to use a prior estimate of the atmospheric state, usually referred to as the background estimate, which contains state estimates at each of the model grid points, and computes a state update based on the observations. Initially, climatology was used as the background, but as numerical weather prediction capabilities improved, short-range forecasts were used as initial conditions. The resulting model state after assimilating observations is typically referred to as the analysis and it represents the best system state estimate with the given information in the observations and background estimate.

The basis of many data assimilation systems, which continue to be used even in modern operational systems today, is known as the analysis cycle. This cycle is essentially one long model integration, in which the model predictions are "corrected" at regular time intervals by assimilating observations, so that system predictions remain close to the state of the real atmosphere. The analysis cycle consists of an analysis computation step, where observations are assimilated to obtain an updated state estimate, and a forecast step where the analysis is used as the initial condition for a short-term forecast that produces the background estimate for the next analysis computation step.

A major difficulty in the assimilation of observations is a reliable representation of the temporal variations in the forecast error statistics. After Edward Lorenz made his fundamental discovery about the chaotic nature of atmospheric flows, which implied that any single deterministic forecast would eventually be rendered useless, (Epstein, 1969) and (Leith, 1974) proposed that a stochastic approach should be taken, where the atmosphere's probability distribution function (PDF) is estimated from a given initial PDF estimate. The Fokker-Planck equation provides a theoretical approach

to achieve this, but the explicit propagation of an operational atmospheric model's PDF is computationally unfeasible.

A more computationally feasible approach is ensemble foresting, which uses a set of individual forecasts generated from statistically equivalent sets of initial conditions and/or model configurations to produce a set of predictions that are a representative random sample of the atmospheric model's PDF. This approach attempts to account for the chaotic evolution of the atmosphere and the imperfections in the atmospheric initial conditions. Ensemble forecasting can typically provide a set of forecasts whose ensemble mean is more accurate within the first few days than the individual forecasts, and it also provides an estimate of the forecast uncertainty (Kalnay, 2003). Ensemble forecasting has been used operationally in the US and Europe since 1992 (Kalnay, 2003, Chapter 1).

The ensemble forecasting approach also helped improve forecast error statistics of the background estimate during the assimilation of observations (Hamill and Snyder, 2000). For example, some data assimilation systems may use a long time series of previous forecasts to specify spatially homogeneous and temporally invariant approximations to background error statistics. These simplified approaches are often employed due to the computational difficulties in associated with estimating flow-dependent error statistics.

A relatively recent and promising approach to data assimilation has been the Ensemble Kalman Filter (Evensen, 1994), which naturally incorporates the benefits of ensemble forecasting into the Kalman filter. This approach implements the analysis cycle for a set (or ensemble) of parallel short-term forecasts and a data assimilation correction updates the ensemble mean and forecast covariance matrix using the classical Kalman filter. This correction then provides an ensemble of initial conditions that are individually integrated with the forecast model. While the forecast

covariance matrix is propagated with linear dynamics in the Kalman filter, the ensemble Kalman filter estimates the forecast error covariance matrix from the sample covariance of the ensemble of model forecasts, which is evolved from the fully nonlinear model dynamics. Under assumptions of linear dynamics and Gaussian errors in the observations and forecast, it can be shown that the EnKF produces the correct background error covariance matrix as the ensemble size increases (Burgers *et al.*, 1998). For smaller ensembles, however, the EnKF is severely rank deficient and its background covariance estimates suffer from sampling errors, including spurious long distance correlations.

In practice, the forecast covariance matrix for contemporary global weather models is enormous: the state vector at operational resolution contains on the order of 10^{10} elements, so it is impractical to compute the analysis explicitly. To circumvent this difficulty, contemporary meteorological data assimilation schemes fall roughly into two classes: variational schemes, which use an iterative method and a linearized version of the forecast model, and ensemble schemes, which use sets of forecasts started from statistically equivalent initial conditions to provide low-rank, Monte Carlo estimates of the forecast error covariance matrix. There is a large literature on both classes. Overviews of the underlying theory may be found in the review articles by (Ghil and Malanotte-Rizzoli, 1991) and (Talagrand, 1997), as well as books by (Lewis *et al.*, 2006), (Daley, 1992), and (Kalnay, 2003). A detailed description of ensemble data assimilation methods may be found in (Evensen, 2009) and at the European Center for Medium-Range Weather Forecasts' 4DVAR operational data assimilation scheme at www.ecmwf.int.

1.2 Data Assimilation in the Ionosphere

The Earth-system component of interest in this dissertation is the ionosphere, which is a layer of the upper atmosphere that is characterized by its significantly ionized composition and electron density population. The ionosphere's electron density structure varies considerably over its spatial extent and displays temporal variability with diurnal, seasonal, and solar cycles. Many climate-like dynamics of the ionosphere have been reproduced with first-principles models, but the prediction and monitoring of space weather is still a major challenge due to the relative lack of observations and an incomplete understanding of the ionosphere's response to its temporally varying external drivers. Specification and prediction of ionospheric space weather is of practical interest due to its potential adverse effects on high-frequency communication systems, navigation systems that use Global Positioning System satellites (Dubey *et al.*, 2006), satellite remote sensing and power grids located in the middle to high latitudes (Stauning, 2013).

The ionosphere is a system that responds rapidly to changing external conditions, which must be continually estimated in addition to the ionospheric state. The primary ionospheric drivers include solar activity, geomagnetic conditions and the state of the charge-neutral upper atmosphere. Of particular interest are extreme space-weather events, which are typically associated with powerful solar events, such as solar flares. During extreme events, the ionosphere and its dynamical drivers undergo significant changes: ionization and heating rates rise sharply, geomagnetic activity undergoes significant disturbances, and the upper atmosphere expands due to enhanced pressure fields and wind circulation that modifies neutral and ionized composition. See (Tsurutani *et al.*, 2009) for an overview of the effects of solar flares on the ionosphere and (Buensanto, 1999) for overview on ionospheric storms.

General circulation models of the ionosphere typically model the ionosphere and thermosphere, and their associated electrodynamical processes, explicitly as a single system (Matsuo and Araujo-Pradere, 2011). The main dynamical variables of interest in this setting include neutral winds and the thermosphere's neutral composition, which mainly consists of atomic and molecular oxygen. However, other important drivers, such as the ionosphere's response to changing solar activity and its coupling with the magnetosphere in high-latitude regions are represented with auxiliary empirical models. Thus the influence of these key dynamical drivers is specified through parameterized inputs, which makes parameter estimation an important component of ionospheric data assimilation.

Relative to the lower atmosphere and oceans, space-weather prediction efforts in the ionosphere through data assimilation have not been as prevalent due to the relative lack of observations. This is quickly changing, however, as observing networks for the ionosphere are growing quickly and this trend is expected to continue (Schunk *et al.*, 2003). As the ionospheric observational network continues to grow, data assimilation is becoming an increasingly important tool for space weather monitoring and modeling. Data assimilation techniques that have proven to be effective in the lower atmosphere, such as ensemble Kalman filter (EnKF) techniques, are also being proposed for the ionosphere. In particular, the mandatory inference of unobserved variables and parameters makes EnKF approaches attractive for the ionosphere.

One of the primary challenges for ionospheric space-weather prediction arises from the ionosphere's compliant dynamics, in which the main variable of interest (electron density) is continually influenced by its external drivers and its updated state has little feed back on their evolution. Thus, predictions of the ionospheric state largely depend on the correct specification of these external drivers, whereas in the lower atmosphere short term predictions have more dependence on the initial state. The

ionosphere's compliant dynamics require that data assimilation systems be able to infer the ionosphere's drivers from the available observations, which are mostly of electron density. Since electron density is largely a passive variable, the challenge for data assimilation systems is to infer the state of the ionosphere's dynamical drivers from the observations of the electron density field that these drivers produced.

Another challenge in ionospheric weather prediction is that the response of the ionosphere to changing solar and geomagnetic activity are typically specified parametrically. The parametric representation of these mechanisms is likely to overly simplify the global response of the ionosphere to changes in these drivers, which may not adequately describe mechanisms associated with the ionosphere's response during extreme events and introduce significant model error. Additionally, the parametric representation of these drivers does not offer any predictive capabilities, which limits its general predictability in the ionosphere, especially during extreme events, where solar and magnetospheric conditions undergo significant changes. Although these issues are more closely related to ionospheric modeling, data assimilation systems can help improve modeling efforts to an extent through the adjustment of solar and magnetospheric mode inputs using the assimilated observations.

This dissertation is composed of several studies in which the local ensemble transform Kalman filter (LETKF) (Hunt *et al.*, 2007) is proposed for ionospheric data assimilation. The LETKF is an ensemble square root filter that computes an analysis (updated forecast) that provides a new set of initial conditions for a subsequent short-range forecast and an updated, low-rank estimate of the forecast error statistics. The analysis is computed independently grid point by grid point, by assimilating nearby observations simultaneously, making the LETKF amenable to efficient implementation on a parallel computer cluster. In the studies presented, the LETKF is used to improve state specification and prediction of the time-evolving three-dimensional

field of electron density by assimilating synthetic observations of electron density into an ionospheric model with operational capabilities. This dissertation presents two primary appendages that are formulated within the LETKF framework that are designed to address some of the challenges in ionospheric data assimilation described above, particularly during extreme space-weather events. Although these strategies are designed and applied in the context of data assimilation for the ionosphere, they are readily applicable to other components of Earth-system models.

The first appendage, presented in Chapter 5, is a methodology that quantifies the relative sensitivity of each analyzed state variable with respect to each of the assimilated observations using the observation influence diagnostic. This formulation is used as the basis of a targeted observation strategy that determines where to optimally place additional observations to improve state specification in a chosen variable. This strategy helps address issues related to the sparsity of observations in the ionosphere by optimally selecting where observations should be placed during extreme space-weather events and also serves as a diagnostic tool for the design of future observing networks. For example, future development of observational networks may use this tool to examine which dynamical variables should be targeted to improve short-term predictions of electron density or other mechanisms of interest during geomagnetic storms. It is particularly notable that this strategy can be used to improve estimates of unobserved dynamical variables (which drive the electron density evolution) using targeted observations of electron density, even though electron density is a largely passive variable. This is promising for ionospheric data assimilation because the correct specification of these drivers is important for prediction, but their observability is much more limited in comparison to electron density.

The second appendage, presented in Chapter 6, develops a strategy for the estimation of model bias to help addresses issues related to model error and mitigating

its effects on electron density predictions. As mentioned above, a primary source of model bias in ionospheric modeling arises from the parametrized representation of solar and magnetospheric drivers, which are often misspecified during extreme events. This proposed bias estimation methodology offers greater flexibility in the representation of systematic errors that arise from the misrepresentation of storm-time effects due to the parametric representation of these drivers. Application of this bias estimation strategy may help remove bias effects from short-term space-weather predictions and may also serve as a tool to discover systematic biases that may be present in ionospheric models. The numerical experiments presented in Chapter 6 show that considerable improvements in electron density predictions may be obtained even when using this bias estimation strategy with a simple model for evolution of the bias estimates..

The layout of this dissertation is as follows. Chapter 2 describes the ionosphere and ionospheric modeling in terms of data assimilation and space-weather prediction. Chapter 3 introduces data assimilation using Kalman filter based approaches and describes the LETKF. Chapter 4 describes an initial study, which is published in *Physica Scripta*, where the LETKF is applied to estimate ionospheric weather prediction with an idealized regional ionospheric model. Chapter 5 presents a targeted observation strategy, based on the observation influence diagnostic, which is applied in observing system experiments to improve ionospheric space-weather specification during a geomagnetic storm event. This targeted observation strategy is published in the *Journal of Geophysical Research: Space Physics*, which is a leading journal in ionospheric studies. Chapter 6 presents a bias correction strategy, designed for sparsely observed and high dimensional system, which is also applied to improve space-weather prediction during a geomagnetic storm event. The proposed bias estimation strategy is under review in the *Journal of Geophysical Research: Space Physics*.

Chapter 2

IONOSPHERIC WEATHER PREDICTION

The ionosphere is an layer of the upper atmosphere that extends from about 60 km to 1000 km in altitude. It is characterized by its significant ionized composition and is partly embedded over multiple atmospheric layers, including the upper mesosphere and thermosphere. Solar radiation is absorbed in these atmospheric layers at different wavelength ranges, primarily by atomic oxygen (O_1), molecular oxygen (O_2) and molecular nitrogen (N_2) gases, which ionizes the atmospheric neutral composition. Ionization patterns vary considerably over the ionosphere's spatial extent and exhibits temporal variability with diurnal, seasonal, and solar cycles. Of particular interest is the population of free electrons resulting from the ionization of thermospheric composition.

Although the electron density population begins to form around 30 km altitude, it begins to have a noticeable effect on radio signals at about 60-90 km altitudes, and that is generally the altitude where the ionosphere begins. The vertical structure of the ionosphere consists of three primary layers, known as the D, E, and F layers. Each layer differs in its influences on radio waves and the manner in which they are created throughout the diurnal cycle. Although the altitude extent of each layer varies with solar activity and location, each layer is generally host to a local electron density peak. Figure 2.1(a) shows the typical structure of an electron density vertical profiles as computed with the thermosphere-ionosphere-electrodynamics general circulation model (TIEGCM). The horizontal axis is electron density in el/cm^3 and the vertical axis is altitude in km. The vertical profiles shown are averaged over the day-time (red) and night-time (blue) sectors at high-latitude regions. Figure 2.1(b) shows similar

vertical profiles over the equatorial regions.

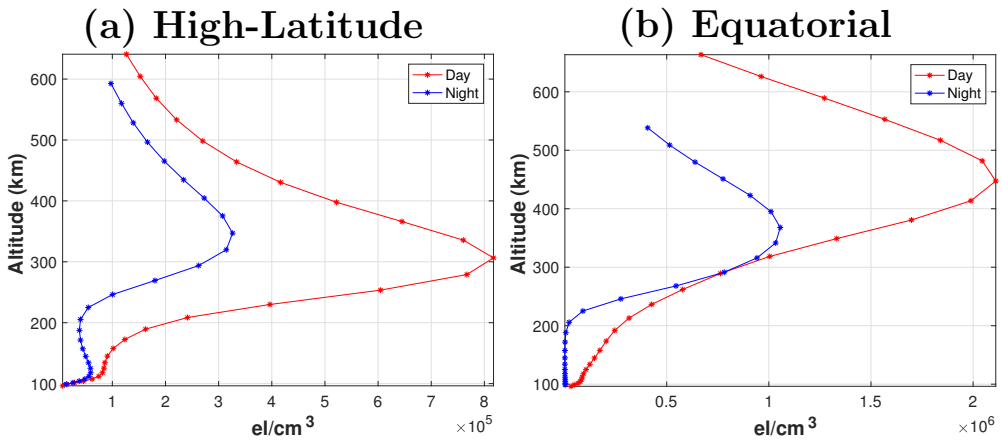

Figure 2.1: (a) Examples of electron density vertical profiles, obtained with the TIEGCM, located at high-latitude regions, averaged over the day (red) and night (blue) sectors. The horizontal axis is electron density in el/cm^3 and the vertical axis is altitude in km. (b) Analogous plot of vertical profile, but located in the equatorial region.

The ionosphere's electrically-charged content is the basis for radio communication. Each ionospheric layer is able to reflect a radio signal up to a certain frequency threshold, called the critical frequency, which is directly proportional to electron density. The D layer is the lowest altitude region, extending from about 60 km to 90 km altitude and is present only during the day, when solar radiation is strong enough to sustain it. Its main effect is the absorption or attenuation of radio communication signals in the low-frequency and medium-frequency ranges. The E layer is the ionospheric layer directly above, existing at about 100 km to 125 km altitude. The E layer also attenuates radio signals, but its ionization is strong enough to refract radio signals to the degree where they may be returned to Earth. Its intensity is decreased during the night, but is still strong enough to attenuate signals in the lower portions of the high frequency (HF) range.

The F layer is the densest ionospheric layer and typically contains the maximum electron density. Its exact altitude varies considerably with location and solar activity.

For example, the F-layer in the vertical profiles shown in Figure 2.1 can vary from about 300 km to 450 km altitude. The F layer reflects signals in the HF range and is used to establish long-distance communication throughout the day and night sectors. Since the F layer is host to global electron density peaks, any signal with frequency higher than the F layer critical frequency will penetrate the ionosphere and continue on to outer space. Signals at that frequency range are used for satellite communication. For a comprehensive description of the physical mechanisms involved in the formation of these ionospheric layers, see the book by (Kelley, 2009).

2.1 Ionospheric Modeling

The ionosphere is a system that exhibits compliant dynamics, in which its state responds rapidly to continually changing external conditions. The nature of these dynamics has important implications on the reliability of ionospheric predictions, since the correct specification of the initial ionospheric state may not produce accurate predictions, if estimates of its time-dependent external drivers are not adequately adjusted Chartier *et al.* (2013). Thus, ionospheric-weather prediction is restricted to the predictability of its external forcing, but understanding of these external forces is often limited, which imposes limitations in the range of phenomena that can be reproduced with ionospheric models ((Siscoe and Solomon, 2006)). Further improvement in ionospheric modeling is thus dependent on the understanding of dynamics of its external drivers, as well as the ionosphere's response to these changing drivers. The most important drivers of the ionosphere include solar activity, geomagnetic conditions, the state of the charge-neutral thermosphere, and planetary gravity waves arising from the mesosphere.

The ionosphere displays climate-like dynamics, which are slow in time-scale, and many of its features have been reproduced with ionospheric models derived from first

principles. However, as the interface between Earth and outer space, the ionosphere experiences significant external forcing from neutral winds, solar radiation and magnetospheric forcing. As a result, the ionosphere's electron density structure displays complex weather-like behavior, commonly referred to as space weather, that can vary on time scales from hours to days. The prediction and monitoring of space weather is still a major challenge due to lack of observations with sufficient spatial and temporal resolution and reliable estimates of ionospheric drivers.

The ionosphere is of great practical interest due to its influence on many communication systems. Uncertainty in ionospheric weather prediction can have adverse effects on high-frequency communication systems, over-the-horizon radars, and navigation systems that use Global Positioning System (GPS) satellites (Coster *et al.*, 2003). Improvement in the estimate of the spatial distribution of electron density as well as mesoscale structures throughout the ionospheric domain can help mitigate these unwanted effects.

2.1.1 Extreme Space-Weather Events

Ionospheric storms refer to events where the ionosphere undergoes large-scale changes in ionization distribution relative to quiet time conditions and are typically caused by massive energy inputs in the upper atmosphere from extreme solar events. As a result of such solar events, the ionospheric state and its drivers undergo significant changes: ionization and heating rates rise sharply, geomagnetic activity undergoes significant disturbances, and the thermosphere expands, creating significant pressure fields and enhanced wind circulation patterns. Thus ionospheric storms are typically associated with geomagnetic storms, which refer to disturbances in the Earth's magnetic field due to the impact of solar wind particles on the Earth's magnetosphere, and thermospheric storms, which refer to the large-scale modification of

neutral composition due to the enhancement of neutral wind circulation as the thermosphere expands. See (Tsurutani *et al.*, 2009) for an overview of the effects of solar flares on the ionosphere and see (Buensanto, 1999) for an overview of ionospheric storms. A introductory description of the space physics in the upper atmosphere is provided by (Prolss and Bird, 2004).

Solar events that are strong enough to cause such events are typically associated with solar flares and coronal mass ejections (CMEs). Solar flares and CMEs are solar events involving enormous eruptions of energy, typically caused from an explosive realignment of the Sun's magnetic field. Solar flares refer to the flash of light associated with the eruption that emit radiation in several bands (white light, ultraviolet, x-ray, gamma rays) and can reach Earth on the order of minutes. CMEs are clouds of magnetized solar particles and plasma that are launched from the sun and typically take 3-4 days to reach the Earth. CME eruptions accelerate energetic particles into the near-Earth space environment, which can disturb the Earth's magnetic field and create currents that drive particles down towards the Earth's poles.

Such variations in solar inputs drive many of the primary ionospheric mechanisms. Solar winds compress the Earth's magnetosphere which creates electric fields in the high-latitude ionosphere, that may extend to lower latitudes during stronger solar events. These electric currents couple the high-latitude ionosphere to the magnetosphere. Strong electric fields and increased ionospheric conductivity from energetic particle impact cause in electric currents and frictional heating due to collisions between accelerated plasma and ambient neutrals. Heating of neutral composition expands the thermosphere, creating pressure gradients, which then modify the global circulation of neutral winds and has a major effect on the ionospheric state. During extreme events, expansion of the neutral thermosphere is sudden and can modify the global thermospheric circulation significantly. In particular, there may be composi-

tion disturbance zones, which travel from high latitudes to mid latitudes, and appear in the ionosphere as large-scale traveling disturbances that can change the F-layer altitude.

Energy inputs to the upper atmosphere typically consist of enhanced electric fields, currents, and energetic particle precipitation. Energetic particles refer to highly energetic electrons, protons, neutrons, and ions that are accelerated into the atmosphere and enter primarily through polar regions. These energetic particles often ionize and dissociate atmospheric constituents. High latitude energy precipitation and magnetospheric convection are important drivers for global ionospheric models. During geomagnetic storms, the cross-polar-cap potential drop, which refers to the difference between the maximum and minimum of electric potential in high latitude regions and is an indicator of the of energy influx from the solar winds, increases significantly, leading to intensification and expansion of the magnetospheric convecting electric fields. These magnetospheric drivers have significant effects on the high-latitude plasma structure.

The behavior of the ionosphere during geomagnetic storms is also influenced by the evolution of the thermosphere, which consists of changes in the neutral winds and composition that lead to changes in rates of production and loss of ionization. This results in ionospheric positive and negative storm effects. Typically, the ionosphere's response is positive during the initial phase of a storm and then is negative afterward, but the duration and magnitude of each phase depends on latitude and season. Due to the intimate coupling between the ionized and charge-neutral composition, models of the upper atmosphere typically model the ionosphere and thermosphere as a single system, to obtain self-consistent solutions for the state of each atmospheric layer.

Recent examples of significant geomagnetic storms include the 1989 storm, which caused a blackout in a large portion of the power grid in Quebec, Canada (Bolduc,

2002), the 2001 Bastille event (Webber *et al.*, 2002), the 2003 Halloween storms (Weaver, 2004), and the storm of 26 September 2011 (Klimenko *et al.*, 2015). Even stronger geomagnetic events are possible. For example, a series of powerful solar flares observed in July 2012 missed the Earth by about a margin of about 9 days, and are estimated to have induced geomagnetic storms that rival the Carrington event of 1859, if the solar ejections had been Earth-directed (Ngwira *et al.*, 2013). This powerful solar event occurred during a relatively quiet solar cycle (Baker *et al.*, 2013), so it is a plausible extreme space-weather scenario for which society should be prepared.

2.1.2 External Driver Representation

Global circulation models of the ionosphere typically represent its interaction with the thermosphere explicitly. However, the ionosphere's response to other key ionospheric drivers is typically represented through parameterized inputs. In particular, the ionosphere's response to solar activity and magnetospheric inputs at high-latitude regions are represented with auxiliary empirical models. The empirical representation of the key dynamical drivers presents a major challenge for ionospheric forecasting, since parameterized inputs offer no predictive capabilities. Additionally the parametric representation of complex ionospheric response to variations in solar and geomagnetic activity may be overly simplified. The solar and magnetospheric inputs to the TIEGCM, which is the ionospheric model of primary focus on in this dissertation, are discussed below.

Solar Conditions

The main source of ionization in the ionosphere, and thus one of its primary dynamical drivers, is the absorption of solar radiation in the thermosphere, in the soft X-ray ultraviolet (XUV, 0.05 nm — 30 nm), extreme ultraviolet (EUV, 30 nm – 120 nm)

and far ultraviolet (FUV, 120 nm — 200 nm) wavelength range, primarily by O_1, O_2 and N_2 neutral gases, through photon ionization and dissociation. The TIEGCM represents effects of solar irradiance and its variability through auxiliary empirical models. The default solar input model, which is used in the TIEGCM configuration for this research, is the EUV flux model for aeronomic calculations (EUVAC) irradiance model (Richards *et al.*, 1994). The EUVAC model specifies important ionospheric processes related to solar activity in the spectral range from 5 nm — 105 nm. The EUVAC model contains a reference spectrum at solar minimum and a wavelength-dependent solar variability that is parameterized by the $F_{10.7}$ index, which is a daily measurement radio flux at 10.7 cm wavelength, and its 81-day average. The solar input directly affects important ionospheric processes such as ionization rates, including direct and photoelectron, dissociation rates, and heating rates of neutrals, electrons and ions, and thus has a major impact on the production of electron density.

The $F_{10.7}$ index is an indicator of solar activity and is one of the longest running records of solar activity (measured since 1947). It can be reliably measured on a day-to-day basis from the Earth's surface in all types of weather, so its data set is robust and has few gaps or calibration issues. It is reported in terms of "solar flux units", which correspond to 10^{-22} W/m^{2}Hz and its values range from below 50 to above 300 to describe different levels of solar activity. Because of its long record, the $F_{10.7}$ index provides climatology of solar activity over six solar cycles. The extreme ultraviolet emissions that impact the ionosphere and modify the upper atmosphere track well with the $F_{10.7}$ index. Many UV emissions that affect the stratosphere and ozone also correlate with the $F_{10.7}$ index. The $F_{10.7}$ index data set can be found at the space weather prediction center (SWPC) division of NOAA.

Geomagnetic Activity

Another key driver of ionospheric dynamics is geomagnetic activity. The ionosphere is strongly coupled to the magnetosphere in high-latitude regions but is often approximated with auxiliary empirical models due its complexity. Geomagnetic activity is typically characterized with the Kp index, which is a widely used index derived from the horizontal component of geomagnetic field disturbances. Kp indices are provided every 3 hours and the index values range from 0 to 9, to describe geomagnetic conditions ranging from quiet to extremely disturbed. The global Kp index is then obtained through an average of the local station Kp indices. Kp indices range from 0 to 9, where indices below 5 correspond to quite or minor geomagnetic activity, indices between 5 and 6 correspond to mild-moderate geomagnetic storms, and indices greater than 6 correspond to strong-severe storms.

The default magnetospheric input, which is used in the TIEGCM configuration for this research, is the Heelis model (Heelis *et al.*, 1982). The Heelis model is parameterized with the Kp index to calculate high-latitude auroral precipitation (Hp), which specifies energy inputs, and the cross-tail potential (Cp), which specifies ion convection patterns. The formulation is shown in Equations (2.1) and (2.2). Historical records of Kp indices are provided by the National Oceanic and Atmospheric Administration (NOAA).

$$Cp(Kp) = 153.13 + 15\mathrm{Kp} + 0.8\mathrm{Kp}^2 \tag{2.1}$$

$$Hp(Kp) = \begin{cases} 16.82e^{(0.32Kp)} - 4.86 & Kp \leq 7 \\ 153.13 + \frac{Kp-7}{9-7}(300 - 153.13) & Kp > 7 \end{cases} \tag{2.2}$$

The Heelis model specifies important high-latitude processes in the ionosphere, such as electric field patterns and auroral energy inputs, for different levels of geomagnetic activity. These inputs are specified directly with the Heelis model above 20° co-latitude and computed explicitly below 35° colatitude. Between 35° and 20° colatitudes, the electric field is computed as a linear combination of the explicit and parametric calculations. The TIEGCM allows for other magnetospheric inputs, such as the Weimer model (Weimer, 2005) and AMPERE model (Anderson *et al.*, 2014), but they are not considered for this study.

A time series of the solar and magnetospheric indices described above is shown in Figure 2.2 during the geomagnetic storm event of 26-27 September 2011, as published by NOAA. The top row shows the indices ten days before and after 26 September 2011. Figure 2.2(a) shows that there is a sudden increase in solar activity prior to 26 September, corresponding to enhanced solar ejections, and Figures 2.2(b)-(c) show a corresponding increase in geomagnetic disturbance slightly after 26 September. The second row shows the same indices during 26-27 September. Figures 2.2(e)-(f) show that the onset of geomagnetic disturbance beings around 10:00 UTC on 26 September, which led to a subsequent geomagnetic storm event

Figure 2.3 shows TIEGCM simulations of the temporal evolution of the global electron density field, in el/cm^3, during 25 September, when geomagnetic conditions were relatively quiet, and after the onset of geomagnetic disturbances on 26 September 2011, at the times indicated. The color scale denotes electron density and the maps are averaged from 250 km to 500 km altitudes. Note the formation of electron density enhancement that forms over high latitude regions during the initial storm phase (14:30 UTC), and that this enhancement grows significantly during the main phase of the storm (17:30 UTC and 20:30 UTC). Although geomagnetic conditions begin to relax after 23:30 UTC, the electron density enhancements still extend into

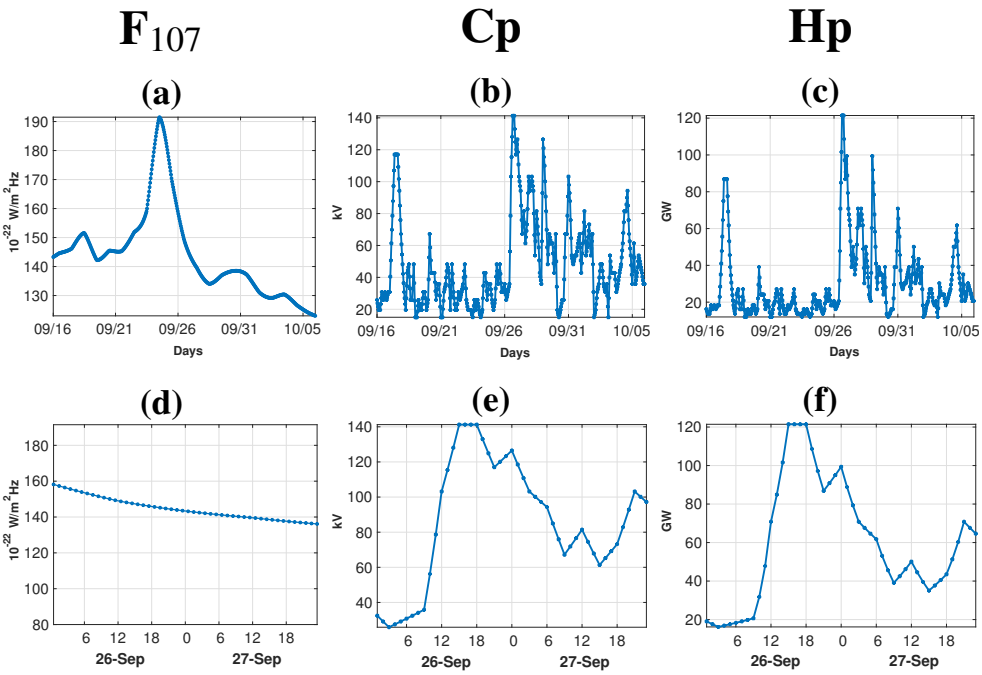

Figure 2.2: (a)-(c) Time series of solar and magnetospheric indices: (a) solar flux ($F_{10.7}$) in solar flux units (10^{-22}W/(m^{2}Hz)), (b) cross-tail potential (Cp) in kV, and (c) hemispheric power (Hp) in GW . The time origin is at 00:00 UTC on 26 September 2011, with the time interval spanning from ten days prior to ten days after. (d)-(f) Time series of the same indices beginning 00:00 UTC on 26 September 2011 and continuing to 23:30 UTC on 27 September 2011.

the higher-latitudes, compared to the period of quiet geomagnetic conditions. This simulated evolution of electron density during geomagnetic disturbances demonstrates the degree to which the global distribution of electron density changes relative to geomagnetically quiet conditions. Data assimilation is becoming an increasingly useful tool to capture and predict these ionospheric deviations from the normal ionosphere during extreme events, so that uncertainties in space-weather specification can be mitigated.

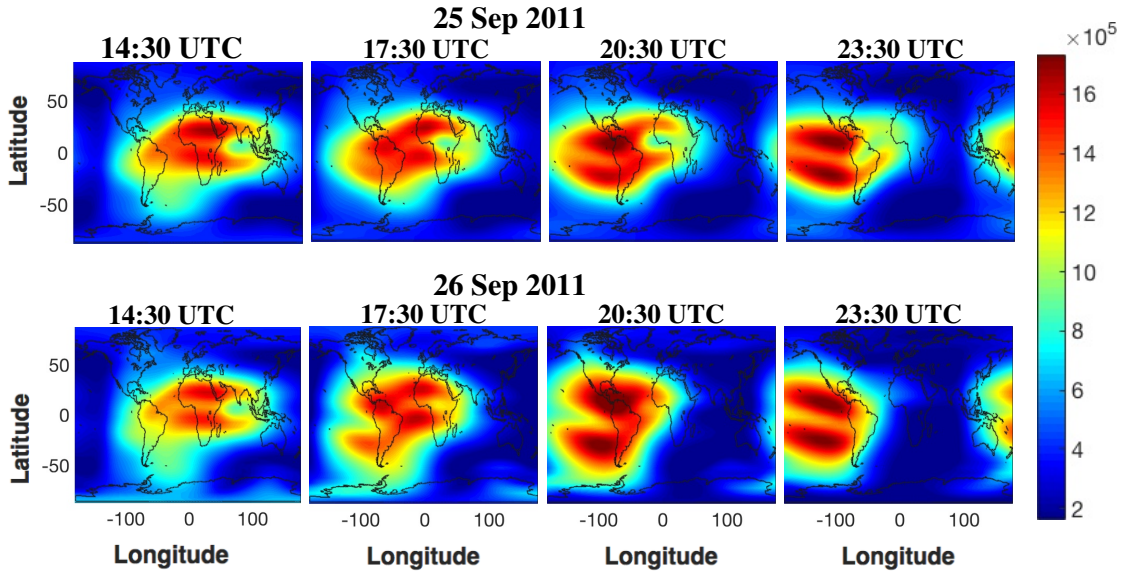

Figure 2.3: (a) Comparison of electron density distribution during quiet (25 September) geomagnetic conditions and during the onset of geomagnetic disturbances (26 September) at the indicated times. The color scale denotes electron density in el/cm^3.

2.2 Data Assimilation in the Ionosphere

While data assimilation has developed into a mature practice for weather prediction in the lower atmosphere and in oceans, development in the ionosphere has been slower, primarily due to the scarcity of observations. The majority of available measurements in the ionosphere are of electron content. The first sources of observations where ionosondes, which reflected signals from the ionosphere at certainty frequency ranges, revealing the electron content and its altitude at different ionospheric layers. Over the past few decades, the observational network has grown vastly, with the development of satellite communication, and it is expected that this trend will continue (Schunk *et al.*, 2003). Newer observation sources include ground-based digisondes, which are generally limited to space above land. Radio occultation techniques on board satellite missions have helped provide global observational coverage of the

ionosphere, including space above oceans. However, observations of other variables, such as winds or temperature remain relatively scarce.

With an increasing ionospheric observation dataset, data assimilation techniques that have proven to be effective for the lower atmosphere, such as ensemble-based techniques, are being tested in the ionosphere. However there are aspects of the ionosphere that are different than the lower atmosphere, which may require more than a simple transplant of data assimilation schemes. Despite the expanding observing network, the observational data set is still significantly smaller than that of the atmosphere and the variety of observed dynamical variables is smaller. Additionally, compliant ionospheric dynamics imply that the predictability of the ionosphere can be advanced only as far as the predictability of its external forcing. These considerations must be accounted for in the development of operational data assimilation systems. Many global-scale characteristics of ionospheric storms have now been reproduced by global first-principles theoretical models, but local features during specific storms are difficult to accurately predict, largely due to uncertainties in the inputs required by the models. Data assimilation is a promising tool to understand and reproduce these ionospheric mechanisms.

A key component of ionospheric weather prediction is the correct representation of the solar and magnetospheric conditions, which are specified with parameterized inputs in the TIEGCM. However, geomagnetic indices are notoriously difficult to measure accurately during periods of magnetospheric disturbance and solar irradiance patterns can deviate significantly from proxy models. Storm time effects can be difficult to model and predict in sign, magnitude and direction. The mechanisms behind these processes are are not completely understood and may pose significant uncertainties, not only due to sub-optimal drivers, but from inherent limitations of the parameterizations. For example, the study by (Pedatella *et al.*, 2009) discusses

the significant enhancement in F2-layer electron density peak and altitude during the geomagnetic storm of 15 December 2006. In particular, these storm time-effects were observed to be very long-lasting after the main phase of this geomagnetic storm. Observational analysis on the initial phase of this geomagnetic storm are presented by (Lei *et al.*, 2008b) and a study about observed traveling ionospheric disturbances are discussed in (Lei *et al.*, 2008a).

Some recent efforts to apply data assimilation to the ionosphere include the Global Assimilative Ionospheric model (USC/JPL GAIM), which includes four-dimensional variational and Kalman filter methods (Wang *et al.*, 2004); the Ionospheric Data Assimilation Three-Dimensional (IDA3D) model, which uses a three-dimensional variational data assimilation scheme (Bust and Mitchell, 2008); the Utah State University Global Model Assimilation of Ionospheric Measurements (USU-GAIM), which uses a Kalman filter (Scherliess *et al.*, 2004); a global three-dimensional ionospheric electron density reanalysis using a Kalman filter to assimilate multi-source data (Yue *et al.*, 2012); a global ionosphere-thermosphere model (GIMT) that uses the Local Ensemble Transform Kalman Filter (LETKF) (Koller *et al.*, 2013); and the Thermosphere-Ionosphere Global Circulation Model (TIEGCM) that uses an ensemble Kalman filter from the Data Assimilation Research testbed (DART) ((Lee *et al.*, 2012).

Chapter 3

DATA ASSIMILATION

In framework of spatiotemporal prediction of Earth-system models, the state estimate of a given system is typically produced with a numerical forecast model, which approximates the spatiotemporal evolution of the system dynamics. Due to the inherent limitations in numerical models, stemming from an incomplete understanding of the system dynamics and physics, it is often necessary to use information obtained directly from measurements or observations of the system to update estimates of the model, and to synchronize the model predictions with the system that produced the observations. The improved state estimate is then used as the initial condition to restart the model and propagate the system state and to produce a forecast at a future time. The procedure by which observations are used to update the system state is broadly referred to as *data assimilation*. Generally, this procedure is a discrete state estimate/update approach in which the forecast model is used to predict the system state and the data assimilation scheme is used to update the system state, based on the available observations

The data assimilation scheme used in this dissertation is the local ensemble transform Kalman filter (LETKF). The LETKF is described in some detail over the next few sections, but the reader is referred to Hunt *et al.* (2007) for a complete description and mathematical justification of this scheme. The LETKF is a type of ensemble square root filter, which is in turn based on the classical Kalman filter. The Kalman Filter is a method by which two estimates of a given state can be optimally combined to produced the best linear unbiased estimate of the "true" state (Talagrand, 1997). The Kalman filter requires information about the statistical error in the two pieces

of information in addition to the estimates themselves. Generally, the procedure is independent of the source of these two pieces of information, but since the focus of this research is on state estimation of Earth-system models, it is assumed that one estimate is the output of a forecast model and the other is from an empirical measurement or observation.

This chapter begins with a simple scalar example of the Kalman filter, and then progress to a more general multi-dimensional example. Practical problems associated with the direct application of the Kalman filter in practice and approaches to circumvent these difficulties are also discussed. Finally, the approach used for LETKF and a discussion of its computational efficiency is presented.

3.1 Kalman Filter: Scalar Example

Consider some unknown scaler quantity of interest, say temperature, T, defined at a point in space and time. Also suppose there is a temperature estimate, $T_b = T + \epsilon_b$, made with a model (say a thermostat) and a temperature measurement, $T_o = T + \epsilon_o$, made with a thermometer. The quantities ϵ_b and ϵ_o are error terms which reflect that T_b and T_o are noisy estimates. The Kalman filter assumes that ϵ_o and ϵ_b are independent, normally distributed random variables with zero mean and variance given by σ_o^2 and σ_b^2, respectively. The probability distribution (PDF) of the errors in each estimate is given by:

$$p_b(t) = \frac{1}{\sqrt{2\pi}\sigma_b} e^{-(t-T_b)^2/(2\sigma_b^2)} \tag{3.1}$$

$$p_o(t) = \frac{1}{\sqrt{2\pi}\sigma_o} e^{-(t-T_o)^2/(2\sigma_o^2)} \tag{3.2}$$

The optimal combination of these two PDFs is given by the most likely state in their joint distribution. T_o and T_b are assumed to be independent, so their joint PDF is

given by

$$p(t) = Ce^{-\left((t-T_b)^2/(2\sigma_b^2)+(t-T_o)^2/(2\sigma_o^2)\right)}, \tag{3.3}$$

where C is the constant acquired through multiplication and normalization of the joint distribution. The Kalman filter finds a temperature, T_a, that maximizes Equation (3.3), or equivalently minimizes

$$J(t) = (t - T_b)^2/\sigma_b^2 + (t - T_o)^2/\sigma_o^2. \tag{3.4}$$

To find the minimizer, the quadratic Equation (3.4) is expanded. Completing the square on the resulting expression yields

$$J(t) = \left(\frac{1}{\sigma_b^2} + \frac{1}{\sigma_o^2}\right)\left[t - \frac{\sigma_o^2 T_b + \sigma_b^2 T_o}{\sigma_b^2 + \sigma_o^2}\right]^2 + F, \tag{3.5}$$

where F is constant with respect to t. The minimizer of $J(t)$ is given by

$$T_a = \frac{\sigma_o^2 T_b + \sigma_b^2 T_o}{\sigma_b^2 + \sigma_o^2}. \tag{3.6}$$

T_a is referred to as the *analyzed* temperature estimate, or simply the analysis. Under the statistical assumptions about the errors in the estimates mentioned above, T_a is the best linear unbiased estimate of the true temperature The same minimizer may be obtained by differentiating Equation (3.4) with respect to t, setting the result to 0 and solving for t, but the approach shown in Equation 3.5 is applicable in the case where temperatures are multi-dimensional, so it is shown here for clarity.

Equation (3.5) also shows that the variance in T_a is given by

$$\frac{1}{\sigma_a^2} = \frac{1}{\sigma_b^2} + \frac{1}{\sigma_o^2} = \frac{\sigma_o^2 + \sigma_b^2}{\sigma_o^2 \sigma_b^2}, \tag{3.7}$$

which implies that the variance in T_a is smaller than both of the background and observed estimates. Defining the quantity

$$K = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_o^2}, \tag{3.8}$$

which is often referred to as the "Kalman gain", Equations (3.6) and (3.7) become

$$T_a = T_b + K(T_o - T_b) \tag{3.9}$$

$$\sigma_a^2 = (1 - K)\sigma_b^2. \tag{3.10}$$

Equations (3.9) and (3.10) provide an intuitive description of the Kalman filter update. The difference between the predicted and observed temperatures, $T_b - T_o$, is commonly referred to as the innovation, or the observational increment, since it represents the new information provided by the observation. According to Equation (3.9), the updated temperature is simply the original model prediction, plus some optimally weighted innovation, K, where $0 \leq K \leq 1$.

If the observed measurement, T_o, is extremely accurate relative to the thermostat measurement ($\sigma_o^2 \ll \sigma_b^2$), then $K \approx 1$ and $T_a \approx T_o$, with uncertainty $\sigma_a^2 \approx \sigma_o^2$. An analogous result is obtained when $\sigma_b^2 \ll \sigma_o^2$. If both temperatures have similar uncertainty $\sigma_b^2 \approx \sigma_o^2$, then $K \approx 1/2$ and $T_a \approx (T_o + T_b)/2$, or simply the average of the two estimates.

3.2 Kalman Filter: Multi-Dimensional Case

Now consider a temperature field, $\mathbf{Q}$, defined on a discrete three-dimensional grid, as would the output of an atmospheric model. The grid points can be ordered into an $m \times 1$ vector, $\mathbf{x}^t$, where m is the number of grid points. Similarly, let $\mathbf{x}^b$ be the m-dimensional vector of background temperature estimates. Suppose there is a collection of ℓ observations, organized in an ℓ-dimensional vector, $\mathbf{y}^o$. The observations are measurements taken from the temperature field, $\mathbf{Q}$, but generally need not be located on the model grid where $\mathbf{x}^b$ is defined.

The *observation* or *forward* operator relates the temperature field defined on the model grid to the observation locations. Since temperature is a model variable and

also the observed variable, H is simply an interpolation procedure. More generally, observed variables may not necessarily be state variables, but functions of the state variables, in which case, H is a functional description of the physical relationship between them, which may be non-linear.

The Kalman filter supposes that the background is related to the system state as $\mathbf{x}^b = \mathbf{x}^t + \boldsymbol{\epsilon}^b$, where $\boldsymbol{\epsilon}^b$ is a $m \times 1$ normally distributed random vector, with zero mean. Similarly, the observation vector is assumed to be of the form $\mathbf{y}^o = H(\mathbf{x}^t) + \boldsymbol{\epsilon}^o$, where $\boldsymbol{\epsilon}^o$ is a normally distributed $\ell \times 1$ vector with zero mean. Furthermore it is assumed that $\boldsymbol{\epsilon}^o$ and $\boldsymbol{\epsilon}^b$ are mutually independent. The uncertainties in each of the estimates are now characterized with covariance matrices $\mathbf{P}^b$ and $\mathbf{R}$, respectively. The objective function analogous to Equation (3.4) is given by quadratic cost function:

$$J(\mathbf{x}) = \left[\mathbf{x} - \mathbf{x}^b\right]^{\mathrm{T}} (\mathbf{P}^b)^{-1} \left[\mathbf{x} - \mathbf{x}^b\right] + \left[\mathbf{y}^o - H(\mathbf{x})\right]^{\mathrm{T}} \mathbf{R}^{-1} \left[\mathbf{y}^o - H(\mathbf{x})\right] \tag{3.11}$$

In the case of linear observation operator and model dynamics, Equation (3.11) has a unique minimizer that can readily be found using a similar "complete the square" approach used in Section 3.1. Expanding Equation (3.11) into its quadratic, linear and constant components yields

$$\begin{aligned} J(\mathbf{x}) &= \left[\mathbf{x} - \mathbf{x}^b\right]^{\mathrm{T}} (\mathbf{P}^b)^{-1} \left[\mathbf{x} - \mathbf{x}^b\right] + \left[\mathbf{y}^o - \mathbf{H}\mathbf{x}\right]^{\mathrm{T}} \mathbf{R}^{-1} \left[\mathbf{y}^o - \mathbf{H}\mathbf{x}\right] \\ &= \mathbf{x}^{\mathrm{T}} \left[(\mathbf{P}^b)^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\right] \mathbf{x} - \mathbf{x}^{\mathrm{T}} \left[(\mathbf{P}^b)^{-1}\mathbf{x}^b + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{y}^o\right] \\ &\quad - \left[(\mathbf{x}^b)^{\mathrm{T}}(\mathbf{P}^b)^{-1} + (\mathbf{y}^o)^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}^{\mathrm{T}}\right] \mathbf{x} + \mathbf{C}_1, \end{aligned} \tag{3.12}$$

where $\mathbf{C}_1$ is the constant with respect to $\mathbf{x}$. The goal is to write Equation (3.12) in the form:

$$\begin{aligned} J(\mathbf{x}) &= \left[\mathbf{x} - \mathbf{x}^a\right]^{\mathrm{T}} (\mathbf{P}^a)^{-1} \left[\mathbf{x} - \mathbf{x}^a\right]^{\mathrm{T}} \\ &= \mathbf{x}^{\mathrm{T}} \left[(\mathbf{P}^a)^{-1}\right] \mathbf{x} - \mathbf{x}^{\mathrm{T}} \left[(\mathbf{P}^a)^{-1}\mathbf{x}^a\right] - \left[(\mathbf{x}^a)^{\mathrm{T}}(\mathbf{P}^a)^{-1}\right] \mathbf{x} + \mathbf{C}_2, \end{aligned} \tag{3.13}$$

where $\mathbf{C}_2$ is constant with respect to $\mathbf{x}$. Equating the quadratic and linear terms of equation (3.13) with those of equation (3.12) yields the Kalman filter state update

and its associated covariance matrix:

$$\mathbf{P}^a = \left[(\mathbf{P}^b)^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\right]^{-1} \tag{3.14}$$

$$\mathbf{x}^a = \mathbf{P}^a \left[(\mathbf{P}^b)^{-1}\mathbf{x}^b + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{y}^o\right] \tag{3.15}$$

Equations (3.15) and (3.14) can be written in many different but equivalent ways. Introducing the Kalman gain matrix:

$$\mathbf{K} = \mathbf{P}^a\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1} \tag{3.16}$$

Equations (3.14) and 3.15 become

$$\mathbf{P}^a = \left[\mathbf{I} + \mathbf{P}^b\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\right]^{-1}\mathbf{P}^b \tag{3.17}$$

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{K}\left(\mathbf{y}^o - \mathbf{H}\mathbf{x}^b\right). \tag{3.18}$$

The details of this reformulation are provided in Appendix F. Similarly to the scalar case, the analyzed updated consists of the background state plus a suitably weighted innovation vector, $\mathbf{y}^o - \mathbf{H}\mathbf{x}^b$.

3.2.1 The Analysis Cycle

The analysis cycle is essentially one long model integration, consisting of discrete state estimation and update steps, with the goal of estimating the state trajectory of a dynamical system $\mathbf{u}_i^t$ that is unknown, with the exception of noisy observations available at a set of discrete times t_i. In the simplest case, the dynamics at time t_i are governed by a linear model, $\mathbf{x}_i^t = \mathbf{M}_i\mathbf{x}_{i-1}^t$, and the observations are linear functions of the system state: $\mathbf{y}^o = \mathbf{H}_i\mathbf{u}_i^t + \boldsymbol{\epsilon}_i^o$, where $\boldsymbol{\epsilon}_i^o$ models observation processing noise as a random vector drawn from a normal distribution of mean 0 and covariance matrix $\mathbf{R}_i$.

Let $\mathbf{u}_i^b$ denote the *background*, i.e., an estimate of the current state vector at time t_i, and let $\mathbf{P}_i^b$ be its associated covariance matrix. In practice, $\mathbf{u}_i^b$ is a short-range forecast produced by a dynamical model. Given a vector of observations $\mathbf{y}_i^o$

obtained at time t_i, the *analysis*, $\mathbf{u}_i^a$, and an updated estimate of its uncertainty, $\mathbf{P}_i^a$, are obtained by minimizing Equation (3.11), as given by Equations (3.17) and (3.18). $\mathbf{u}_i^a$ serves as an updated initial condition for a subsequent short-term integration of the dynamical model to time t_{i+1}, which yields a new background state vector $\mathbf{u}_{i+1}^b$ and associated covariance matrix, $\mathbf{P}_{i+1}^b$:

$$\mathbf{u}_{i+1}^b = \mathbf{M}_{i+1}\mathbf{u}_i^a \tag{3.19}$$

$$\mathbf{P}_{i+1}^b = \mathbf{M}_{i+1}\mathbf{P}_i^a\mathbf{M}_{i+1}. \tag{3.20}$$

The model is integrated to a time t_{i+1}, where there is a new collection of observations $\mathbf{y}_{i+1}^o$, to be assimilated and the analysis procedure is repeated to compute $\mathbf{u}_{i+1}^a$ and $\mathbf{P}_{i+1}^a$. This cycle alternates between a forecast step and an update step indefinitely, ideally frequently enough such that model state trajectory remains close to that of the system of interest.

3.2.2 Practical Difficulties

The cycle analysis procedure described in Section 3.2.1 provides an approach that in principle yields optimal state estimation and uncertainty quantification of high-order dynamical systems, but there are numerous practical difficulties that make its direct application unfeasible in operational numerical weather prediction (NWP) settings.

The primary difficulty is associated with the enormous state size of operational atmospheric models. In particular, the computation of Equations (3.17) and (3.18) is highly impractical due to the state size. To circumvent this difficulty, iterative methods are often used to minimize Equation (3.11) directly. This is the basis of

variational data assimilation and it typically requires the development of a tangent linear model of the forecast model to make the minimization problem feasible.

The number of observations in NWP settings is usually considerably smaller than the state size, and observations are often taken from independent sets of instruments, so the associated observation error covariance matrix, $\mathbf{R}_i$ is diagonal or block diagonal; operations involving $\mathbf{R}_i$ or its inverse are not problematic. However, the state size for NWP is typically about $m = 10^6 - 10^7$ depending on model resolution, making operations involving its $m \times m$ covariance matrix, $\mathbf{P}_i^b$, in the minimization of (3.11) highly impractical. Additionally, $\mathbf{P}_i^b$ is often not known precisely and in any case, would be difficult or impossible to invert given its dimension. Furthermore, if the observation operator, $\mathbf{H}_i$ is non-linear, which is often the case in practice, Equation (3.11) is not guaranteed to have a unique unbiased minimizer.

Another major source of practical difficulties is the propagation of the forecast uncertainty. Even in the case of linear dynamics, computational costs are still prohibitive due to the state size. $\mathbf{P}^b$ can be propagated as shown in Equation (3.20), but this approach would require the equivalent of $2m$ model integrations, where m is the number of grid points.

These practical limitations make the classical Kalman filter approach impossible to apply in NWP settings. The forecast uncertainty estimate is important for accuracy since it describes important dynamical relationships among dynamical variables and balance constraints. Due to the complexity of operational atmospheric models, the temporal evolution of the forecast covariance matrix is often not known precisely, and thus simplified version are often employed. For example, model covariance matrices are typically approximated as spatially homogeneous and isotropic (Wang *et al.*, 2006). They may be may also be held constant or updated less often, but this ap-

proach comes at the expense of deteriorated analysis estimates due to "errors of the day" that may be missed (Kalnay, 2003).

Although the optimality of the Kalman filter scheme can be rigorously proven only in the case of linear dynamics and complete knowledge of the forecast and observation covariance matrices, data assimilation schemes using variations of Equation (3.11) have been effective in many operational settings (Kalnay, 2003). Data assimilation approaches based on sub-optimal Kalman filter updates, such as ensemble Kalman filters, have also shown promise. In the following sections, an overview of these types of approaches is provided.

3.3 Ensemble Kalman Filter

Correct specification of the forecast covariance matrix, $\mathbf{P}_i^b$ is essential for accuracy in the computation of the analysis, but it cannot be feasibly computed and propagated at its full complexity in numerical weather prediction settings throughout the data assimilation cycle. The ensemble Kalman filter (EnkF) attempts to address both of these issues, by using the sample statistics of a set, or *ensemble* of model realizations. In particular, the most likely state is estimated by the ensemble mean and the model covariance is approximated by the sample covariance of the ensemble. The EnKF addresses the challenge of propagating the forecast uncertainty by evolving each individual ensemble member according to the possibly non-linear dynamics of the model, which presumably yields a reasonable flow-dependent approximation of the forecast uncertainty to account for the "errors of the day", provided that the ensemble size is large enough. Additionally, the EnKF formulation is relatively simple, compared to variational approaches, since there is no need for the maintenance of a model dependent tangent linear model. Additionally, EnKF schemes are essentially

independent of the specific forecast model used, making them highly adaptable to Earth-system models.

In the EnKF formulation, the ensemble mean and the forecast covariance matrix are updated according to the classical Kalman filter equations, given by Equations (3.17) and (3.18). Thus there is an underlying Gaussian assumption on the model and observation errors, as well as linearity in the model dynamics. In the case of a nonlinear model, if each forecasting step of the assimilation cycle is relatively short, the model propagation is approximately linear, so the initial Gaussian distribution can be approximately retained and can be used in the next analysis step. A typical formulation for EnKF schemes is provided below. The EnKF evolves an ensemble of forecasts according to the model dynamics, and when observations become available to be assimilated, the entire ensemble is adjusted in tandem, reflecting the new state estimate as dictated by the observations. The discussion describes the analysis calculations at a fixed time, so the time notation is dropped for clarity.

Denote the background ensemble of size k by $\{\mathbf{x}^{b(i)}\}_{i=1}^{k}$. It is assumed that its ensemble mean,

$$\bar{\mathbf{x}}^{b} = k^{-1} \sum_{i=1}^{k} \mathbf{x}^{b(i)} \tag{3.21}$$

is the best available state estimate before observations are assimilated. Construct the matrix of background perturbations, $\mathbf{X}^{b}$, whose ith column to be $\mathbf{x}^{b(i)} - \bar{\mathbf{x}}^{b}$. Then the forecast covariance matrix is approximated by the sample covariance of the ensemble as

$$\mathbf{P^{b}} = (k-1)^{-1} \mathbf{X}^{b} (\mathbf{X}^{b})^{\mathrm{T}} \tag{3.22}$$

The EnKF produces an "analysis" ensemble, $\{\mathbf{x}^{a(i)}\}_{i=1}^{k}$, whose ensemble mean

$$\bar{\mathbf{x}}^{a} = k^{-1} \sum_{i=1}^{k} \mathbf{x}^{a(i)} \tag{3.23}$$

represents the new state estimate after observations have been assimilated and its associated uncertainty is given by the analyzed covariance matrix

$$\mathbf{P^a} = (k-1)^{-1}\mathbf{X}^a(\mathbf{X}^a)^{\mathrm{T}}, \tag{3.24}$$

where the matrix $\mathbf{X}^a$, is the matrix of analysis perturbations, constructed so that its ith column is given by $\mathbf{x}^{a(i)} - \bar{\mathbf{x}}^a$. The transformation from background ensemble $\{\mathbf{x}^{b(i)}\}_{i=1}^k$ to analysis ensemble $\{\mathbf{x}^{a(i)}\}_{i=1}^k$ is done by updating the background ensemble mean $\bar{\mathbf{x}}^b$ and its associated sample covariance matrix $\mathbf{P}^b$ according to the classical Kalman filter, such that it produces an updated ensemble whose ensemble mean $\bar{\mathbf{x}}^a$ and covariance matrix $\mathbf{P}^a$, match Equations (3.17) and (3.18).

Computing the Kalman gain, $\mathbf{K}$, as it is written in Equation (3.16), is of order ℓ^3 and computing $\mathbf{P}^a$ involves operations on $m \times m$ matrices, where ℓ is the number of observations and m is the number of grid points. By construction, the columns of $\mathbf{X}^b$ sum to zero, implying that its rank is at most $k-1$, and thus the sample forecast covariance matrix as given in Equation 3.22 is also at most rank $k-1$. This implies that $\mathbf{K}$ and $\mathbf{P}^a$ are both at most rank $k-1$, which is typically much smaller than the number of observations and model grid points for NWP applications. Thus the analysis equation can be computed efficiently, for example with an eigenvalue or singular value decomposition.

To construct the analysis ensemble, one must find a collection of forecasts, whose perturbation matrix, $\mathbf{X}^a$, satisfies Equation (3.24) and has the correct ensemble mean. For a given background ensemble and collection of observations, all EnKF-based algorithms yield the same update for the analyzed ensemble mean, $\bar{\mathbf{x}}^a$, and covariance matrix, $\mathbf{P}^a$. However, the ensemble of forecasts that have these two statistics is not unique, and this is where many EnKF-based schemes differ. One method for

computing the analysis ensemble is through an update of the form $\mathbf{X}^a = \mathbf{X}^b\mathbf{Y}$, where $\mathbf{Y}$ is a matrix to be determined.

Each EnKF scheme is based on the choice of $\mathbf{Y}$ and many types of ensemble filters may be constructed with different choices of $\mathbf{Y}$, since for any arbitrary $k \times k$ unitary matrix, $\mathbf{U}$, analysis ensembles constructed as

$$\mathbf{X}^a = \mathbf{X}^b\mathbf{Y}\mathbf{U}, \tag{3.25}$$

will yield the same sample covariance matrix, $\mathbf{P}^a$. It is demonstrated by (Tippet *et al.*, 2003) that these types of schemes are part of a larger class of methods known as square-root filters. Although these square root filters yield the same update on the ensemble mean and sample covariance, higher order statistical moments can be different, which is an important aspect to consider when using a non-linear forecast model. Another important consideration, particularly for NWP settings, is that square root filters may differ on how the computational complexity scales as the number of observations or the state size grows. See (Tippet *et al.*, 2003) for more discussion on these two last points.

Examples ensemble square root filters are the Local Ensemble Kalman Filter (Ott *et al.*, 2004), Ensemble Adjusted Kalman Filter (Anderson, 2001), the Ensemble Transform Kalman Filter (Bishop *et al.*, 2001) and the Local Ensemble Transform Kalman Filter (Hunt *et al.*, 2007). Square root filters are also know as a deterministic EnKF (Whitaker and Hamill, 2001), as opposed to earlier versions of EnKF schemes ((Evensen, 1994), (Burgers *et al.*, 1998),(Houtekamer and Mitchell, 1997)), which are stochastic in the sense that they perturb the observations randomly in generating each ensemble member.

3.3.1 Challenges With Ensemble Kalman Filters

EnKF-based schemes offer a computationally tractable methodology for computing and propagating a flow-dependent forecast covariance matrix, $\mathbf{P}^b$. However $\mathbf{P}^b$ is at most rank $k-1$, and thus can only account for forecast uncertainty in at most $k-1$ directions. Computational limitations restrict the ensemble size to be much smaller than the state size, which yields a severely rank-deficient approximation of the actual covariance matrix of the global state estimate. This rank deficiency can introduce degraded analysis quality due to sampling errors. Since a larger ensemble may not be an option due to practical limitations, there are other approaches to remedy sampling errors as described in the following sections.

Covariance Localization

In operational NWP settings, model state updates using data assimilation occur every one to six hours, depending on the model. Covariance localization suppresses correlations among model variables beyond some prescribed spatial distance to a given grid point so that the update state at that grid point depends only on observations and model variables within a finite distance. This is a desirable approach since it is expected that the nearby observations and model variables have a stronger correlation to the quantities defined at the given grid point so that their covariance are better represented with a reasonably sized ensemble of forecasts. Localization also avoids influence from distant observations, which despite not having a physically meaningful correlations with the given grid point, may have spurious correlations generated from sampling errors of the undersized ensemble. Locally, atmospheric dynamics tend to evolve in a much smaller subspace (Patil *et al.*, 2001), and thus not only can a local region contain the majority of the dynamical information that is relevant to

the grid point, but it will also have an effectively larger-rank approximation of the covariance matrix within that local space. This enables a modestly sized ensemble to effectively capture local model dynamics and directions of growing uncertainty in a computationally affordable way.

Although covariance localization may effectively suppress long-distance spurious correlations, using a localization radius that is too small may also degrade the analysis, since it may suppress physically and temporally meaningful correlations that would be beneficial for the analysis accuracy. A small localization radius may also introduce imbalances or undesired small-scale noise into the analysis, especially if the observation network is sparse. These discontinuities may occur at points where adjacent model grid points are updated using sets of observations with little overlap, that is, where one or more observations are included at one grid point and excluded at the next. A typical approach to avoid this is multiply the EnKF background error sample covariances by a weighting factor that decreases smoothly to zero at finite distance, so that the influence of observations decreases to zero smoothly.

The optimal localization size depends on physical length scales present in the model dynamics as well as the ensemble size that is permitted by practical considerations. The localization radius should ideally be large enough to capture physically meaningful correlations but small enough to exclude long distance spurious correlations as much as possible. (Miyoshi, 2014) suggests that rather than simply increasing the localization radius, it may be more beneficial to adaptively choose a localization geometry in an ellipsoid shape that encapsulates all the physically related points only. This localization geometry is dependent on the dynamics being modeled, and may be different from grid point to grid point, so an adaptive scheme that can distinguishing physically meaningful correlations from spurious ones would be required.

Covariance Inflation

Ensemble Kalman filters approximate the model forecast covariance matrix through the sample covariance of its ensemble. Although, this gives computationally feasible approximation of the temporally varying model covariance, model errors and non-linearities may be sources of errors in this representation of the model uncertainty. However the primary source of error arises from sampling errors due to the size of the ensemble. In the case of large non-linear forecast models, the ensemble size may be several orders of magnitude smaller than the model covariance matrix, which may cause the ensemble to not be statistically representative of the global model dynamics and is likely to misrepresent long distance correlations (Hamill *et al.*, 2001). Strong non-linearities and model error may contribute to this error but such sampling errors can also occur in perfect model scenarios (Anderson, 2007).

Covariance localization compensates for sampling errors to a degree, but the ensemble may still misrepresent model covariance locally in some regions due to the non-linearities of the model. After multiple analysis cycles, the ensemble covariance may not be representative of the actual model uncertainty. This overconfidence may lead to filter divergence, where the ensemble covariance inadequately underestimates the model uncertainty and the influence of observations is reduced greatly, or even ignored.

Covariance inflation is an *ad hoc* approach to compensate for this, where the background perturbation matrix, $\mathbf{X^b}$, is multiplied by a constant factor $\rho > 1$ to increase its ensemble spread, which effectively reduces the confidence that the EnkF scheme has on the background ensemble. Covariance inflation has been helpful in for a variety of ensemble filters but it may be difficult to apply since it can be costly to find covariance inflation values that perform well for a given application, especially

for large models. (Hunt *et al.*, 2007) discusses more details about covariance inflation in the LETKF when applied to the Global Forecast System model. It may be the case that covariance inflation needs to change with space, time, and for different state variables (Anderson, 2009), so adaptive covariance inflation schemes have been proposed and are discussed in (Anderson, 2009; Li *et al.*, 2009; Petrie and Dance, 2010). The adaptive covariance inflation scheme used in this study is the one proposed by (Miyoshi, 2011), and it is briefly described in Section 3.4.4.

3.4 Local Ensemble Transform Kalman Filter

A detailed description and mathematical derivation of the local ensemble transform Kalman filter (LETKF) data assimilation scheme can be found in (Hunt *et al.*, 2007). A simplified derivation of the LETKF approach is provided here and an outline of the algorithm is given in Appendix A. The following discussion describes how the analysis is computed for a fixed analysis time step, the time notation is dropped for simplicity. The LETKF is a local ensemble square-root filter that transforms the background ensemble at a fixed time, $\{\mathbf{u}^{b(j)}\}_{j=1}^{k}$, into an analysis ensemble, $\{\mathbf{u}^{a(j)}\}_{j=1}^{k}$. Each model grid point is processed independently, assimilating only the observations that lie within a prescribed distance, to produce an ensemble of analyzed local state vectors as is described below. The collection of analyzed local ensembles are assembled to form an ensemble of global analyzed states.

The LETKF addresses challenges seen in EnKF scheme through its localized approach and also offers considerable computational efficiency that scales well to large dynamical systems. Each local analysis needs only to account for forecast uncertainty locally, which is typically much lower dimensional than global uncertainty. Thus the global analysis may be accurately computed using only a moderately sized ensemble. Each grid point is updated independently, making the LETKF a naturally parallel

algorithm. Additionally, the operations involved in the an lysis calculation scale with ensemble size and number of observations and not the state size, which allows this approach to be readily scaled to large models.

3.4.1 Analysis Computation

Consider a fixed model grid point, indexed as L and associate a local window, consisting of the region within a prescribed horizontal and vertical distance from its location. The subscript L is used to emphasize that the following quantities are associated with a specific grid point and its associated local window. Denote the d components of the jth background state, $\mathbf{u}^{b(j)}$ that are located within the local window by the d-vector, $\mathbf{x}_L^{b(j)}$. The vector $\mathbf{x}_L^{b(j)}$ contains all the state variables defined at each of the grid points in the local window that are being updated, or analyzed. The local background ensemble mean is given by $\bar{\mathbf{x}}_L^b = k^{-1}\sum_{j=1}^{k}\mathbf{x}_L^{b(j)}$. The $d \times k$ matrix of local background ensemble perturbations, $\mathbf{X}_L^b$, is the matrix whose jth column is given by $\mathbf{X}_L^{b(j)} = \mathbf{x}_L^{b(j)} - \bar{\mathbf{x}}_L^b$. The local model covariance matrix is approximated by the sample covariance of the ensemble as $\mathbf{P}_L^b = (k-1)^{-1}\mathbf{X}_L^b(\mathbf{X}_L^b)^{\mathrm{T}}$.

As discussed in Section 3.3, the analysis equations may be computed efficiently using a low-rank approximation of the forecast covariance matrix, $\mathbf{P}_L^b$. This low-rank approximation of $\mathbf{P}_L^b$ is not invertible, but the minimization problem given in Equation (3.30) has a $(\mathbf{P}^b)^{-1}$ factor. To ensure that the minimization problem is well defined, (Ott *et al.*, 2002, 2004) compute a reduced-rank analysis in the space S spanned by the background ensemble perturbations, $\mathbf{X}_L^b$, using the eigenvectors of $\mathbf{P}_L^b$ as an orthonormal basis of S. Thus they write $\mathbf{P}_L^b = \mathbf{Q}\tilde{\mathbf{P}}_L^b\mathbf{Q}^{\mathrm{T}}$ where $\tilde{\mathbf{P}}_L^b$ is a $(k-1)\times(k-1)$ diagonal matrix of the non-zero eigenvalues of $\mathbf{P}_L^b$, and $\mathbf{Q}$ is an $n \times (k-1)$ orthogonal matrix of the corresponding eigenvectors. The matrix

$\tilde{\mathbf{P}}_L^b$ represents the background covariance matrix in the chosen orthogonal coordinate system of S and makes the $(\mathbf{P}_L^b)^{-1}$ term in Equation (3.11) well-defined.

A similar approach is taken in the LETKF, but with the coordinate system constructed with $\mathbf{Q} = \mathbf{X}_L^b$, such that $\mathbf{P}^b = \mathbf{X}_L^b \tilde{\mathbf{P}}_L^b (\mathbf{X}_L^b)^{\mathrm{T}}$ and

$$\tilde{\mathbf{P}}_L^b = (k-1)^{-1}\mathbf{I}_k, \tag{3.26}$$

where $\mathbf{I}_k$ is the k-dimensional identity matrix. Thus the LETKF seeks to minimize an objective function similar to Equation (3.11), but in the column space, S, of $\mathbf{X}_L^b$. This same decomposition of $\mathbf{P}_L^b$ is used in the ETKF (Bishop *et al.*, 2001).

Although $\mathbf{X}_L^b$ spans the space S, its columns are necessarily linearly dependent by construction, so they do not form a basis for S. However, as a real symmetric matrix, $\mathbf{P}_L^b$ is one-to-one in its column space, S, which is also the column space of $\mathbf{X}_L^b$. Thus $\mathbf{X}_L^b$ is a linear transformation from a k-dimensional space $\tilde{S}$ to the $(k-1)$-dimensional space S, relative to which $\mathbf{P}_L^b$ has a well defined inverse.

Consider a Gaussian random vector $\mathbf{w} \in \tilde{S}$ with mean $\mathbf{0}$ and covariance matrix $\tilde{\mathbf{P}}_L^b = (k-1)^{-1}\mathbf{I}_k$. Then $\mathbf{x}_L = \bar{\mathbf{x}}_L^b + \mathbf{X}_L^b\mathbf{w}$ is also a Gaussian random vector with mean $\bar{\mathbf{x}}^b$ and covariance matrix $\mathbf{P}_L^b$ (Hunt *et al.*, 2007). This motivates a cost function analogous to Equation (3.30) in the space $\tilde{S}$:

$$J^*(\mathbf{w}) = (k-1)^{-1}\mathbf{w}^{\mathrm{T}}\mathbf{w} + [\mathbf{y}^o - \mathbf{H}(\bar{\mathbf{x}}_L^b + \mathbf{X}_L^b\mathbf{w})]^{\mathrm{T}}\mathbf{R}^{-1}[\mathbf{y}^o - \mathbf{H}(\bar{\mathbf{x}}_L^b + \mathbf{X}_L^b\mathbf{w})] \tag{3.27}$$

Denote the ℓ-vector of observations within the local window as $\mathbf{y}_L^o$, and its associated $\ell \times \ell$ covariance matrix as $\mathbf{R}_L$. The background observation predictions for the jth local forecast are denoted by the ℓ-vector

$$\mathbf{y}_L^{b(j)} = \mathbf{H_L}\left(\mathbf{u}^{b(j)}\right), \tag{3.28}$$

where $\mathbf{H}_L$ is the local forward operator that relates model state quantities to the local observations. The operator defined by $\mathbf{H}_L$ may be a linear interpolation, or it may be a non-linear function whose linearization is assumed to provide a good approximation of $\mathbf{H}_L$ within the local window. The linearization is constructed by computing the $\ell \times 1$ ensemble mean of the predicted observations, $\bar{\mathbf{y}}_L^b = k^{-1} \sum_{i=1}^{k} \mathbf{y}_L^{b(i)}$, and the $\ell \times k$ observation perturbation matrix, $\mathbf{Y}_L^b$, whose jth column is given by $\mathbf{Y}_L^{b(j)} = \mathbf{y}_L^{b(j)} - \bar{\mathbf{y}}_L^b$. so that

$$\mathbf{H}(\bar{\mathbf{x}}_L^b + \mathbf{X}_L^b \mathbf{w}) = \bar{\mathbf{y}}_L^b + \mathbf{Y}_L^b \mathbf{w}, \tag{3.29}$$

Equation (3.27) becomes

$$\mathbf{J}^*(\mathbf{w}) = (k-1)^{-1} \mathbf{w}^{\mathrm{T}} \mathbf{w} + [\mathbf{y}_L^o - \bar{\mathbf{y}}_L^b - \mathbf{Y}_L^b \mathbf{w}]^{\mathrm{T}} \mathbf{R}_L^{-1} [\mathbf{y}_L^o - \bar{\mathbf{y}}_L^b - \mathbf{Y}_L^b \mathbf{w}], \tag{3.30}$$

which has the the same form as the original cost function $\mathbf{J}$ given by Equation 3.11, but with $\tilde{\mathbf{P}}_L^b = (k-1)^{-1}\mathbf{I}$, $\bar{\mathbf{w}} = \mathbf{0}$, and $\mathbf{Y}_L^b$ replacing the linear observation operator, $\mathbf{H}$. The analogue of the analysis Equations (3.17) and (3.18) in the space $\tilde{S}$ are given by

$$\tilde{\mathbf{P}}_L^a = \left[(k-1)\mathbf{I}_n - (\mathbf{Y}_L^b)^{\mathrm{T}} \mathbf{R}_L^{-1} \mathbf{Y}_L^b \right]^{-1} \tag{3.31}$$

$$\bar{\mathbf{w}}_L^a = \tilde{\mathbf{P}}_L^a (\mathbf{Y}_L^b)^{\mathrm{T}} \mathbf{R}_L^{-1} \left(\mathbf{y}_L^o - \mathbf{y}_L^b \right) \tag{3.32}$$

In particular, if $\bar{\mathbf{w}}_L^a$ minimizes $\mathbf{J}^*$, then $\mathbf{x}_L = \bar{\mathbf{x}}_L^b + \mathbf{X}_L^b \mathbf{w}_L^a$ minimizes the original cost function $\mathbf{J}$ in model space given by Equation (3.11) (Hunt *et al.*, 2007).

The ensemble mean of the analyzed local state at the grid point L is an adjustment of the background ensemble mean, consisting of a linear combination of local forecast perturbations given by $\bar{\mathbf{w}}_L^a$. The analysis in model state space is given by

$$\mathbf{P}_L^a = \mathbf{X}_L^b \tilde{\mathbf{P}}_L^a (\mathbf{X}_L^b)^{\mathrm{T}} \tag{3.33}$$

$$\overline{\mathbf{x}}_L^a = \overline{\mathbf{x}}_L^b + \mathbf{X}_L^b . \bar{\mathbf{w}}_L^a \tag{3.34}$$

The analysis calculations above are computationally efficient since they scale with the number of observations and ensemble size, and not the state size. The corresponding model state and covariance matrix in S can then easily be found with Equations (3.33) and (3.34) respectively.

3.4.2 Analysis Ensemble Construction

The final stage of the analysis computation at the fixed grid point L is the construction of an ensemble, $\{\mathbf{x}_L^{a(j)}\}_{j=1}^k$, whose perturbation matrix, $\mathbf{X}_L^a$, has an ensemble mean given by $\bar{\mathbf{x}}_L^a$ and whose covariance satisfies $\mathbf{P}_L^a = (k-1)^{-1}\mathbf{X}_L^a(\mathbf{X}_L^a)^{\mathrm{T}}$. This construction is accomplished by choosing $\mathbf{X}_L^a$ such that

$$\mathbf{X}_L^a = \mathbf{X}_L^b \mathbf{W}_L^a, \tag{3.35}$$

where $\mathbf{W}_L^a$ is the matrix square root of $(k-1)\tilde{\mathbf{P}}_L^a$, or in other words, $\mathbf{W}_L^a(\mathbf{W}_L^a)^{\mathrm{T}} = (k-1)\tilde{\mathbf{P}}_L^a$. With this choice of analysis ensemble, the analyzed covariance matrix is constructed as needed:

$$\begin{aligned}
(k-1)^{-1}\mathbf{X}_L^a(\mathbf{X}_L^a)^{\mathrm{T}} &= (k-1)^{-1}(\mathbf{X}_L^b\mathbf{W}_L^a)(\mathbf{X}_L^b\mathbf{W}_L^a)^{\mathrm{T}} \\
&= (k-1)^{-1}\mathbf{X}_L^b\mathbf{W}_L^a(\mathbf{W}_L^a)^{\mathrm{T}}(\mathbf{X}_L^b)^{\mathrm{T}} \\
&= \mathbf{X}_L^b\tilde{\mathbf{P}}_L^a(\mathbf{X}_L^b)^{\mathrm{T}} \\
&= \mathbf{P}_L^a
\end{aligned}$$

Also, with this construction, the columns of $\mathbf{X}_L^a$ add to zero (Hunt *et al.*, 2007), so the correct ensemble mean can be obtained by adding $\bar{\mathbf{x}}_L^a$ to each of its columns. The choice of symmetric square root, $\mathbf{W}_L^a$ as the symmetric square root for the construction of $\mathbf{X}_L^a$ is important due to the localized approach in the LETKF. This choice of $\mathbf{W}_L^a$ ensures that each local analysis covariance matrix, $\mathbf{P}_L^a$, varies continuously among the grid points and it also ensures that the resulting analysis ensemble of

perturbations is the closest to the background ensemble of perturbations, so that the resulting global analysis is smooth, provided that there is enough overlap of observations from grid point to grid point. See (Ott *et al.*, 2004) and (Hunt *et al.*, 2007) for further discussion on this choice of matrix square root.

The analyzed global ensemble consists of the collection of analyzed local ensembles, $\{\mathbf{x}_L^{a(j)}\}_{j=1}^k$, at each model grid point L. All of the above steps can be performed independently at every grid point, which allows the LETKF to be implemented efficiently on parallel computers. The LETKF has only three adjustable parameters: the ensemble size k and the localization radius associated with each grid point L. The third adjustable parameter is the covariance inflation factor ρ, which can vary with space and time, as described in Section 3.4.4.

3.4.3 Unobserved Variables and Parameter Estimation

For each grid point, L, the LETKF formulation provides a low-rank estimate of the local model covariance, $\mathbf{P}_L^b$, which approximates the cross-correlations among all state variables and parameters within the associated local window. This enables the estimation of unobserved state variables and model parameters with the state augmentation approach, as described below.

Denote the parameter being estimated as $\mathbf{p}$ and for simplicity assume it is a global parameter, such that it is physically correlated to the observed quantities at all grid points in the domain. Associated with the ensemble of global background vectors,$\{\mathbf{u}^{b(j)}\}_{j=1}^k$, is an ensemble of background parameter estimates,$\{\mathbf{p}^{b(j)}\}_{j=1}^k$. Consider a fixed grid point L and the ensemble of local background state vectors within its associated local window, $\{\mathbf{x}_L^{b(j)}\}_{j=1}^k$, as described in Section 3.4. The ensemble

of local state vectors is augmented with the ensemble of parameters $\{\mathbf{p}^{b(j)}\}_{j=1}^{k}$, such that the jth local state vector is given by

$$\mathbf{z}_L^{b(j)} = \left[\left(\mathbf{x}_L^{b(j)} \right)^T \left(\mathbf{p}^{b(j)} \right)^T \right]. \tag{3.36}$$

The analyzed local state vector, $\mathbf{z}_L^{a(j)}$, is updated as in Equation (3.34), and contains an ensemble of analyzed parameter estimates at the grid point L, which is denoted as $\{\mathbf{p}_L^{a(j)}\}_{j=1}^{k}$. Repeating this process at each grid point yields an ensemble of local analyzed parameter estimates at each grid point. Each value of $\{\mathbf{p}_L^{a(j)}\}_{j=1}^{k}$, is calculated at each grid point, L, using observations located within their respective local windows, so these quantities vary in space and time.

The ensemble of analyzed parameters, $\{\mathbf{p}^{a(j)}\}_{j=1}^{k}$, that is used in the propagation of the forecast is typically obtained through a suitable choice of spatial average among the ensemble of analyzed parameters at each grid point. In the case of non-global parameters, a spatial average over the grid points within a specific region is used. Multiple model parameters may be estimated in a similar manner, by augmenting their values to each of the local states. An overview of parameter estimation strategies in data assimilation is presented in (Ruiz *et al.*, 2013) and additional parameter estimation strategies with Ensemble Kalman filters for non-global parameters are discussed in (Bellsky *et al.*, 2014).

3.4.4 Covariance Inflation

As discussed in Section 3.3.1, localization compensates for sampling errors to a degree, but the ensemble may still misrepresent model covariance locally in some regions due to the non-linearities of the model. After multiple analysis cycles, the ensemble covariance may underestimate the actual model uncertainty. This puts the data assimilation system at risk of filter divergence, where the ensemble covariance

inadequately underestimates the model uncertainty and the influence of observations is reduced greatly.

Approaches to mitigate these effects typically involve the artificial inflation of the forecast covariance matrix. Covariance inflation is typically applied by manually tuning it, which may be unfeasible with complex models. (Hunt *et al.*, 2007) discusses more details about covariance inflation in the LETKF when applied to the Global Forecast System model. Adaptive covariance inflation schemes have been proposed for this and are discussed in (Anderson, 2009) and (Li *et al.*, 2009). The adaptive covariance inflation scheme derived by (Miyoshi, 2011), which is designed to work specifically for the LETKF, is used in this study. A brief description of its implementation is provided below and the reader is referred to (Miyoshi, 2011) for a thorough description.

The covariance inflation is computed for each grid point L along with each local analysis ensemble, yielding spatially and temporally varying analysis covariance inflation factors. Each inflation factor is computed using the update

$$\alpha_L^a = \frac{\alpha_L^b v_L^o + \alpha_L^o v_L^b}{v_L^o + v_L^b}, \tag{3.37}$$

where α_L^b and α_L^o are the current and observed inflation factors, respectively. Their respective variances are v_L^b, which is a fixed parameter used to control temporal variations of adaptive covariance inflation values, and v_L^o, which is updated at each analysis time based on new observations. The factors α_L^o and v_L^o are computed under the assumption that the local innovation, $\mathbf{d}_L = \mathbf{y}_L^o - \bar{\mathbf{y}}_L^b$, has expected value $\langle \mathbf{d}\mathbf{d}^{\mathrm{T}} \rangle = \mathbf{H}_L \mathbf{P}_L^b \mathbf{H}_L + \mathbf{R}_L$. The covariance inflation scheme tries to find an inflation factor α_L^o such that $\alpha_L^o \mathbf{P}_L^b$ approximately satisfies the innovation relation. These considerations yield the local "observation covariance" update

$$\alpha_L^o = \frac{\mathrm{tr}\left(\mathbf{d}_L \mathbf{d}_L^{\mathrm{T}} \circ \mathbf{R}_L^{-1}\right) - \ell}{\mathrm{tr}\left(\mathbf{H}_L \mathbf{P}_L^b \mathbf{H}_L^{\mathrm{T}} \circ \mathbf{R}_L^{-1}\right)}, \tag{3.38}$$

where ℓ is the number of local observations, tr denotes the trace, and $\circ$ is the Hadamard (elementwise) product. The variance v_L^o of α_L^o is updated as

$$v_L^o = \frac{2}{\ell}\left(\frac{\alpha_L^b \,\mathrm{tr}(\mathbf{H}_L\mathbf{P}_L^b\mathbf{H}_L \circ \mathbf{R}_L^{-1}) + \ell}{\mathrm{tr}(\mathbf{H}_L\mathbf{P}_L^b\mathbf{H}_L^{\mathrm{T}} \circ \mathbf{R}_L^{-1})}\right). \tag{3.39}$$

In the LETKF, the analyzed covariance inflation factor for a grid point, α_L^a, is computed prior to the analysis calculation at L and is efficiently incorporated in the calculation by modifying $\tilde{\mathbf{P}}^a$ in Equation (3.32) by replacing the $(k-1)\mathbf{I}_k$ term with $(k-1)\mathbf{I}_k/\alpha_L^a$.

Chapter 4

OBSERVING SYSTEM EXPERIMENTS WITH AN ELECTRODYNAMICS MODEL

The work in this chapter is based on (Durazo *et al.*, 2016), which is published in the journal *Physica Scripta*.

4.1 Introduction

This chapter presents numerical experiments to test the skill of the Local Ensemble Transform Kalman Filter (LETKF) (Hunt *et al.*, 2007) for assimilating synthetic ionospheric observations in cases where the domain is sparsely sampled. In particular, the LETKF is applied to a mesoscale model that incorporates the effect of neutral turbulence as a driving force, whose timescales are fast compared to those in a global weather forecast model.

The ionospheric model used here is a regional idealized model, called iDiablo, that extends from the E region to part of the F region (approximately 80 km to 440 km in altitude). It is built from a three-dimensional direct numerical solver for the Navier-Stokes equations of neutral flow (Taylor, 2008) and simulates the time evolution of an ion density field as it interacts with neutral winds from an idealized gravity wave and electrostatic forcing from the movement of charged ions. In particular, this model focuses on the interaction between neutral and charged flows due to collision forces. It should the noted that the real ionosphere is significantly influenced by external solar conditions and is thus compliance-driven to some extent. Ionospheric models typically account for solar conditions parametrically, and ensemble Kalman filters, including the LETKF, can incorporate changes in these parameters through state-

augmentation techniques, as described in Section 3.4.3. However, iDiablo does not include dynamics resulting from changes in solar conditions. Thus this dynamical system, which includes a local ionosphere region and its interaction with the propagating gravity waves, is more persistence driven than to compliance driven.

At lower altitudes (80 km to 130 km), ion- and electron-neutral collisions can be significant due to the large density of neutral particles, so the specification of neutral winds within this region is important. The ion density structure at these lower altitudes has been difficult to model due to the high number of collisions and resulting complex coupled electrostatic fields. In addition, the altitude of this region is too low for orbiters and too high for radiosondes to make direct measurements, so observations in this region are especially sparse. Although the use of satellite-derived observations with the radio occultation technique, such as those provided by the COSMIC satellites, provides near global coverage, these observations have been shown to have significant uncertainties at these altitudes (Yue *et al.*, 2010; Liu *et al.*, 2010). At higher altitudes (above 200 km), the ion- and electron-neutral collision processes become much weaker as the density of neutral particles decays with altitude. The iDiablo model was chosen for this investigation because it focuses on the interaction of an ion density field with neutral winds, and the neutral winds can be modified easily to study their effects on the ion density field.

The goal in this study is to to forecast mesoscale structures of a regional ion density field as it interacts with the driving force of neutral winds. These structures are large enough to be of practical interest and small enough to be resolved in global ionospheric models. Instead of deducing the neutral winds from observations of charged content, the time evolution of the neutral winds is specified within an iDiablo model realization and is used to generate synthetic observations. These synthetic observations are then assimilated into a set of forecast ensembles at fixed time intervals

using the LETKF scheme to obtain an estimate of the time evolution of the neutral winds. The estimates are compared with the original model realization to assess the accuracy of the forecast. In other words, iDiablo is assumed to perfectly represent the ionospheric dynamics; errors arising in forecasts restarted from the analyses reflect the performance of the data assimilation system. Of course, an actual forecast system for the Earth's ionosphere must cope with errors in the model as well as in the data, but this approach lets us separate the two issues for this initial investigation.

Section 4.2 describes the iDiablo model, Section 4.3 outlines the initialization of the forecast ensembles and the data assimilation parameters that are varied. Section 4.4 describes the numerical results, and conclusions are presented in Section 4.5.

4.2 Ionospheric Model: iDiablo

4.2.1 Overview

The ionospheric model used in this paper, which is called iDiablo, is a modification of a three-dimensional direct numerical simulation package that is explained in more detail in (Tang and Mahalov, 2013). This model simulates the time evolution of a three-dimensional ion density and its electrostatic potential field on a $44 \times 44 \times 361$ grid as it is perturbed by neutral winds from an idealized gravity wave. The three physical directions in the model are the zonal (X), meridional (Z), and vertical (Y) directions. Physically, this domain spans 80 km to 440 km altitudes in the vertical and a 44 km $\times$ 44 km region in the horizontal direction, so it captures the E region and the part of the F region. The initial condition for the ion density field is obtained from the IRI2007 model, on 25 June 1991 at 20:00 LT, at 35° N, 136° W in geographic latitude and longitudes (Bilitza and Reinisch, 2007).

The ion density field is uniform at first and its vertical profile is shown in Fig-

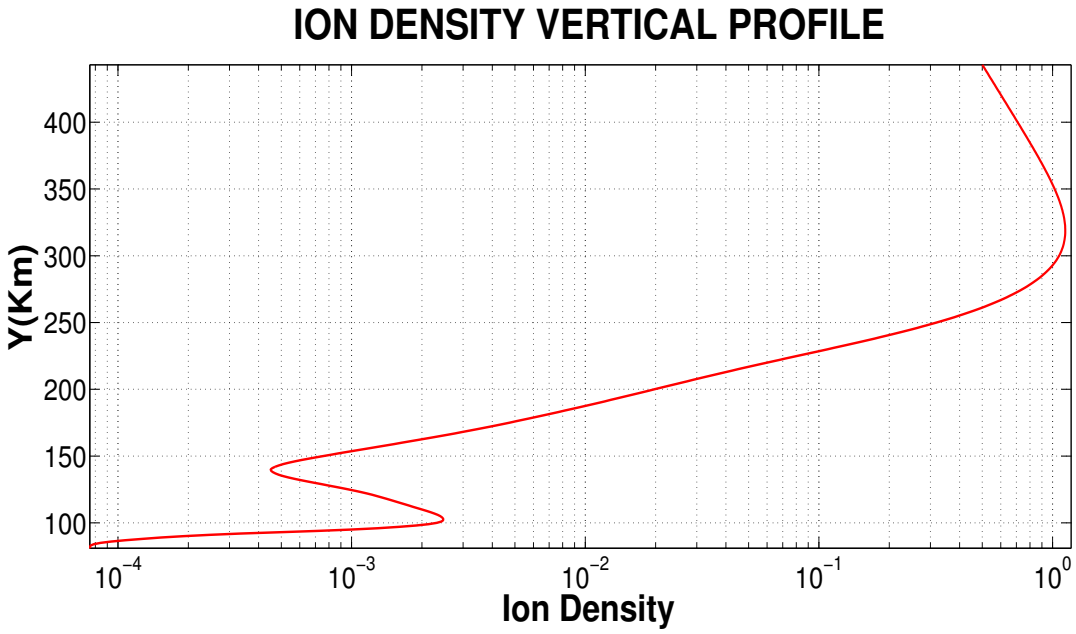

Figure 4.1: Mean ion density field profile at the beginning of the simulation. The vertical axis is altitude in kilometers and the horizontal axis is the ion density scaled by $10^6\,\mathrm{cm}^{-3}$.

ure 4.1. There is a local maximum in the ion density in the E region at around 105 km altitude and another in the F region at around 335 km altitude. The E-region density peak experiences significant forcing from the gravity wave via ion- and electron-neutral particle collisions. However, this process weakens as the altitude increases, especially at the ion density peak in the F-region. Figure 4.2 shows that the gravity wave has a damped sinusoidal magnitude in each of its three components as altitude increases, reflecting this collision process. The gravity wave has a period of 15 minutes and a speed of $15\,\mathrm{m\,s}^{-1}$.

The movement of charged particles by the gravity wave creates an electric field, captured by its electrostatic potential, which is shown in Figure 4.3 for the lower altitudes of the domain (80 km to 150 km). The greatest magnitudes of the electrostatic potential correspond to the strongest neutral winds, which are located in the lower part of the domain. As the altitude increases, the gravity wave winds dampen and

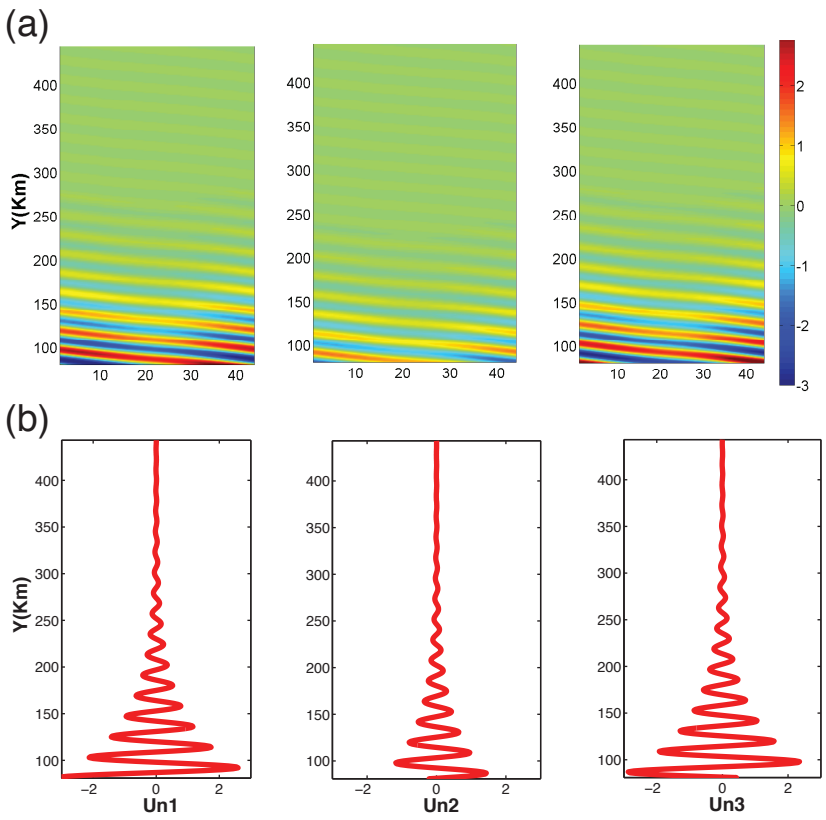

Figure 4.2: (a) Mean Y-Z plane and (b) mean vertical profile of each component of the gravity wave used as the forcing in the truth run. For each component, the vertical axis is altitude and the horizontal axis is the position in the Z direction in kilometers. The wind speed is shown in scaled units ($1 = 15$ m s^{-1}).

the electromagnetic forcing is reduced accordingly. Together with the wind forcing from the gravity wave, the electrostatic potential contributes to the structure of the resulting ion density field as time evolves.

The ion density field has periodic boundary conditions in the horizontal directions. The boundary condition at the top of the domain is set so that diffusive equilibrium is maintained and the bottom has zero-flux boundary conditions. Spectral methods are used to compute derivatives in the horizontal directions, and finite differences are used for derivatives in the vertical direction. The time evolution of the ion density field is

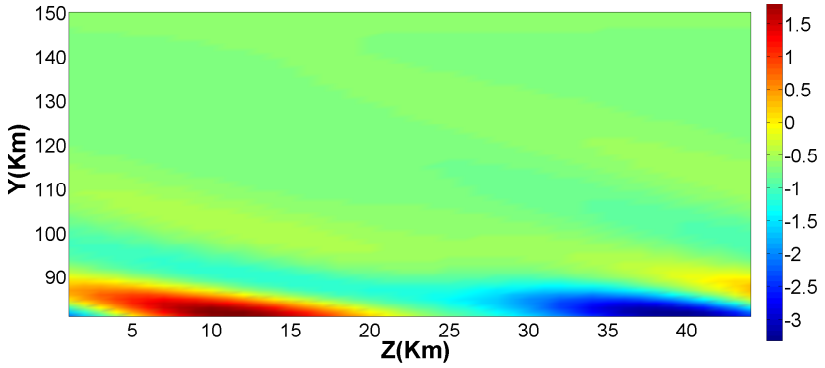

Figure 4.3: Mean Y-Z plane of the electrostatic potential field for the altitudes of interest. The vertical axis is the altitude in kilometers and the horizontal axis is the horizontal position in kilometers. The color coding shows the potential in scaled units ($1 = 5.5876 \times 10^3$ volts).

done through an explicit Runge-Kutta-Wray method. The time step corresponds to about 1% of the gravity wave period.

4.2.2 Model Dynamics

The "perfect model" assumption is made here so that a "truth run" of the model can be compared with ensemble forecasts to assess the skill of the data assimilation system. The "truth run" starts with a pre-specified initial condition that includes the ion density field and the neutral winds. The dynamics in iDiablo are driven by the gravity wave, and the ion density field does not feed back on the neutral dynamics. Thus, to predict the structure of the ion density field, the gravity wave forcing must be specified accurately. Synthetic observations of one component of the neutral winds are generated from the truth run and are assimilated at intervals corresponding to about half the gravity-wave period.

4.3 Ensemble Initialization and Data Generation

Because the observational coverage of the ionosphere is sparse, the main goal in the numerical experiments described below is to study how the accuracy of ensemble analyses of the ion density field varies with the density of assimilated neutral wind observations. Each component of the gravity wave used in these numerical experiments is uniform in the X (zonal) direction, sinusoidal in the Z (meridional) direction, and damped sinusoidal in the Y (vertical) direction. Each component consists of the sum of two wave modes that may have different amplitudes, as follows:

$$U_{n_1} = A_1 \cos\left(Z + Y + H_1 + \frac{t}{p + H_2}\right) + A_2 \cos\left(\lambda Z + Y + H_3 + \frac{t}{p + H_4}\right) \quad (4.1)$$

$$U_{n_2} = -A_3 \cos\left(Z + Y + H_1 + \frac{t}{p + H_2}\right) + A_4 \cos\left(\lambda Z + Y + H_3 + \frac{t}{p + H_4}\right) \quad (4.2)$$

$$U_{n_3} = A_5 \sin\left(Z + Y + H_1 + \frac{t}{p + H_2}\right) + A_6 \sin\left(\lambda Z + Y + H_3 + \frac{t}{p + H_4}\right) \quad (4.3)$$

where H_1, H_2, H_3, and H_4 are perturbation terms, t is time, p is the wave period, and each A_i is a wave amplitude.

In the truth run, the gravity wave is produced by Equation (4.1) with $H_1 = H_2 = H_3 = H_4 = 0$. Synthetic observations are generated from the truth run at selected model grid points by adding a Gaussian random noise with zero mean. The noise variance is scaled to be 10% of the magnitude of the truth run's neutral winds at the grid point where the observation is generated (larger variances are considered the third set of the numerical experiments below). Although all three components of the gravity wave are analyzed, the synthetic observations are only of the U_{n_1} component.

Ensemble forecasting schemes start from sets of initial conditions that are assumed to be statistically equivalent. In the numerical experiments below, an initial condition for each ensemble solution is obtained by choosing values of H_1, H_2, H_3, and H_4 from a Gaussian distribution with mean zero and variance 0.5; the same values of H_i are applied to each component of the gravity wave. The H_1 and H_3 terms perturb the

phase of each of the gravity wave mode and the H_2 and H_4 terms perturb the period. The amplitude of each mode also is perturbed by adding a random Gaussian random variable with zero mean and variance equal to 20% of the amplitude. A set of 40 initial conditions is generated in this way. The period of the gravity wave is about 15 minutes; observations are assimilated every half-wave period, and the forecast-analysis cycle lasts for 8 wave periods (about 2 hours).

All ensemble Kalman filters require a sufficiently large forecast ensemble to adequately sample the forecast covariance. The LETKF presumes that the dynamics are locally low dimensional, that is, the short-term forecast error lies in a relatively low-dimensional subspace compared to the size of the space spanned by all the components of the model state vector within a given local region. This behavior is characteristic of numerical weather models (Patil *et al.*, 2001), and 40 ensemble members provided good results when the LETKF was applied to the Global Forecast System model (Szunyogh *et al.*, 2005). However, the minimum ensemble size required for the forecast-analysis system described here is not quantified in this study.

4.4 Numerical Experiments

In the numerical experiments described below, a set of 40 ensemble initial conditions is generated and the forecast-analysis cycle run as described in Section 4.3 with observations at a randomly chosen set of 25%, 10%, 5%, 2% and 1% of the model grid points. The performance of the LETKF is then assessed with respect to changes in the size of the local regions and the amplitude of the observational noise. The final experiment considers a situation in which all observations lie in vertical columns at randomly chosen horizontal locations.

4.4.1 Measures of Accuracy

The principal measure of analysis accuracy used is the root-mean-square (RMS) difference between the "truth run," $T_{i,k,j}(t_n)$ and the analysis ensemble mean, $A_{i,k,j}(t_n)$, in various regions of the three-dimensional model grid; i, k, j index the grid points in the zonal, meridional, and vertical directions, respectively, and t_n is the analysis time. There are $N_x = N_z = 44$ grid points in the zonal and meridional directions and $N_y = 361$ vertical levels; each simulation consists of $M = 16$ assimilation cycles.

The *overall RMS* (ORMS) error at each t_n is the RMS difference between the analysis mean and the truth over the entire model grid:

$$\mathrm{ORMS}(t_n) = \left[N_x^{-1} N_z^{-1} N_y^{-1} \sum_{i,k,j} \left(T_{i,k,j}(t_n) - A_{i,k,j}(t_n) \right)^2 \right]^{1/2} . \tag{4.4}$$

The *altitude RMS* (ARMS) error at the jth vertical level is the time average of the RMS difference between the analysis and truth over the X-Z plane:

$$\mathrm{ARMS}(j) = M^{-1} \sum_{n=1}^{M} \left[N_x^{-1} N_z^{-1} \sum_{i,k} \left(T_{i,k,j}(t_n) - A_{i,k,j}(t_n) \right)^2 \right]^{1/2} . \tag{4.5}$$

The *plane RMS* (PRMS) error is also defined to explore the error structures that form as time advances. It is defined for each vertical and meridional level in the model grid as the time-averaged value of the RMS difference between the analysis mean and the truth:

$$\mathrm{PRMS}(k, j) = M^{-1} \sum_{n=1}^{M} \left[N_x^{-1} \sum_{i} \left(T_{i,k,j}(t_n) - A_{i,k,j}(t_n) \right)^2 \right]^{1/2} . \tag{4.6}$$

4.4.2 First Set of Numerical Experiments

All the numerical experiments described in this paper are variations on this initial one, which uses synthetic observations of one component of the neutral wind as described in Sec. 4.3. At each grid point, observations are assimilated from a local

region consisting of 7 grid points in each of the horizontal directions and 15 grid points in the vertical, corresponding to 16% of the model grid extent in each horizontal direction and about 9% in the vertical. The observations are taken randomly selected subsets of 25%, 10%, 5%, 2% and 1% of the model grid points. A covariance inflation factor of 15% ($\rho = 1.15$ in Step 2 of the LETKF cookbook (Hunt *et al.*, 2007)) is used to compensate for the effect of nonlinearities and the underestimation of forecast covariance that is typical of ensemble Kalman filters.

Forecast Errors in the Neutral Winds

First, the time stability of the LETKF scheme in forecasting the neutral winds is examined. Figure 4.4 shows the ORMS error of each neutral wind component for all the cases of observational coverage. The vertical axis is scaled relative to the speed of the gravity wave, which is 15 $\mathrm{m\,s^{-1}}$; that is, an ORMS value of 1 corresponds to 15 $\mathrm{m\,s^{-1}}$. The results are plotted for every model time step; each period of the gravity wave is about 15 minutes, corresponding to 100 time steps; data are assimilated every 50 time steps.

Although only observations of the U_{n_1} component of the wind are assimilated, the ORMS value of each wind component is less than half of that obtained in the absence of data assimilation. The filter appears to remains stable for all choices of observational coverage considered here, although there is a slow upward drift in the RMS error of the vertical wind component, particularly for the sparsest (1% and 2%) levels of observational coverage. The ORMS error rises as the observation density is decreased, but the results for 25%, 10%, and 5% observational coverage are similar. These results suggest that in the case of randomly located observations, the minimal coverage required for good forecast accuracy is between 2% and 5% of the grid points. Similar results were found in perfect-model experiments of the LETKF

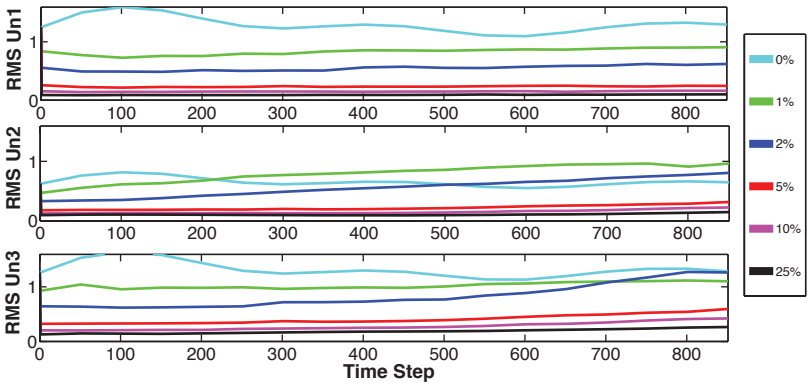

Figure 4.4: ORMS of the three components of the neutral wind field for forecast/analysis cycles of varying observational coverage. In each panel, the vertical axis is the ORMS value in scaled units ($1 = 15$ m s^{-1}) and the horizontal axis is the time step; the time interval corresponds to 8 periods of the gravity wave (about 2 hours).

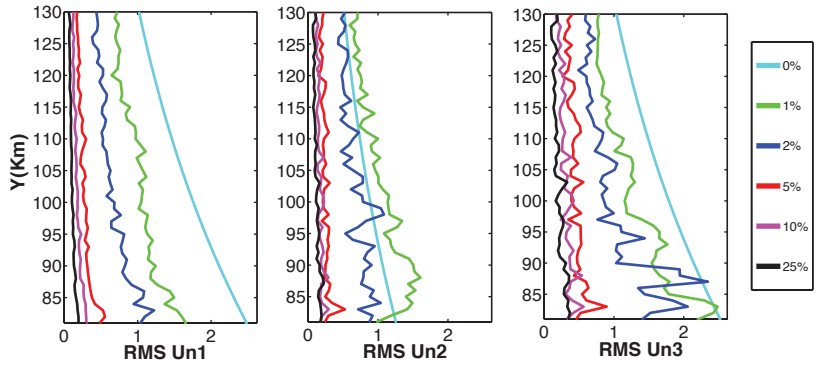

Figure 4.5: Altitude RMS (Equation 4.5) of the three components of the neutral winds as a function of observational coverage. In each panel, the vertical axis is the altitude in kilometers and the horizontal axis is the ARMS value in scaled units as for Figure 4.4.

with the Global Forecast System (GFS) model (Szunyogh *et al.*, 2005).

Figure 4.5 shows the altitude RMS (ARMS) errors, Equation 4.5, for the neutral winds. The errors are similar for observational coverage down to 5% of the model grid points. Even with observational density at 2% of the grid points, the ARMS error is still significantly smaller compared to the case with no data assimilation. However,

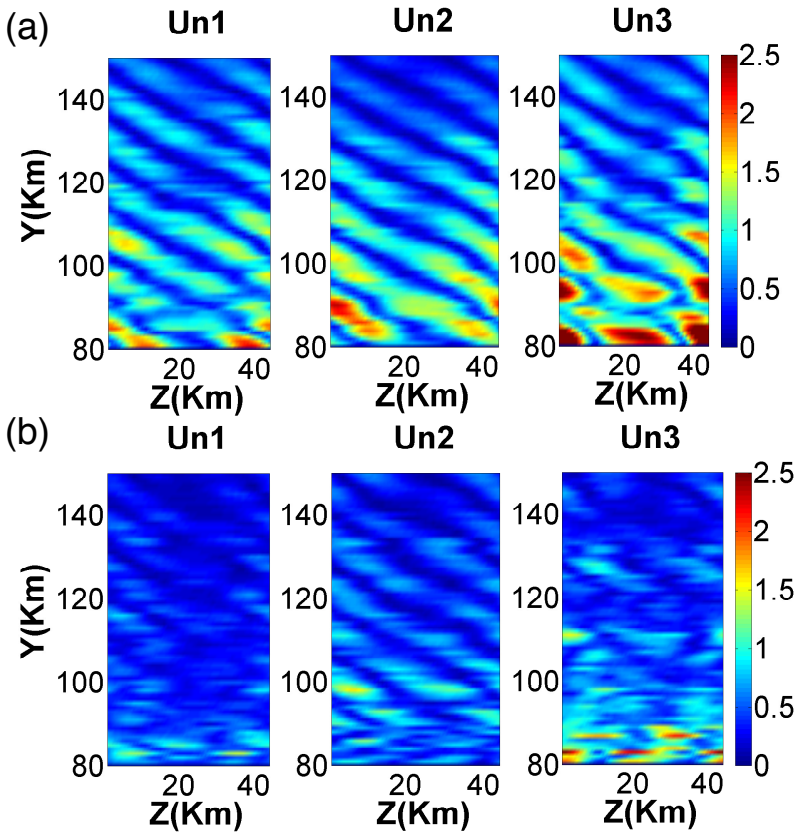

Figure 4.6: Planar RMS error, Equation 4.6, of the three components of the neutral wind at (a) 1% and (b) 2% observational density. The vertical axis is the altitude and the horizontal, the meridional direction in kilometers. The color bar represents magnitudes in scaled units, as in Figure 4.4.

there are noticeable local maxima as a function of height at the lowest observtional densities.

Figures 4.6(a) and 4.6(b) show the corresponding planar RMS (PRMS) error distribution, Equation 4.6, for the cases of 1% and 2% observational coverage, respectively. The error distributions reflect the damped sinusoidal magnitude of the gravity wave in all three components. The magnitude of the error is considerably larger at 1% coverage than at 2% coverage, and in both cases is largest in the vertical (U_{n_3}) component of the wind.

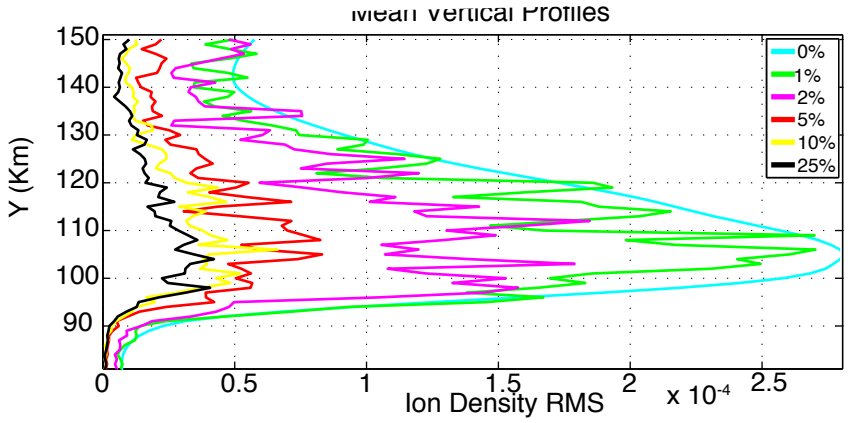

Figure 4.7: ARMS error of the vertical profile of the analyzed ion density field averaged over time. Each curve corresponds to a different observational density as indicated.

Errors in the Ion Density and Electrostatic Potential Fields

Next consider the electrostatic and ion density analyses. Figure 4.7 shows a more pronounced increase in the ARMS of the ion density, compared to the neutral winds, as the observational density decreases. However, as long as the observational density remains at 2% or more of the grid points, the RMS error of the ion density is less than the case where no observations are assimilated. The largest contribution to the analysis errors occurs near the E-region ion density peak. The neutral winds from the gravity wave still have significant magnitude at this altitude. These altitudes also are the regions where the ensemble spread is largest, as shown in Figure 4.8.

Figure 4.9 shows the PRMS errors for (a) the ion density field and (b) the electrostatic field. The errors in the ion density field follows those of the gravity wave winds in cases of low observational coverage. The gravity wave is strongest in magnitude near the bottom of the domain, and these error structures become smaller as the neutral winds get weaker at altitudes of 100–130 km. The errors in the electrostatic field are largest at the bottom of the domain and are approximately halved when the observational coverage is increased to 2% from 1%.

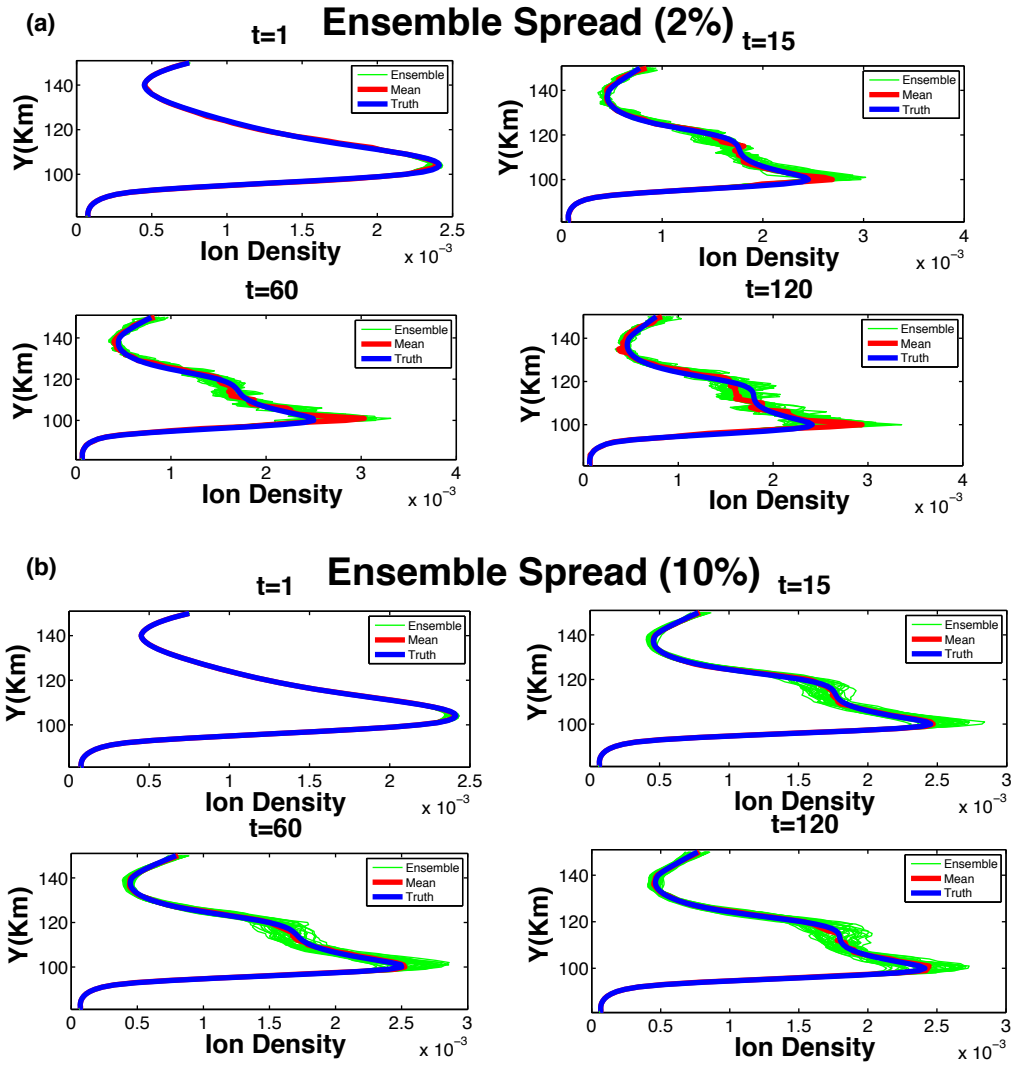

Figure 4.8: Vertical profiles of the analysis ion density field at (a) 2% and (b) 10% observational density. The blue and red curves show the ion density from the "truth run" and the analysis ensemble mean, respectively; green curves show the values for selected ensemble solutions. The plots are for 1, 15, 60 and 120 minutes (0, 1, 4, and 8 gravity-wave periods), respectively. In each plot, the vertical axis is the altitude in kilometers and the horizontal axis is the ion density.

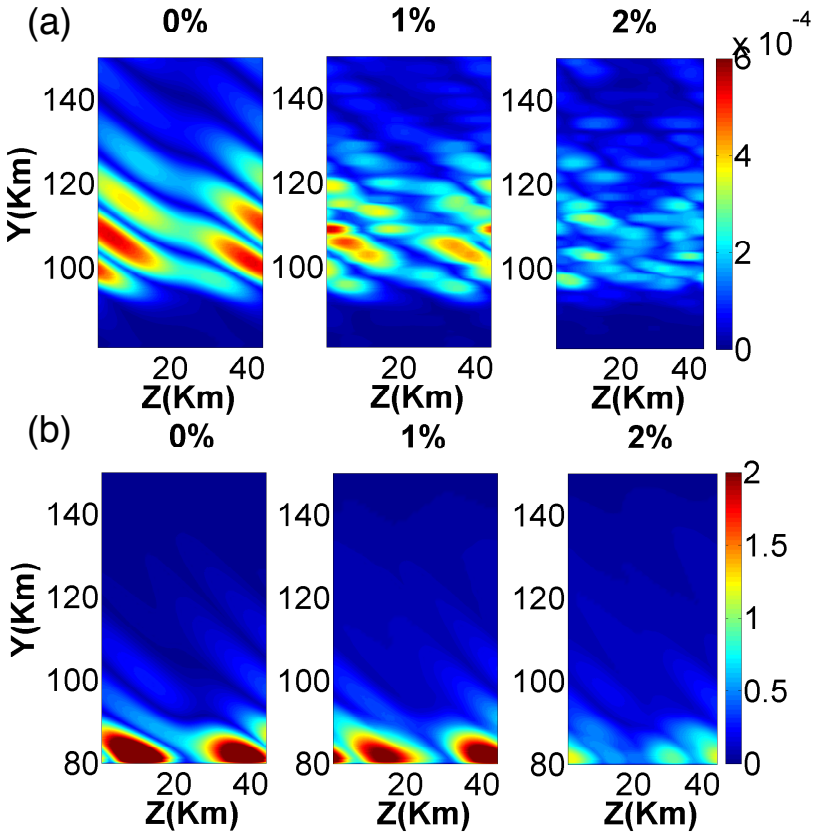

Figure 4.9: PRMS error of (a) the ion density field and (b) the electrostatic field in the case of no data assimilation (left panel) and data assimilation with observations located at a random choice of 1% and 2% of the model grid points, respectively. For each panel, the vertical and horizontal axes are altitude and horizontal location respectively in kilometers. The colorbar depicts the magnitude of the errors as indicated.

4.4.3 Set 2: The Effect of Local Region Size

In the remainder of this paper, the first set of numerical experiments is repeated for varying selected parameters of the LETKF data assimilation scheme. The remaining figures show the ion density ARMS and PRMS errors.

When the observational coverage becomes sufficiently sparse, the local regions surrounding many grid points lack any observations. This set of experiments assesses the effect of approximately doubling the extent of the local regions in each direction.

Specifically, 15 grid points are included in each horizontal direction (about 32% of the horizontal extent) and 53 grid points in the vertical direction (about 15% of the vertical extent). Although a larger local region allows more observations to be assimilated, it may also include greater model nonlinearities, the ensemble sample covariance may fail to adequately approximate the actual forecast uncertainty, and more distant observations may be less correlated with the dynamics near each analyzed grid point. In addition, for a fixed ensemble size, the computational complexity of the LETKF is linear in the number of observations at each grid point, so more observations increase the computational expense.

Despite these potential difficulties, an increase in the local region size has beneficial effects in the perfect-model simulations at sparser observational densities. Figure 4.10(a) shows the analyzed vertical profiles of ion density, whose accuracy is significantly improved compared to those in Figure 4.7. Figure 4.10(b) shows the same PRMS error profiles of the ion density as Figure 4.9(b). Even at observational densities as low as 1%, the local errors are substantially reduced compared to the initial choice of local region size.

4.4.4 Set 3: Larger Observational Noise

This set of numerical experiments is the same as the previous one, except that the noise in the synthetic observations is quadrupled to 40% of the magnitude of the neutral wind in the truth run at the grid points where observations are generated. Figure 4.11 displays the same quantities as Figure 4.10. The errors have a similar structure, but they are larger at each altitude compared to those in Set 2. These results suggest that the LETKF can provide good analyses for the iDiablo model when the data has as few as 2 significant bits of accuracy.

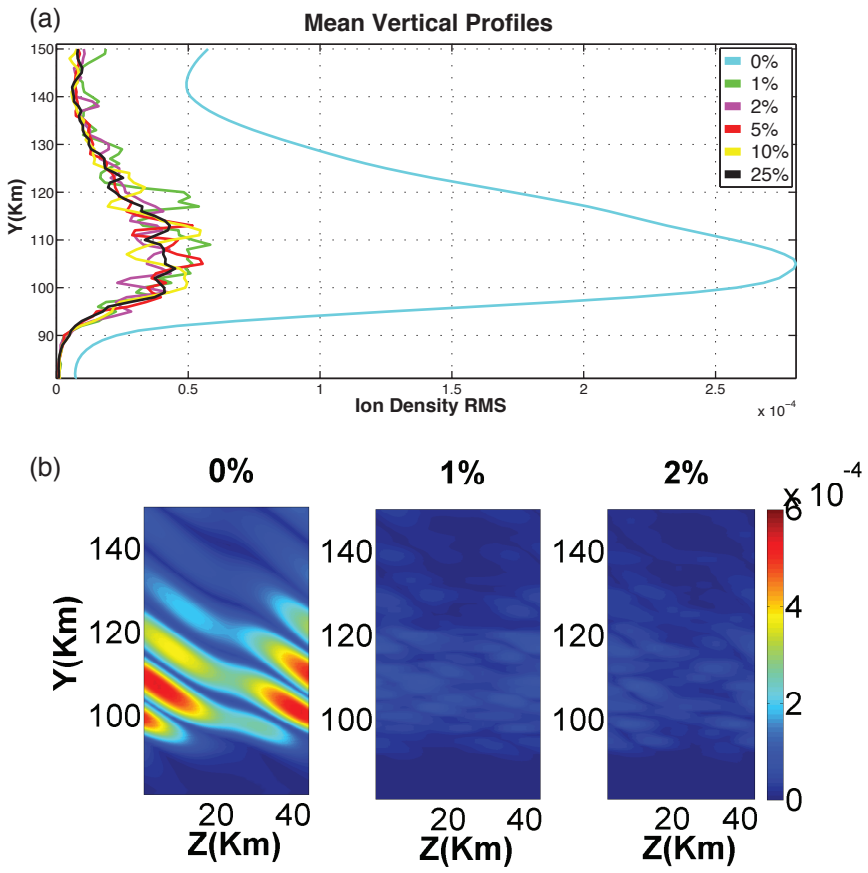

Figure 4.10: (a) Analysis mean of ion density as a function of altitude using larger local regions in the LETKF. (b) Corresponding PRMS plots of the ion density with no data assimilation and with the LETKF at 1% and 2% observational coverage.

4.4.5 Set 4: Structured Observational Network

Finally, an observing network in which measurements are taken throughout a randomly located set of vertical columns is considered. The observational noise levels and local region size are as in Set 2. There are 44^2 horizontal grid points, so the observing network at 1% coverage consists of 19 randomly selected horizontal locations at which synthetic observations at each vertical model level are generated. Figure 4.12 is analogous to Figure 4.11; (a) shows the RMS errors in ion density as a function of altitude, and (b) shows the structure of the RMS error averaged in the zonal direction

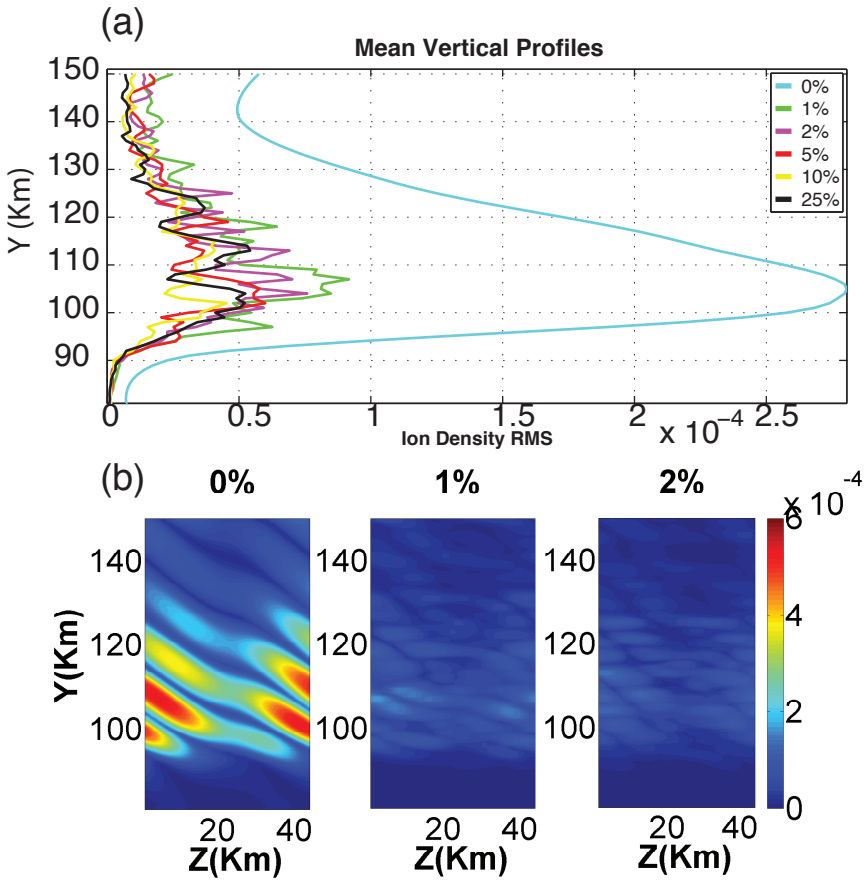

Figure 4.11: (a) Analysis mean of ion density as a function of altitude using observations with larger noise levels. (b) Corresponding PRMS plots of the ion density with no data assimilation and with the LETKF at 1% and 2% observational coverage.

at each vertical level. With a suitably large local region, the analyses at each grid point are reasonably accurate all the way to 1% observational coverage.

4.5 Discussion and Conclusions

In this paper, the efficacy of the Local Ensemble Transform Kalman Filter (LETKF) is tested on an idealized ionospheric model that includes the interaction of an ion density field and a gravity wave, along with the subsequent electromagnetic forcing caused by the movement of charged particles. Observations of one component of

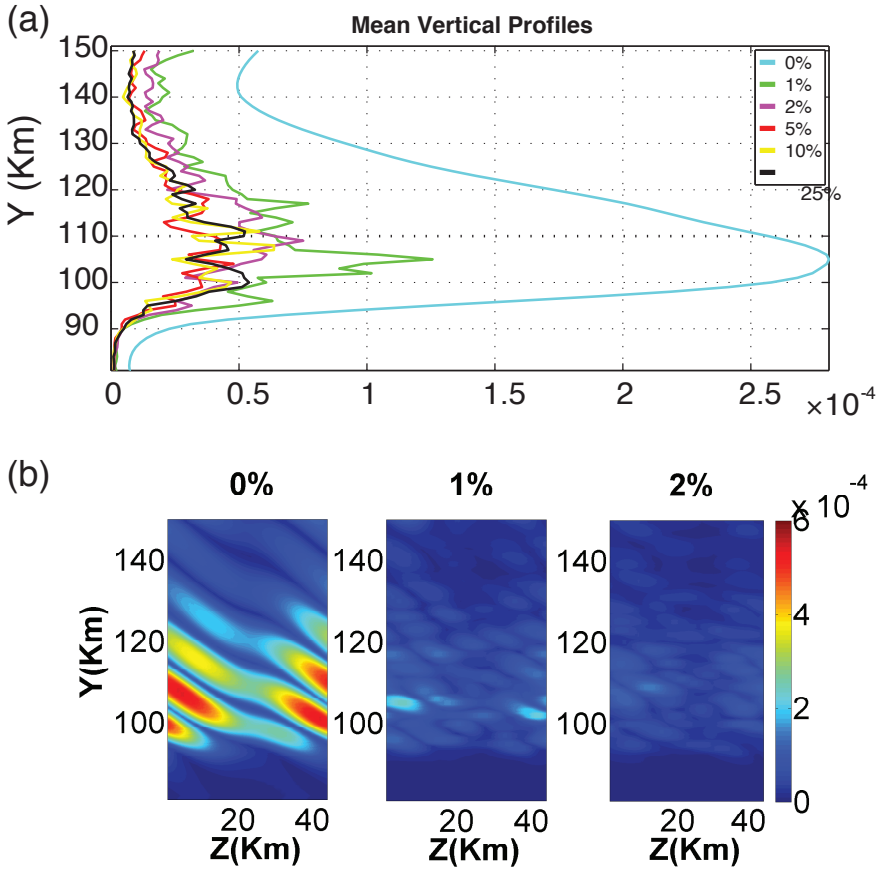

Figure 4.12: (a) Analysis mean of ion density as a function of altitude using observations throughout randomly located vertical columns in the model domain. (b) Corresponding PRMS plots of the ion density with no data assimilation and with the LETKF at 1% and 2% observational coverage.

the neutral wind are assimilated to compute updated initial conditions that include the ion density field. In these numerical experiments, a gravity wave with a linear structure is used as the forcing. Initial results suggest that the LETKF is effective when the observational density is sparse. The analyses of the wind and ion density fields are almost as good with 2% observational coverage as they are with many more observations, at least in this perfect-model scenario. Good results are also obtained when the noise levels in the synthetic data are quadrupled from their original values. Because the LETKF assimilates observations only in a local region about each model

grid point, its performance depends on the size of the local regions, particularly when the observational density is low.

One of the main attractions of the LETKF scheme is its relatively low computational cost. The model grid used here consists of three neutral wind components defined on 702,768 grid points. Although the LETKF is amenable to efficient implementation on parallel computers, it can also be implemented efficiently in MATLAB, which was the computing platform used for this initial investigation. In the first set of experiments, which used the smaller local regions, the LETKF required about 2 minutes per data assimilation step with observations at 5% of the grid points; the larger local regions required about 15 minutes per assimilation step. Significantly greater performance is possible in compiled implementations using multiple processors.

Like all ensemble Kalman filter schemes, the forecast model must be run from many initial conditions. This is usually the most expensive step; in this investigation, the iDiablo model was compiled and run on a high-performance computing cluster. The MATLAB program imported the model fields and implemented the LETKF to generate the updated initial conditions.

The results suggest that the LETKF can perform well with rather sparse observational networks, provided that the density is reasonably uniform and that the local regions are chosen sufficiently large. Future work will address questions of model error and more complex forcing. With the amount of observational coverage expected to increase over the next few years, data assimilation will play a larger role in ionospheric space weather forecasting. However, observational coverage is likely to remain sparse over many regions of the globe. For these reasons, the LETKF merits further consideration as a data assimilation system for space weather forecasting.

Chapter 5

LETKF FOR IONOSPHERIC DATA ASSIMILATION: OBSERVATION INFLUENCE ANALYSIS DURING A GEOMAGNETIC STORM EVENT

This chapter presents observing system experiments in which a targeted observation strategy is used to estimate the global distribution of electron density during the 26 September 2011 geomagnetic storm. The work in this chapter is based on (Durazo *et al.*, 2017) and is published in the *Journal of Geophysical Research: Space Physics*.

5.1 Introduction

The ionospheric model used for this study is the Thermosphere Ionosphere Electrodynamics Global Circulation Model (TIEGCM). Synthetic observations of electron density are assimilated using the local ensemble transform Kalman filter (LETKF) data assimilation scheme, and the proposed targeted observation strategy is developed using the observation influence matrix diagnostic (Cardinali *et al.*, 2004). The observation influence is used to identify regions where additional observed vertical profiles may have a significant impact on forecast accuracy. In particular, a strategy for choosing locations to place these additional vertical profiles to target errors in specific state variables is developed and the impact of this choice on forecast accuracy is examine. This strategy may be useful for deployment of observations in real time to mitigate unwanted effects due to space-weather uncertainty or to inform future development of observational infrastructure in the future.

This approach is novel in that the analogue of the influence matrix has not been used with ensemble Kalman filters or the LETKF, and it has not been used as the basis for a targeted observation strategy in space weather forecasting. The use of

targeted locations for observations provides a considerable improvement in electron density estimation within a 600 km radius compared to randomly chosen locations for augmented vertical profiles. Our targeting strategy may also be used to improve accuracy in the estimation of other state variables or ionospheric drivers, which is demonstrated here for neutral wind estimates.

The layout of this chapter is as follows. An overview of the data assimilation system used in this work is provided in Section 5.2, Section 5.3 describes the influence matrix implementation in the LETKF and how it is used for targeted observations, Section 5.4 describes the observing system experiments with targeted observations and presents results, and a discussion and conclusions of these results are given in Section 5.6.

5.2 Overview

The data assimilation scheme used in these numerical experiments is the LETKF. For discussion purposes of this chapter, the main components of the assimilation procedure are briefly mentioned below, but the reader is referred to Section 3.4.1 for a more complete description of the LETKF procedure. In the following discussion, the analysis calculations are performed at time t_i only and the time subscript is excluded for clarity.

The LETKF procedure begins with an ensemble of k forecast (background) vectors, $\{\mathbf{u}^{b(j)}\}_{j=1}^{k}$, whose ensemble mean $\bar{\mathbf{u}}^b$ may be regarded as a maximum likelihood estimator of the "true" state $\mathbf{u}^t$. For each fixed grid point, L, there is an associated local window. Denote the d components of $\mathbf{u}^{b(j)}$ that represent the model state variables within the associated local window as $\mathbf{x}_L^{b(j)}$. The forecast covariance matrix within the local window is given by the $d \times d$ matrix, $\mathbf{P}_L^b = (k-1)^{-1}\mathbf{X}_L^b(\mathbf{X}_L^b)^{\mathrm{T}}$, where $\mathbf{X}_L^b$ is the $d \times k$ matrix of background ensemble perturbations

Let $\mathbf{y}_L^o$ be the ℓ-vector of observations in the local region and $\mathbf{R}_L$ the associated $\ell \times \ell$ covariance matrix. In a similar way, let $\mathbf{y}_L^{b(j)} = \mathbf{H}_L \mathbf{u}^{b(j)}$ denote the ensemble of predicted (background) measurements associated with the observation operator within the local region, with $j = 1, 2, \ldots, k$. Given the mean, $\bar{\mathbf{y}}_L^b = k^{-1} \sum_{i=1}^{k} \mathbf{y}_L^{b(i)}$, construct the $d \times k$ observation perturbation matrix $\mathbf{Y}_L^b$ whose jth column is $\mathbf{Y}_L^{b(j)} = \mathbf{y}_L^{b(j)} - \bar{\mathbf{y}}_L^b$, with $j = 1, 2, \ldots, k$. The analysis procedure is then carried out at each of the grid points as discussed in Section 3.4.1.

5.2.1 Data Assimilation System

The TIEGCM-LETKF data assimilation system runs as a forecasting step, which estimates the state of the ionosphere-thermosphere (I-T) system using the TIEGCM up to a time $t_n + \Delta t$, followed by an analysis step in which observations in the time window $[t_n - \Delta t, t_n + \Delta t]$ are assimilated using the LETKF. In all numerical experiments presented in this paper, the radius of the analysis window is taken to be $\Delta t = 0.5$ hrs, as has been done in other I-T data assimilation systems (Lee *et al.*, 2012) (Matsuo *et al.*, 2013). The result of the analysis step is an updated I-T state estimate at t_n that serves as the initial condition for the subsequent forecasting step. This forecast/analysis sequence is repeated at 1-hr intervals for the duration of the simulation starting at at 00:30 UTC on 26 September 2011. (Forecast/analysis time steps of less than 1-hr have been suggested by (Chen *et al.*, 2016), but they are not consider here.)

Each integration of the TIEGCM yields a 1-hr forecast of the I-T state vector. During each analysis step, a subset of the components are *analyzed*, i.e., updated using the assimilated observations. This subset is referred to as the *LETKF state vector*. The components of the TIEGCM state vector that are not included in the LETKF state vector are left unchanged during the analysis step. In our numerical

experiments, the LETKF state vector is composed of electron density (N_e), neutral temperature (T_n), zonal (U_n) and meridional (V_n) components of neutral winds, as well as atomic (O_1) and molecular (O_2) mass mixing ratios. This choice of LETKF state vector is done following the work of (Matsuo *et al.*, 2013),(Lee *et al.*, 2012). Prior to each analysis calculation, each component of the LETKF state vector is divided by their respective background global mean to yield a non-dimensionalized value. Following the analysis calculation, the respective factors for each state variable are used to dimensionalize the analyzed LETKF state vector.

5.2.2 Ionosphere Model

The ionospheric model used in this study is the TIEGCM, which is a first-principles, three-dimensional, non-linear model of the coupled I-T system that solves the momentum, energy, and continuity equations for neutral and plasma species, accounting for the electrodynamical processes governing their coupling. The coordinate system is spherical, with horizontal resolution of $5^\circ \times 5^\circ$ in longitude and latitude. The vertical resolution is of half-scale height, extending from about 97 km to about 600 km, depending on solar activity. A thorough model description can be found at the NCAR HAO website (http://www.hao.ucar.edu/modeling/tgcm/tie.php).

The ionosphere's interactions with thermospheric variables, such as neutral wind circulation, temperature, and oxygen composition, are represented explicitly. The effects of solar activity on the I-T system are represented parametrically using the daily solar radio flux at 10.7 cm ($F10.7$). In high-latitude regions, geomagnetic activity is parameterized using the Heelis model (Heelis *et al.*, 1982) and the Kp index. In particular, the Kp index is used to compute the hemispheric power (Hp) index, which parameterizes high-latitude auroral precipitation, and the cross-tail potential (Cp)

index, which parameterizes high-latitude ion convection. See Section 2.1.2 for more thorough description of these parameter inputs.

5.2.3 Generation of Synthetic Observations

Observations are synthetically generated from a "truth run", which is taken to be a TIEGCM simulation driven with the set of solar and geomagnetic parameters published by NOAA for 26 September 2011. The temporal evolution of these parameters is shown in the top two rows of Figure 5.1. The results shown in this chapter (Section 5.4) represent a "perfect model" scenario, in which the TIEGCM perfectly describes the state trajectory of the I-T system, and any forecast errors that arise are solely due to the performance of the data assimilation system. In reality, data assimilation systems must cope with errors and biases in the forecast model and in the observations, as they can contribute significantly to forecast uncertainty. The perfect-model assumption separates these issues from the performance of our targeted observation strategy and data assimilation system for this investigation. The true state trajectory is referred to as $\{\mathbf{u}_i^t\}_{i=1}^N$ and the associated $F_{10.7}$, Hp and Cp indices used to generate it are referred to as $\{\mathbf{p}_i^t\}$.

Synthetic observations are generated by sampling the electron density component of $\mathbf{u}_i^t$, as described in Section 5.2.3 at the same times and locations as the observing network of the Constellation Observing System for Meteorology, Ionosphere, and Climate/ Formosa satellite 3 (COSMIC) (Rocken *et al.*, 2000) during 26 September 2011. The COSMIC electron density vertical profiles typically vary from about 80 km to 800 km in altitude, with a vertical resolution of about 10 km. About 85 profiles are available for assimilation during each hour.

For the purposes of our targeted observation strategy, an augmented network of

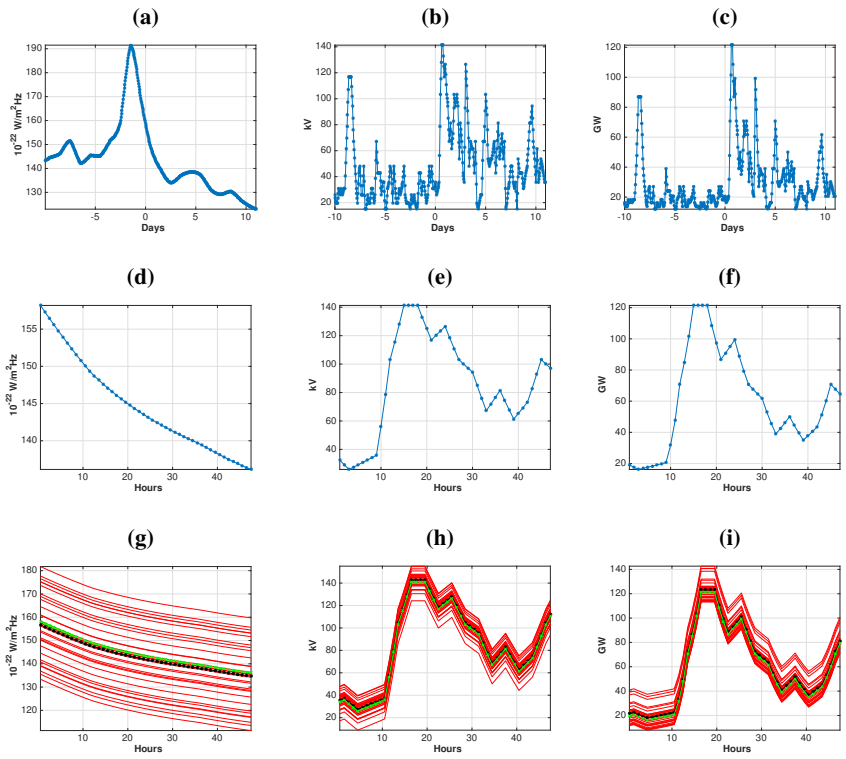

Figure 5.1: (a)-(c) Time series of the forcing parameters: (a) solar flux ($F_{10.7}$) in solar flux units (10^{-22}W/(m^{2}Hz)), (b) cross-tail potential (Cp) in kV, and (c) hemispheric power (Hp) in GW . The time origin is at 00:00 UTC on 26 September 2011, with the time interval spanning from ten days prior to ten days after. (d)-(f) Time series of the "true" forcing parameters beginning 00:00 UTC on 26 September 2011 and continuing to 00:00 UTC on 28 September 2011. (g)-(i) Analogous time series of the forcing parameter ensemble. The thin red curves represents individual ensemble members, the thick black curves are the respective ensemble means, and the green curves is the "truth".

observations, consisting of 62 globally distributed synthetic electron density vertical profiles are considered for assimilation during selected analysis steps. These vertical profiles are also generated from electron density component of $\mathbf{u}_i^t$ and extend from 100 km to 550 km in altitude with resolution of 5 km. The locations of the vertical profiles in the augmented network are fixed throughout the simulation.

The vertical profiles in the COSMIC and augmented networks are generated with additive Gaussian noise to represent observation processing error: $\mathbf{y}^o = \mathbf{H}\mathbf{u}^t + \boldsymbol{\epsilon}$, where $\mathbf{H}$ interpolates the electron density component of $\mathbf{u}_i^t$ to the respective observation

locations and times, and $\boldsymbol{\epsilon}$ is a Gaussian random vector with zero mean and covariance matrix $\mathbf{R}$.

The observing errors are assumed to be independent and to scale as a percentage of the "truth" at the observation locations, i.e., $\mathbf{R}$ is diagonal and its main diagonal is given by $\sigma^2\mathbf{H}\mathbf{u}^t$. Although observations obtained through retrievals, such as the COSMIC profiles, are typically correlated, they are treated as uncorrelated in our simulations for simplicity, but with an inflated value of the variance as described below. COSMIC retrievals have also been shown to have significant uncertainties associated with the Abel transform used in their inversion process, specifically at low altitudes in the ionosphere (Yue *et al.*, 2010),(Liu *et al.*, 2010).

The retrieval errors are approximated from the data in the third column of Figure 3 of (Liu *et al.*, 2010), which shows the horizontal error structure at at 150, 200, 250 and 300 km altitudes from 60°S-60°N latitudes at 06:00 UTC, on April-June 2008. Our reconstruction of these error structures at the same altitudes and hour are shown in panels (a)-(d) of Figure 5.2. The retrieval errors are shown in terms of percentage of relative error, which ranges from -80% to 80%. The horizontal retrieval errors are linearly interpolated to obtain the vertical structure shown in Panels (e)-(f) at local midnight (60°W) and local noon (60°E), respectively. These vertical error structures compare well with the third column of Figure 4 in (Liu *et al.*, 2010).

To apply the error structures shown in Figure 5.2 in the data assimilation system, this data is interpolated to all the grid points of the TIEGCM domain. Between 150 km and 300 km, Abel errors are linearly interpolated to all model altitudes and above 300 km altitudes, the errors are linearly extrapolated to zero. The errors are kept constant below 150 km altitudes and at latitudes above 60°N and south of 60°S. Temporally, it is assumed that this error structure remains the same and simply shifts

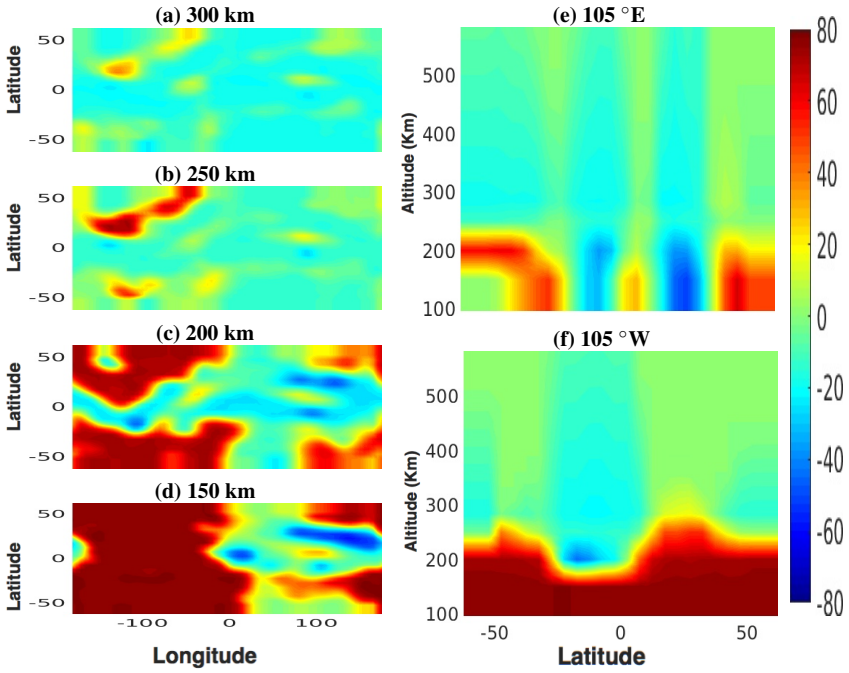

Figure 5.2: (a)-(d) Maps of estimated retrieval errors due to the Abel transform at 06:00 UTC. The maps are shown at 150, 200, 250 and 300 km altitudes from 60°S-60°N latitudes. (e)-(f) Altitudinal structure of retrieval errors from 100 to 700 km altitudes at local midnight (105°W) and local noon (105°E). The color scale is relative error given as a percentage and is the same in all panels.

towards the west at 2.5 deg/hr. The resulting error structure is linearly interpolated to the locations of the vertical profiles in the COSMIC and augmented networks to obtain σ_{inv}.

In reality, the retrieval errors are correlated in ways that depend in possibly complicated ways on the nature of the satellite observing platforms. To accommodate this reality, an extra 10 percent is added to the overall error variance. In all numerical experiments, the noise level used to generate the synthetic observations in the COSMIC and augmented networks is 10% measurement error plus the estimated error associated with the retrieval ($\sigma = 0.10 + \sigma_{inv}$). A similar approach to specifying observation error to COSMIC-like synthetic profiles is done in (Hsu *et al.*, 2014). Although the error structures related to the inversion process may differ considerably

during a geomagnetic storm, the error structure used here provides a reasonable representation of the dependence that the retrieval errors have on altitude. Panels (e) and (f) of Figure 5.2 show that the largest retrieval errors occur at altitudes below 250 km, especially during the nighttime. The results presented in Sections 4 and 5 have been computed with synthetic data generated with the error structures shown in Figure 5.2.

An ensemble of 40 forecasts is utilized in all numerical experiments presented in this chapter. Each member of the forecast ensemble is driven with parameter values of $F_{10.7}$, Hp and Cp drawn from a normal distribution centered around $\{\mathbf{p}_i^t\}_{i=1}^N$ at each forecast time. The standard deviation of the distribution of $F_{10.7}$ is taken to be its 21-day standard deviation during the spin-up period, which is taken to be from the 5 to 25 September 2011, and the standard deviation for the geomagnetic parameters is taken from ± 1.0 units of Kp. The respective standard deviations are 13×10^{-22} W/m^{2}Hz for $F_{10.7}$, 8.4 kV for Cp, and 7.2 GW for Hp. A similar approach to specifying standard deviations for these forcing parameters is used by (Matsuo and Araujo-Pradere, 2011),(Lee *et al.*, 2012),(Chen *et al.*, 2016). The third row of Figure 5.1 shows the temporal evolution of the forcing parameter ensemble throughout the simulation. Starting from the same climatology conditions provided by NCAR, each ensemble member is spun up from 00:00 UTC on 5 September 2011 to 00:30 on 26 September 2011, which is when the first batch of observations is assimilated.

5.3 Targeted Observation Strategy Using the Influence Matrix

As described in Section 5.2.3, a set of synthetic observations, consisting of electron density vertical profiles, is partitioned into two observation networks. The first network consists of observations sampled according to the COSMIC observing network on 26 September 2011, and the second network consists of a group of 62 globally

distributed vertical profiles. These two observation networks are referred to as the COSMIC and augmented observation networks, respectively, and the combined set of observations as the full observation network.

In numerical experiments presented, the COSMIC network is regarded as the base set of observations that are assimilated during 26 September 2011, and the potential benefits of assimilating additional synthetic electron density vertical profiles is considered. To begin, consider the analysis obtained when assimilating the full network of observations. The influence matrix is used to quantify the relative contribution to the analysis from each vertical profile in the augmented network, which is used to rank the contribution of the augmented vertical profiles according to the metric in Section 5.3.2. A formulation to compute the analysis adjustment due to the assimilation of augmented vertical profiles is also provided, which can be used to retrieve the baseline analysis obtained from the COSMIC network only. Below, the formulation of the influence matrix for data assimilation with the LETKF and how it might be used as a targeted observation strategy for ionospheric weather forecasting during the 26 September 2011 geomagnetic storm is presented.

5.3.1 Influence Matrix Formulation

The influence matrix is a diagnostic tool that quantifies the contribution of each assimilated observation on the analysis. It has typically been used in the context of linear estimators, and (Cardinali *et al.*, 2004) introduced its analogue in the context of data assimilation for numerical weather prediction using the Kalman filter. In this section, a brief derivation of its analogue for the LETKF is presented; additional derivation details may be found in Appendix C. The influence of the assimilated observations is computed independently for each grid point, along with the computation of each of the local analyses. As done in the assimilation procedure described in Sec-

tion 3.4.1, consider a fixed grid point L, and use L as a subscript to denote quantities associated with the grid point or its associated local region.

Suppose there are ℓ observations from the full observation network within the local region associated with L. Denote the vector of local observations and their associated Gaussian errors with the $\ell \times 1$ vectors $\mathbf{y}_L^F$ and $\boldsymbol{\epsilon}_L^F$, respectively, and the observation covariance matrix by $\mathbf{R}_L^F$. The F superscript emphasizes that these observational quantities are associated with the full observation network. Consider the d components in the LETKF state vector. Denote the background ensemble mean at L by the $d \times 1$ vector $\bar{\mathbf{x}}_L^b$, its $d \times 1$ vector of Gaussian errors by $\boldsymbol{\epsilon}_L^b$, and its associated $d \times d$ local forecast covariance matrix by $\mathbf{P}_L^b$. The local observation operator is denoted by the $\ell \times d$ matrix $\mathbf{H}_L^F$. See Section 3.4.1 for additional details on the construction of these matrices.

Following the notation and derivation in (Cardinali *et al.*, 2004), let $\mathbf{z}_L^F = \left[(\mathbf{y}_L^F)^{\mathrm{T}}\ (\bar{\mathbf{x}}_L^b)^{\mathrm{T}}\right]^{\mathrm{T}}$ and $\boldsymbol{\epsilon}_L = \left[(\boldsymbol{\epsilon}_L^F)^{\mathrm{T}}\ (\boldsymbol{\epsilon}_L^b)^{\mathrm{T}}\right]^{\mathrm{T}}$ be the $(\ell + d) \times 1$ augmented state and error vectors, respectively. Also define the $(\ell + d) \times d$ matrix $\mathbf{Z}_L^F = \left[(\mathbf{H}_L^F)^{\mathrm{T}}\ \mathbf{I}_d\right]^{\mathrm{T}}$, where $\mathbf{I}_d$ is the $d \times d$ identity matrix. The Kalman filter analysis problem may be posed as the weighted linear regression problem

$$\mathbf{z}_L^F = \mathbf{Z}_L^F \mathbf{x}_L^t + \boldsymbol{\epsilon}_L, \tag{5.1}$$

where $\mathbf{x}_L^t$ refers to the d components of the state vector corresponding to the "truth" at the grid point L. The observation and background errors are assumed to be mutually independent, so the system covariance matrix is given by the $(\ell + d) \times (\ell + d)$ block diagonal matrix $\boldsymbol{\Omega}_L^F = \begin{bmatrix} \mathbf{R}_L^F & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_L^b \end{bmatrix}$. The least squares solution of problem (5.1) is given by $\widehat{\mathbf{z}}_L = \mathbf{Z}_L^F \widehat{\mathbf{x}}_L$, where

$$\widehat{\mathbf{x}}_L = \left((\mathbf{Z}_L^F)^{\mathrm{T}} (\boldsymbol{\Omega}_L^F)^{-1} \mathbf{Z}_L^F\right)^{-1} (\mathbf{Z}_L^F)^{\mathrm{T}} (\boldsymbol{\Omega}_L^F)^{-1} \mathbf{z}_L^F \tag{5.2}$$

is equivalent to the analysis ensemble mean, $\bar{\mathbf{x}}_L^{a(F)}$, of the Kalman filter (Cardinali

et al., 2004). The $(\ell + d) \times (\ell + d)$ influence matrix for the regression problem (5.1) is defined to be

$$\mathbf{S}_L = \frac{\partial \widehat{\mathbf{z}}_L}{\partial \mathbf{z}_L} = \begin{bmatrix} \frac{\partial(\mathbf{H}_L^F \bar{\mathbf{x}}_L^a)}{\partial \mathbf{y}_L^F} & \frac{\partial(\mathbf{H}_L^F \bar{\mathbf{x}}_L^a)}{\partial \bar{\mathbf{x}}_L^b} \\ \frac{\partial \bar{\mathbf{x}}_L^a}{\partial \mathbf{y}_L^F} & \frac{\partial \bar{\mathbf{x}}_L^a}{\partial \bar{\mathbf{x}}_L^b} \end{bmatrix} = \begin{bmatrix} \mathbf{H}_L^F \mathbf{P}_L^{a(F)} (\mathbf{H}_L^F)^{\mathrm{T}} (\mathbf{R}_L^F)^{-1} & \mathbf{H}_L^F \mathbf{P}_L^{a(F)} (\mathbf{P}_L^b)^{-1} \\ \mathbf{P}_L^{a(F)} (\mathbf{H}_L^F)^{\mathrm{T}} (\mathbf{R}_L^F)^{-1} & \mathbf{P}_L^{a(F)} (\mathbf{P}_L^b)^{-1} \end{bmatrix}. \quad (5.3)$$

The entries of $\mathbf{S}_L$ are the partial derivatives of the local analysis, $\bar{\mathbf{x}}_L^{a(F)}$, in observation and model space, with respect to each of the ℓ observations and d analyzed components of the background state vector at L. Denote the lower and upper left sub-matrices of $\mathbf{S}_L$ as $\mathbf{S}_L^{XF}$ and $\mathbf{S}_L^{FF}$, respectively. $\mathbf{S}_L^{XF}$ is a $d \times \ell$ matrix of partial derivatives whose jth column quantifies the influence of the jth observation on the d components of $\bar{\mathbf{x}}_L^{a(F)}$. Similarly, $\mathbf{S}_L^{FF}$ is an $\ell \times \ell$ matrix whose jth column quantifies the influence of the jth observation on each component of the $\ell \times 1$ analyzed observation vector, $\mathbf{H}_L^F \bar{\mathbf{x}}_L^{a(F)}$. The subscript L emphasizes that the influences are derived strictly from observations in the local region about the grid point L; the superscripts denote that the influences are from the full network of observations and act upon the analyzed state vector in model space (XF) and in observation space (FF), respectively.

Denote the matrices of ensemble perturbations for the background forecast and observations associated with L as $\mathbf{X}_L^b$ and $\mathbf{Y}_L^F$, respectively, and the analysis covariance matrix in the space of the local ensemble solutions as $\widetilde{\mathbf{P}}_L^{a(F)}$ (see Section 3.4.1). Using these quantities together with Equation (3.33) yields $\mathbf{S}_L^{XF} = \mathbf{X}_L^b \widetilde{\mathbf{P}}_L^{a(F)} (\mathbf{Y}_L^F)^{\mathrm{T}} (\mathbf{R}_L^F)^{-1}$ and $\mathbf{S}_L^{FF} = \mathbf{Y}_L^F \widetilde{\mathbf{P}}_L^{a(F)} (\mathbf{Y}_L^F)^{\mathrm{T}} (\mathbf{R}_L^F)^{-1}$. As long as the set of assimilated observations varies smoothly among neighboring grid points, $\widetilde{\mathbf{P}}_L^{a(F)}$ also varies smoothly (see Section 3.4.1), and so does the observation influence.

Suppose that the observations from the COSMIC and augmented networks are given by the $\ell_C \times 1$ vector $\mathbf{y}_L^C$ and the $\ell_A \times 1$ vector $\mathbf{y}_L^A$, respectively. For the purpose of exposition, the observations in the full network are partitioned as $\mathbf{y}_L^F = \left[(\mathbf{y}_L^C)^{\mathrm{T}} \, (\mathbf{y}_L^A)^{\mathrm{T}} \right]^{\mathrm{T}}$, but it is emphasized that the following results do not depend on

this ordering. The associated partitions on $\mathbf{Y}_L^F$ and $\mathbf{R}_L^F$ are given by $\mathbf{Y}_L^F = \left[(\mathbf{Y}_L^C)^{\mathrm{T}} \, (\mathbf{Y}_L^A)^{\mathrm{T}} \right]^{\mathrm{T}}$ and $\mathbf{R}_L^F = \begin{bmatrix} \mathbf{R}_L^C & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_L^A \end{bmatrix}$. The partitioned structure of the influence matrices $\mathbf{S}_L^{XF}$ and $\mathbf{S}_L^{FF}$ is then

$$\mathbf{S}_L^{XF} = \begin{bmatrix} \mathbf{S}_L^{XC} \, \mathbf{S}_L^{XA} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_L^b \widetilde{\mathbf{P}}_L^{a(F)} (\mathbf{Y}_L^C)^{\mathrm{T}} (\mathbf{R}_L^C)^{-1} & \mathbf{X}_L^b \widetilde{\mathbf{P}}_L^{a(F)} (\mathbf{Y}_L^A)^{\mathrm{T}} (\mathbf{R}_L^A)^{-1} \end{bmatrix} \quad (5.4)$$

$$\mathbf{S}_L^{FF} = \begin{bmatrix} \mathbf{S}_L^{CC} \, \mathbf{S}_L^{CA} \\ \mathbf{S}_L^{AC} \, \mathbf{S}_L^{AA} \end{bmatrix} = \begin{bmatrix} \mathbf{Y}_L^C \widetilde{\mathbf{P}}_L^{a(F)} (\mathbf{Y}_L^C)^{\mathrm{T}} (\mathbf{R}_L^C)^{-1} & \mathbf{Y}_L^C \widetilde{\mathbf{P}}_L^{a(F)} (\mathbf{Y}_L^A)^{\mathrm{T}} (\mathbf{R}_L^A)^{-1} \\ \mathbf{Y}_L^A \widetilde{\mathbf{P}}_L^{a(F)} (\mathbf{Y}_L^C)^{\mathrm{T}} (\mathbf{R}_L^C)^{-1} & \mathbf{Y}_L^A \widetilde{\mathbf{P}}_L^{a(F)} (\mathbf{Y}_L^A)^{\mathrm{T}} (\mathbf{R}_L^A)^{-1} \end{bmatrix} \quad (5.5)$$

The sum of the kth row of $\mathbf{S}_L^{XC}$ and $\mathbf{S}_L^{XA}$ are respectively referred to as the cumulative influence of the COSMIC and augmented networks on the kth component of $\bar{\mathbf{x}}_L^{a(F)}$. Computing $\mathbf{S}_L^{XC}$ and $\mathbf{S}_L^{XA}$ at each grid point yields the spatial structure of the cumulative observation influence of each observation network on the kth analyzed component of the global state vector. The top row of Figure 5.3 shows maps of the cumulative COSMIC observation influence on the analyzed (a) electron density, (b) zonal component of neutral winds, and (c) molecular oxygen composition, at 370 km altitude, at 12:30 UTC on 26 September 2011. The magenta markings denote the locations of the COSMIC vertical profiles. The white horizontal lines in the top row mark the 77.5°N latitude, where the observation influence variations with respect to altitude are shown for the same variables in the bottom row. The white lines in the bottom row marks the 370 km altitude. As described in Section D.2.3, the analyzed state variables and observations are non-dimensionalized, so the observation influences shown are in non-dimensional units.

A positive observation influence at L indicates that the analysis would increase (decrease) if the assimilated COSMIC observations were to increase (decrease), at the rate given by the cumulative influence. For example, over the Persian Gulf region, where electron density has positive observation influence (red areas) of about 1 unit,

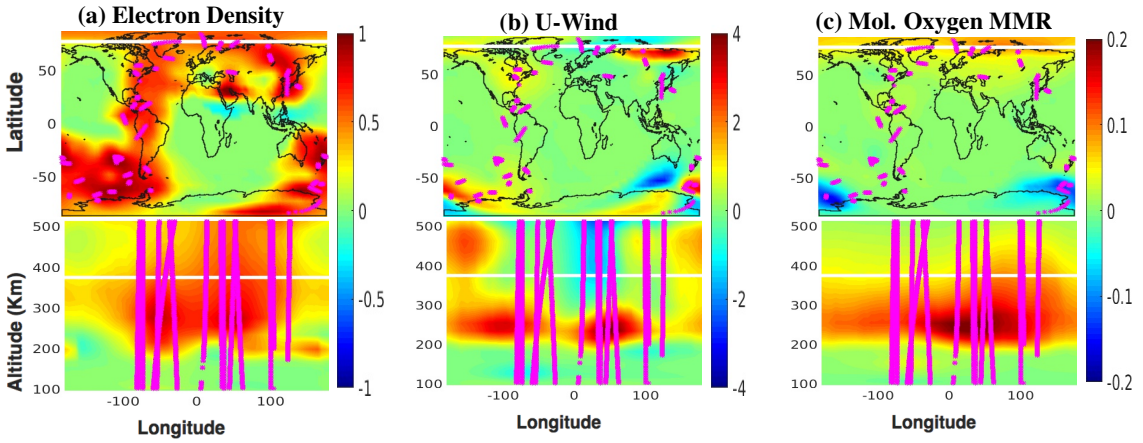

Figure 5.3: (Top row) Global maps of satellite observation influence at 12:30 UTC on 26 September 2011 at altitude 370 km for (a) electron density, (b) zonal component of neutral winds and (c) molecular oxygen composition mass mixing ratio (MMR). (Bottom row) Maps of the same variables and time as the top, showing variations of satellite observation influence at a fixed latitude of 77.5°N, as denoted by the white dotted line in the top row. The white line in the second row denotes the 370 km altitude. The location of the synthetic satellite observations are marked by the magenta lines. All figures are in non-dimensional units (see Section 3).

a cumulative increase (decrease) of 1 unit in the local COSMIC observations would result in an increase (decrease) of 1 unit in the analyzed electron density. The Siberian region has negative observation influence (blue areas) of -4 units for the zonal component of neutral winds (shown in the second column), so a cumulative increase (decrease) of 1 unit in the local COSMIC observations would result in an decrease (increase) of 4 units in the analyzed zonal component of neutral winds at L. Regions where the observation influence is near 0 (light green areas) are regions where COSMIC observations have little contribution to the analysis, relative to the augmented observations or the background forecast.

Let $\bar{\mathbf{x}}_L^{a(F)}$ and $\bar{\mathbf{x}}_L^{a(C)}$ denote the local analysis ensemble means determined by assimilating observations in the local region about L from the full and COSMIC observation

networks, respectively. The difference between them can be computed from a rank-ℓ_A update given by

$$\bar{\mathbf{x}}_L^{a(F)} - \bar{\mathbf{x}}_L^{a(C)} = \mathbf{S}_L^{XA} \left[\mathbf{I} - \mathbf{S}_L^{AA} \right]^{-1} \mathbf{r}_L^A, \tag{5.6}$$

where $\mathbf{r}_L^A = \mathbf{y}_L^A - \mathbf{H}_L^A \bar{\mathbf{x}}_L^{a(F)}$ is the $s \times 1$ vector of analysis residuals at the observation locations from the augmented network. The derivation of Equation (5.6) and an equivalent but more computationally efficient formulation to compute the inverse factor is provided in Appendix C. The quantity given by Equation (5.6) is referred to as the analysis adjustment due to the augmented observations at the grid point L and may be used to retrieve $\bar{\mathbf{x}}_L^{a(C)}$ without explicitly computing it.

5.3.2 Using the Influence Matrix to Target Observations

Maps of cumulative observation influence from augmented observations, which are constructed at each grid point L by summing over the columns of $\mathbf{S}_L^{XA}$ (see Equation (5.4)), can identify where vertical profiles from the augmented network have the highest degree of influence on the analyzed components of the global state vector. However, these maps do not quantify the total influence that each individual augmented vertical profile has on the components of the analyzed global state vector. For example, at a fixed grid point, if multiple augmented vertical profiles are assimilated, the distribution of influence among the augmented profiles may be roughly equal, or one of the profiles may have a majority of the influence. The distribution of influence among augmented vertical profiles also generally changes for each grid point, and while some vertical profiles may influence the global analyzed state vector at some of the same grid points, they will generally also influence different sets of grid points. In what follows, the metric used to measure the total influence of each

augmented vertical profile on the global analyzed state vector is presented. Consider a vertical profile, P, from the augmented network. Generally the observations in P belong to the local region associated with a set of grid points G_p. Choose a grid point $L \in G_P$. The influence from each augmented observation on $\bar{\mathbf{x}}_L^{a(F)}$ is given by $\mathbf{S}_L^{XA}$.

If P is the only augmented vertical profile within the local region associated with L, then the entries of $\mathbf{S}_L^{XA}$ give the influence of each observation in P on L. The sum of the kth row in $\mathbf{S}_L^{XA}$ yields the cumulative influence of the whole vertical profile, P, on the kth component of $\bar{\mathbf{x}}_L^{a(F)}$, which is denote by $\mathbf{I}_{L(k)}^{(P)}$. The total influence contribution of P on the kth component of the analyzed global state vector is given by summing the cumulative influence of P over all grid points $L \in G_P$:

$$\mathbf{T}_k^{(P)} = \sum_{L \in G_P} \left| \mathbf{I}_{L(k)}^{(P)} \right|, \tag{5.7}$$

where the vertical bars denote the absolute value. If there are $s > 1$ augmented vertical profiles in the local region associated with L, partition $\mathbf{S}_L^{XA}$ as $\mathbf{S}_L^{XA} = \left[\mathbf{S}_L^{XA_1} \, \mathbf{S}_L^{XA_2} \cdots \mathbf{S}_L^{XA_s} \right]$, where $\mathbf{S}_L^{XA_i}$ contains the observation influence of each observation in the ith augmented vertical profile on $\bar{\mathbf{x}}_L^{a(F)}$. Then the cumulative influence of the ith augmented vertical profile is obtained by summing over the columns of $\mathbf{S}_L^{XA_i}$.

The metric defined by Equation (5.7) is used to rank the total influence of each vertical profile from the augmented network on the analyzed state variables. As seen in Figure 5.3, the observation influence is generally different for each analyzed state variable, which directly affects the ranking of vertical profiles from the augmented network. The kth state variable is said to be *targeted* when the m highest ranked augmented vertical profiles for the kth state variable are selected to be included in the analysis procedure.(The value $m = 5$ is taken in the numerical experiments shown

in the next section.) A comparison of forecast improvement when targeting different state variables is also given in the next section.

5.4 Numerical Experiments and Results

This section described the results of several observing system experiments in which the geomagnetic storm of 26 September 2011 is simulated with the TIEGCM-LETKF data assimilation system described in Section 5.2.1. In particular, the targeted observation strategy discussed in Section 5.3.2 is employed to optimally select the most influential vertical profiles from the augmented network of observations. The benefits of assimilating targeted observations during the main geomagnetic storm phase are quantified by comparing the results to the forecast obtained without targeted observations.

The first row of Figure 5.4 shows the temporal evolution of the altitude-averaged background estimates of electron density during the geomagnetic storm at 12:30, 16:30, and 20:30 UTC on 26 September 2011. At the beginning of the geomagnetic storm (12:30 UTC), the peak electron density is located mostly in the daytime equatorial region and extends into the mid-latitudes. The background estimates of electron density overestimate the true electron density in the daytime low-to mid-latitudes and also at the polar regions as indicated by the positive (red) deviations from the truth shown in the third row of Figure 5.4.

During peak geomagnetic disturbance conditions (16:30 UTC), electron density is significantly enhanced in daytime mid-to high-latitude regions, especially in the southern hemisphere, and also in the nighttime high-latitude regions. In the day time regions, electron density is overestimated in the mid-latitude regions, especially over the Atlantic Ocean area and also in the nighttime equatorial regions. Electron density is significantly underestimated in the southern polar region as indicated by

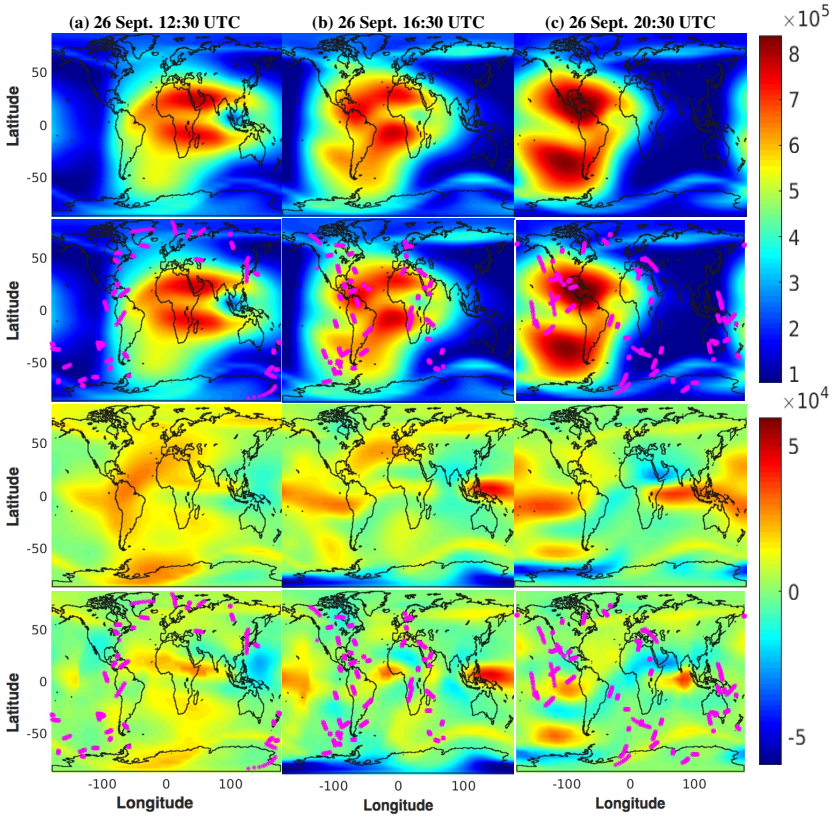

Figure 5.4: Global maps of background (first row) and analysis (second row) estimates of electron density, in el/cm^3, averaged from 200 km to 500 km altitudes at 12:30, 16:30 and 20:30 UTC on 26 September 2011. Rows 3-4: Analogous global maps of electron density deviations from the truth for the background and analyzed estimates, in el/cm^3, at 12:30, 17:30 and 20:30 UTC on 26 September 2011. The analyzed electron density (second row) and its deviation from the truth (fourth row) are obtained by assimilating only synthetic COSMIC observations whose locations are shown by magenta markings. The color scale in the first and second rows differs from the third and fourth rows.

the negative (blue) deviations from the truth in the third row. Finally, as geomagnetic conditions relax (20:30 UTC), the electron density remains overestimated in the daytime mid-latitude and nighttime equatorial regions, but high-latitude estimates improve.

The second and fourth rows of Figure 5.4 show the analysis obtained by assimilating the COSMIC network of observations and its the deviation from the truth, respectively. The locations of the assimilated vertical profile are shown by the ma-

genta markings, The assimilation of these observations reduces many of the electron density errors seen in the background estimates, except in regions of sparse data coverage and also in the southern polar region during the peak geomagnetic disturbance conditions (16:30 UTC). The goal of our targeted observation strategy is to optimally choose vertical profiles from the augmented network and add them to the set of assimilated COSMIC observations to reduce these electron density errors throughout the geomagnetic storm.

Now consider the potential impact of assimilating vertical profiles from the augmented network. The first row of Figure 5.5 shows global maps of cumulative observation influence from augmented vertical profiles on the analyzed electron density estimates at 12:30, 16:30, and 20:30 UTC. The locations of the vertical profiles in the augmented network are denoted by the white circles in each of the panels. These maps are constructed by summing over the columns of each $\mathbf{S}_L^{XA}$ (see Equation (5.4)) as described in Section 5.3.1. Augmented vertical profiles generally have positive influence on the analyzed electron density and the influence is strongest in areas of sparse COSMIC network coverage.

The cumulative observation influence from augmented observations on the analyzed estimates of U_n and O_1 is shown in the second and third rows of Figure 5.5. The observation influence for U_n is generally restricted to high-latitude regions, where the strongest neutral wind activity occurs. At 50° latitude and higher there are nearly adjacent regions of positive and negative observation influence on the analyzed estimate of U_n. The influence of augmented observations on O_1 is strongest at high latitudes although it does extend to the mid-latitudes as well. Each panel of Figure 5.5 also denotes the location of the five most influential vertical profiles for each respective analyzed variable with the white circles, as computed by the procedure described in Section 5.3.2. The locations of the five most influential vertical profiles vary consid-

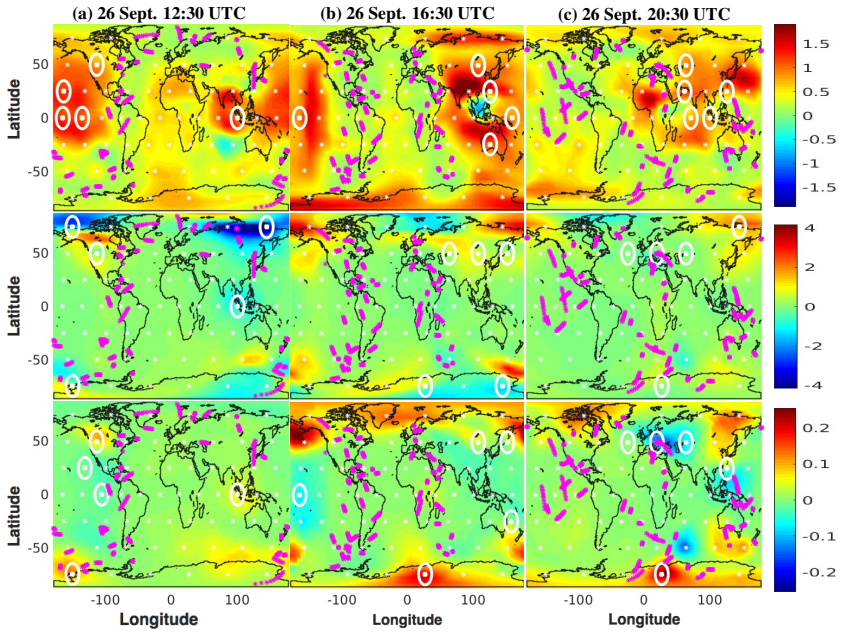

Figure 5.5: Global maps of influence from the augmented vertical profiles on the analysis, averaged over 200 km to 500 km altitudes at 12:30, 16:30, and 20:30 UTC. The observation influence fields are constructed and non-dimensionalized as described in Section D.2 for (row 1) electron density, (row 2) the zonal components of the neutral winds and (row 3) molecular oxygen composition. The color scale differs for each of the rows. COSMIC and augmented vertical profile locations are marked in magenta and white, respectively. The five most influential augmented profiles for each respective variable, as described in Section 5.3.2, are circled in white in each panel.

erably depending on the analyzed variable being considered. A state variable is said to be targeted when the five vertical profiles from the augmented network with the largest influence on that variable are included in the assimilation procedure.

Now focus on the impact of assimilating the five most influential vertical profiles from the augmented observation network on the analyzed electron density. The panels in Figure 5.6(a) show the temporal evolution of the altitude-averaged analyzed electron density from 11:30 UTC to 21:30 UTC on 26 September 2011. The five vertical profiles from the augmented network selected to target electron density during each analysis time are included in the assimilation procedure, and their locations are

shown by the thick white circles. A few of the selected vertical profiles are located in

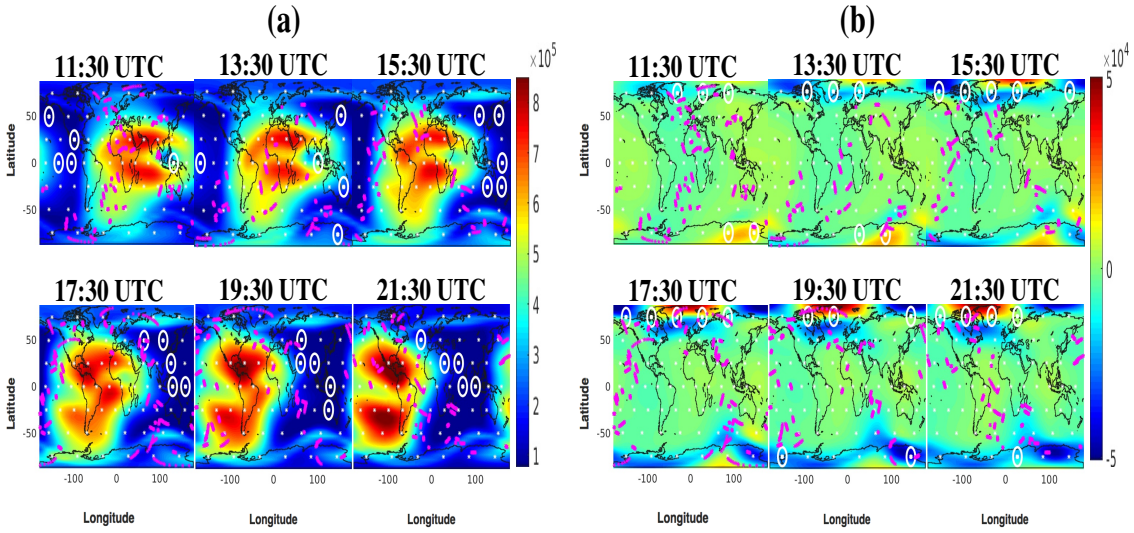

Figure 5.6: (a) Global maps of analyzed electron density, in units of el/cm^3, obtained when using the targeting scheme described in Section 5.3.2 to target electron density. The global maps are shown starting from 11:30 UTC to 21:30 UTC on 26 September 2011 and each map is averaged from 200 to 500 km in altitude. (b) Analogous maps of analyzed zonal component of the neutral winds ($\mathbf{U_n}$) obtained when targeting the zonal component of neutral winds ($\mathbf{U_n}$). Synthetic COSMIC and augmented vertical profile locations in each case are shown in magenta and white, respectively, with the top five most influential augmented vertical profiles in each case circled in white.

equatorial regions, for example at 11:30 UTC and 13:30 UTC, but the majority are located in the nighttime high-latitude regions, where the rapidly changing geomagnetic conditions occur.

The panels in Figure 5.6(b) show the analyzed altitude-averaged analyzed U_n in the case where the five most influential vertical profiles for U_n are included in the assimilation procedure. The locations of these vertical profiles correspond to regions where the influence of augmented vertical profiles have the most benefit on the analyzed U_n estimate. The locations of the vertical profiles selected to target U_n are considerably different than those selected to target electron density, though the majority of the targeted vertical profiles are also located in the high latitude regions.

The impact of targeted vertical profiles is evaluated with the RMSE difference of the forecast from the truth, defined as

$$RMSE = \sqrt{\frac{\sum_{L \in I} \left(\mathbf{e}_L - \mathbf{e}_L^t\right)^2}{N}}, \tag{5.8}$$

where $\mathbf{e}_L$ is the electron density component of the LETKF state vector, $\mathbf{x}_L$, at the grid point L and $\mathbf{e}_L^t$ is the corresponding truth. The summation is done over the set I of N grid points located within a 600 km horizontal radius of the vertical profiles selected to target electron density, shown by the thick white circles in Figure 5.6(a). Figure 5.7(a) shows the time series of the root mean square error (RMSE) of analyzed electron density estimates obtained without assimilating any vertical profiles from the augmented network (blue) and when assimilating five vertical profiles selected to target electron density (red).

Augmented vertical profiles are assimilated starting at 11:30 UTC, which corresponds to the onset of disturbed geomagnetic conditions, so the RMSE is the same in both cases before 11:30 UTC. As geomagnetic conditions become increasingly disturbed, the RMSE of analyzed electron density increases sharply when no augmented vertical profiles are assimilated, peaking at 14:30 UTC and 16:30 UTC, and then decreasing as geomagnetic activity relaxes, suggesting that areas of large electron density uncertainty are identified during peak geomagnetic disturbance conditions. On the other hand, the RMSE of analyzed electron density with the five vertical profiles targeting electron density remains relatively constant throughout the geomagnetic storm, demonstrating that the errors in these regions of large electron density uncertainty are reduced. The largest improvement due to assimilation of the five targeted

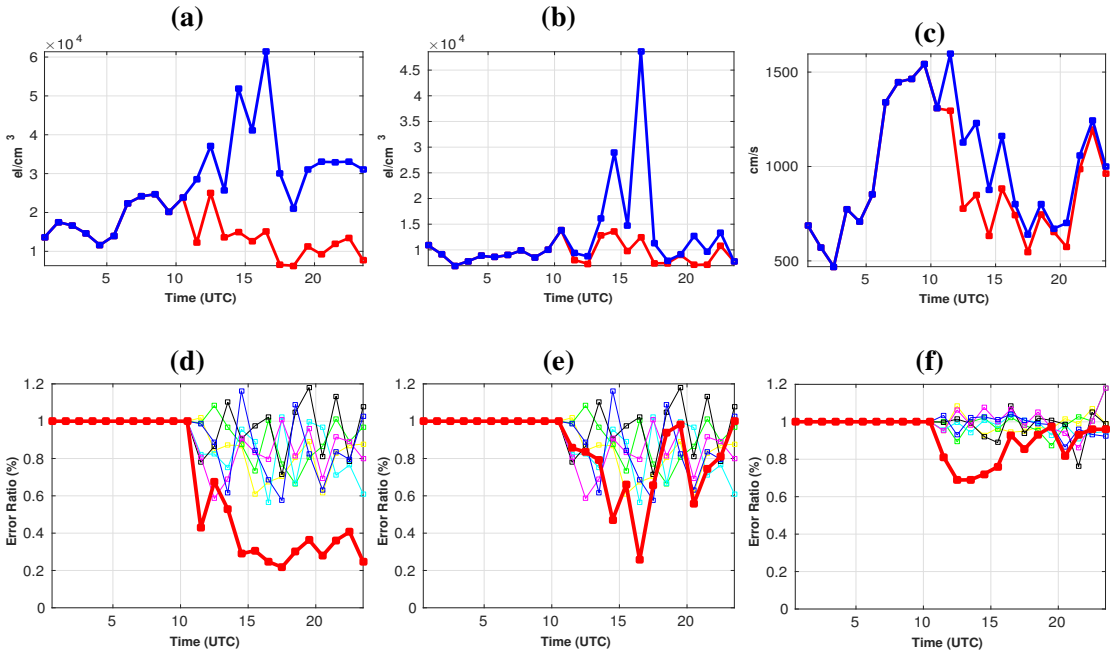

Figure 5.7: (a) Time series of root-mean-square error (RMSE) of analyzed electron density in units of el/cm^3, averaged vertically from 100 km to 550 km altitudes and horizontally over 600 km regions centered around the 5 augmented vertical profiles used to target electron density. The red and blue curves correspond to the analyzed electron density RMSE with and without the 5 targeted vertical profiles, respectively. (b)-(c) Analogous plots of the RMSE of (a) analyzed electron density and (c) the zonal component of neutral winds (U_n) centered around the 5 vertical profiles used to target the zonal component of neutral winds. (d) Time series of RMSE ratios of analyzed electron densities obtained with and without five augmented electron density vertical profiles. The thick red curve corresponds to the ratio of the red to blue curves in panel (a) and thin curves show the analogous RMSE ratios obtained for six different sets of five randomly chosen vertical profiles. (e)-(f) Analogous plots of analyzed electron density and U_n RMSE ratios corresponding to the RMSE curves shown in (b) and (c), respectively. In panels (e)-(f), the same sets of randomly chosen augmented vertical profiles are assimilated.

vertical profiles occurs at 16:30 UTC, where the analyzed electron density RMSE is reduced by about 80% over the 600 km regions.

Figure 5.7(b) shows an analogous RMSE time series of analyzed electron density, with I in Equation (6.11) corresponding to 600 km regions centered around the five vertical profiles selected to target U_n, whose locations are shown by the thick white

circles in Figure 5.6(b). These are regions where the analyzed estimates of U_n are relatively uncertain and could potentially cause significant uncertainties in electron density estimates. The largest improvement in RMSE is of about 60% and it occurs during the main phase (16:30 UTC). Figure 5.7(c) shows an analogous time series of RMSE for the analyzed estimates of U_n, computed over the same regions as in Figure 5.7(b). The RMSE values calculated here are done with a similar equation to (6.11), but for the U_n component of the LETKF state vector. The largest improvements in the analyzed neutral wind estimates are about 30% as geomagnetic conditions become increasingly disturbed between 11:30 and 15:30 UTC.

As expected, RMSE values of analyzed electron density and neutral winds do improve within the 600 km region centered around the augmented vertical profiles. The relative benefit of using our targeting procedure is compared against the benefit of using randomly selected vertical profiles from the augmented network. To do so, conisder the ratio of analyzed electron density RMSE values that are obtained with and without the assimilation of the five vertical profiles selected to target electron density. For example, the thick red curve in Figure 5.7(d) is the ratio of the time series of analyzed electron density RMSE values that correspond to the curves shown in Figure 5.7(a). For comparison, the analogous time series of RMSE ratio values of analyzed electron density is computed for six different sets of five vertical profiles that are randomly selected from the augmented observation network. Since the region I over which the RMSE values are computed in each case is different, the RMSE ratio values yield a direct comparison of improvement in analyzed electron density RMSE values. The RMSE ratio time series for the each set of randomly chosen vertical profiles is shown by the thin curves in Figure 5.7(d).

Figure 5.7(d) shows that using augmented vertical profiles that are randomly chosen reduces electron density RMSE values on average by about 15% over 600 km

regions throughout the geomagnetic storm, which is considerably less than the average of 65% RMSE reduction seen when assimilating vertical profiles chosen to target electron density. Additionally, the RMSE reduction obtained with targeted vertical profiles is more consistent for each analysis calculation during the geomagnetic storm and the RMSE reduction is at the very least 60% after 14:30 UTC, demonstrating that the regions of largest electron density uncertainty are consistently identified and improved.

Figure 5.7(e) shows the analogous time series of analyzed electron density RMSE ratio obtained when using vertical profiles to target U_n, where the thick red curve corresponds to the ratio of the curves shown in Figure 5.7(b). The RMSE ratio for the same six sets of randomly chosen vertical profiles is also shown. Assimilating five augmented vertical profiles chosen to target U_n generally yields larger reduction of RMSE in analyzed electron density than the groups of randomly vertical profiles, most notably during the sharp increase in geomagnetic disturbance from 13:30 to 16:30 UTC, where improvements on average are about 40% and are as large as 70%. After 16:30 UTC, the RMSE reduction is on average comparable to those obtained with randomly selected vertical profiles, which may indicate that there are no regions with outstanding levels of uncertainty in U_n, compared to the peak geomagnetic disturbance conditions that occur from 13:30 to 16:30 UTC.

Figure 5.7(f) shows the time series of RMSE ratio values for analyzed U_n estimates when targeting U_n, with the thick red curve corresponding to the ratio of RMSE values on Figure 5.7(c). Similarly to electron density estimates at these locations as Figure 5.7(e), there is considerable reduction in RMSE during peak geomagnetic disturbance conditions from 12:30 to 16:30 UTC compared to when using the randomly chosen vertical profiles. The RMSE reduction in analyzed estimates of U_n during this time period are on average about 25%, whereas the average RMSE reduction when

using randomly chosen vertical profiles during this same time period is only of about 2%. The RMSE reduction in analyzed $\mathbf{U_n}$ obtained with targeted vertical profiles is on average comparable to that of the randomly selected vertical profiles, as geomagnetic conditions begin to relax after 16:30 UTC, which is also seen in Figure 5.7(e). These RMSE time series results demonstrate that our targeting strategy may be used to effectively target areas of large uncertainty in electron density, and other state variables such as neutral winds to considerably improve forecast estimates.

5.5 Analysis Adjustment Due to the Assimilation of Augmented Observations

In this section, the observation influence is used to efficiently compute the exact change in the analysis at a grid point L due to the assimilation of the augmented observations. Consider the analysis at the grid point $\bar{\mathbf{x}}_L^{a(F)}$ obtained when assimilating the full network of observations ($\mathbf{y}_L^F$). Due to the mathematical formulation of the influence matrix, the effect of assimilating the augmented set of observations is quantified with the change in $\bar{\mathbf{x}}_L^{a(F)}$ when *excluding* these observations from the assimilation procedure. Begin by first computing $\bar{\mathbf{x}}_L^{a(F)}$. The observing network when excluding the augmented network is simply the COSMIC network ($\mathbf{y}_L^C$), which yields the analysis denoted by $\bar{\mathbf{x}}_L^{a(C)}$. Suppose there are ℓ_A augmented observations being assimilated. The adjustment on the analysis state vector is derived from a rank-ℓ_A update given by

$$\bar{\mathbf{x}}_L^{a(F)} - \bar{\mathbf{x}}_L^{a(C)} = \mathbf{S}_L^{XA} \left[\mathbf{I} - \mathbf{S}_L^{AA} \right]^{-1} \mathbf{r}_L^A, \tag{5.9}$$

where $\mathbf{r}_L^A = \mathbf{y}_L^A - \mathbf{H}_L^A \bar{\mathbf{x}}_L^{a(F)}$ is the $s \times 1$ vector of analysis residuals corresponding to the observations from the augmented network. Equation (5.9) gives the analysis adjustment due to augmented observations at L. (If $\ell_A = 0$, no observations from the augmented network are in the local region, and so the analysis adjustment is $\mathbf{0}$.) Repeating this procedure for every grid point yields the spatial structure of the

analysis adjustment due to the augmented network on the analyzed components of the global state vector. Additionally, Equation (5.9) can be used to readily retrieve $\bar{\mathbf{x}}_L^{a(C)}$ without its explicit computation. A similar approach can be taken to derive the analysis adjustment due to the assimilation of any subset of the full network of observations.

The first row of Figure 5.8 shows the altitude-averaged global analysis adjustment due to the assimilation of the augmented network for electron density at 12:30, 16:30 and 20:30 UTC on 26 September 2011. Regions where the analysis adjustment is positive (negative) indicate that the analyzed electron density is increased (decreased) when assimilating vertical profiles from the augmented network, relative to when only assimilating observations from the COSMIC network. Comparison with the third row of Figure 5.4 ishows that in regions where the analysis adjustment tends to correct the analysis errors obtained when only assimilating the COSMIC network. For example, electron density in the southern polar region in the third row of Figure 5.4 at 16:30 UTC is underestimated (blue) and in this same region, the analysis adjustment due to augmented observations is positive. The analysis adjustment of the zonal component of neutral winds ($\mathrm{U_n}$) and molecular oxygen composition ($\mathrm{O_1}$) are shown in the second and third rows of Figure 5.8, respectively.

Equation (5.9) requires the inversion of the $\ell_A \times \ell_A$ matrix, $\left[\mathbf{I} - \mathbf{S}_L^{AA}\right]$, which may be numerically costly if ℓ_A is large. An equivalent but typically more efficient formulation for this inversion (derived in the next section) is given by

$$\left[\mathbf{I} - \mathbf{S}_L^{AA}\right]^{-1} = \mathbf{I} + (\mathbf{R}_L^F)^{-1}\mathbf{Y}_L^F\widetilde{\mathbf{P}}_L^{a(C)}(\mathbf{Y}_L^F)^{\mathrm{T}}, \tag{5.10}$$

where $\widetilde{\mathbf{P}}_L^{a(C)}$ is constructed exactly as the covariance matrix in Equation (3.32) in the main text, but using only the ℓ_C observations in $\mathbf{y}_L^C$. The computation of $\widetilde{\mathbf{P}}_L^{a(C)}$ requires the inversion of an $m \times m$ matrix, where m is the ensemble size, and so

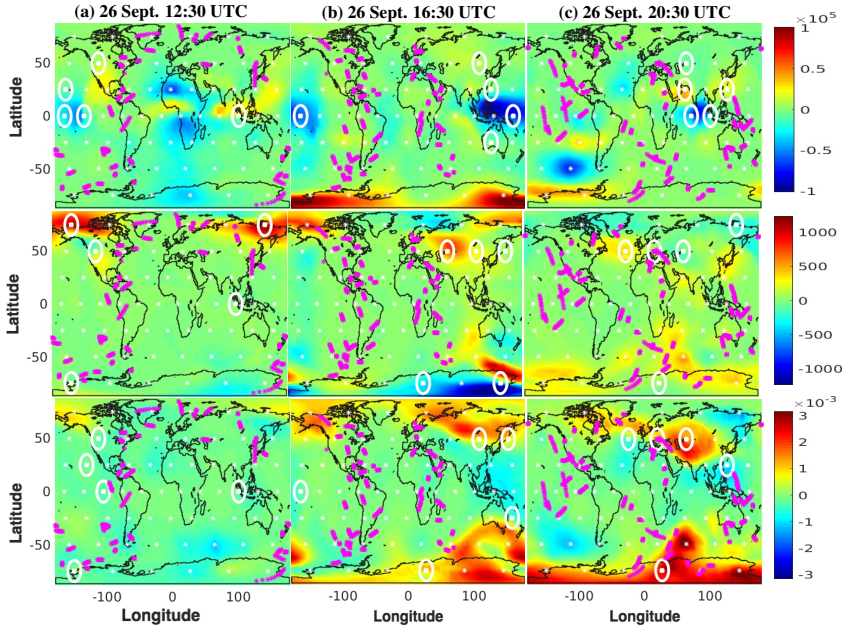

Figure 5.8: Row 1: Global maps of the analysis adjustment for electron density, in el/cm^3, averaged over 200 km to 500 km altitudes at 12:30, 16:30 and 20:30 UTC. The analysis adjustments are constructed as described in Section 4 and are defined as difference between the analyses obtained when assimilating the full network of observations (augmented and COSMIC) and the COSMIC observations only (see Equation (5.6). Rows 2-3: Analogous maps of analysis adjustments for the zonal component of the neutral winds in cm/s and molecular oxygen composition in mass mixing ratio units, at the same times. The color scale differs for each row. The locations of vertical profiles from the COSMIC and augmented observing networks are marked in magenta and white, respectively. The top 5 most influential augmented vertical profiles for each respective variable are denoted by the thick white circles.

this approach is more efficient in computing equation (5.9) if the ensemble size is smaller than the number of observations being excluded ($m < \ell_A$), which can be expected in an operational setting. In general, computing the analysis adjustment using equation (5.9) is more computationally efficient than computing $\bar{\mathbf{x}}_L^{a(F)}$ and $\bar{\mathbf{x}}_L^{a(C)}$ explicitly, and then finding the difference.

5.6 Discussion

A targeted observation strategy based on the influence matrix diagnostic is proposed for ionospheric data assimilation with the LETKF scheme. A forecast ob-

tained when assimilating a set of synthetic vertical profiles that represent those of the COSMIC network is first considered. Using this targeted observation strategy, the observation vector is augmented with additional vertical profiles whose locations are optimally chosen to mitigate forecast errors during the geomagnetic storm of 26 September 2011. The augmented observations come from a network of 62 globally distributed vertical profiles (shown by the white dots in Figure 5.6(a)) and the contribution each of these vertical profiles would have if they were assimilated is quantified with the influence matrix diagnostic. The five augmented vertical profiles with the largest influence on the global analysis, according to the strategy described in Section 5.3.2, are then selected to be included in the assimilation procedure. The targeted observation strategy is used at each analysis time from 11:30 UTC to 23:30 UTC on 26 September 2011, which corresponds to the onset and main phase of the geomagnetic storm.

This chapter develops a strategy to identify regions where additional observations may yield the largest forecast improvement. This diagnostic tool may help identify regions that should be observed in future extreme events as the ionospheric observing network expands. However, there are also practical aspects that should be considered, such as ease of observability and quality of the observations, that should be considered for future work. In operational settings, the deployment of additional observations in near-real time should also be addressed, but this also restricts the type and location of these observations.

In the augmented network chosen for this chapter, the longitude distance between vertical profiles ranges from 30° at the equator to 60° in the polar regions, and their latitude distance is 25°. With the $5° \times 5°$ horizontal grid resolution used in the TIEGCM, at least one vertical profile is located within the local region of every grid point during each analysis calculation. A higher resolution augmented observation

network may be beneficial since it would offer more choices from which the most influential vertical profiles are chosen. In future work, the benefits of a higher resolution network of augmented vertical profiles with a higher resolution version of the TIEGCM will be examined.

In the numerical experiments presented, augmented vertical profiles located in the nighttime and mid-to high-latitude regions have the most influence on the forecast, especially in regions of sparse COSMIC data coverage. The benefit of assimilating the five most influential vertical profiles is evaluated with the RMSE of analyzed electron density within 600 km regions centered around the selected vertical profiles. Compared to when no vertical profiles from the augmented network are assimilated, analyzed electron density RMSE values are reduced on average by about 65% when using five vertical profiles selected to target electron density, whereas using five randomly selected vertical profiles from the augmented network reduces the RMSE of analyzed electron density by about 15% on average. Similarly, selecting five augmented vertical profiles to target the zonal component of neutral winds (U_n) reduce RMSE values of analyzed U_n on average by about 25% during the transition from quiet to disturbed geomagnetic conditions (11:30 UTC to 16:30 UTC), whereas using five randomly selected vertical profiles reduces the RMSE values of analyzed U_n by about 2% on average during the same time period.

These results demonstrate that the proposed targeted observation strategy can optimally select locations for augmented vertical profiles that yield the most improvement in electron density estimates, as well as other state variables. The scope of this paper is to propose a targeting strategy based on the observation influence and demonstrate its application to reduce forecast errors for electron density and U_n estimates. Selecting targeted observations to improve estimates in other state variables may also be beneficial. In future work, approaches were subsets of the available

augmented vertical profiles are used to simultaneously target groups of different state variables, rather than using all augmented vertical profiles to target a single state variable, will be explored.

In this study, it is assumed that the temporal evolution of the parameters associated with solar conditions and geomagnetic activity is known throughout the geomagnetic storm. The estimation of these parameters is a challenging and important component of ionospheric data assimilation that must be accounted for in an operational setting, especially during extreme events. The proposed targeting strategy may be used to identify locations of augmented vertical profiles that would be the most influential to the estimation of these parameters and the benefits of assimilating these targeted vertical profiles may then be quantified. In future work, the data assimilation system and targeted observation strategy will be expanded to include a parameter estimation component.

Chapter 6

DATA ASSIMILATION FOR IONOSPHERIC SPACE-WEATHER FORECASTING IN THE PRESENCE OF MODEL BIAS

This chapter presents observing system experiments in which a model bias estimation strategy for ionospheric space-weather forecasting is proposed. The work presented in this chapter has been submitted for publication in the *Journal of Geophysical Research: Space Physics*.

6.1 Introduction

Modeling efforts in the ionosphere require the continual estimation of its time-dependent external drivers. The benefits of modeling the ionosphere and thermosphere, and their associated electrodynamical processes, with a single general circulation model have been observed (Matsuo and Araujo-Pradere, 2011). However, other important ionospheric drivers are specified with auxiliary empirical models. The ionosphere is strongly coupled to the magnetosphere in high-latitude regions, and the related processes associated with the electric field patterns and energy precipitation for different geomagnetic conditions are represented empirically. (Edwin *et al.*, 2011). The ionospheric response to solar activity is also typically modeled empirically. The empirical representation of these dynamical drivers puts great importance on the parameterizations and their respective inputs used to drive the ionosphere, especially during extreme events. Although the quality of parametric models used to represent these key dynamics is an important consideration, the scope of this paper focuses on the optimization of an already established configuration of parameterized inputs.

Inputs to solar and geomagnetic parameterizations are typically measured directly and are physically meaningful, but it is not always clear if these are the optimal indices to drive an ionospheric model for a selected time period; they often need to be adjusted, especially during extreme events. Although solar activity does not typically vary significantly on a daily time-scale, extreme ionospheric space-weather events are usually related to powerful solar events that are difficult to predict, and the associated solar inputs must be specified accordingly. Geomagnetic conditions are also difficult to represent during moderate to strong perturbation periods, and geomagnetic indices have been shown to have uncertainties, ranging from limited station observability coverage to limited representation of associated physical mechanisms (Xu, 2008). Additionally, geomagnetic indices are issued over 3 hour intervals, which is longer than the typical time-scales of ionospheric responses during extreme events (Du *et al.*, 2010). The development of improved indicators of geomagnetic activity, such as the methods described in (Wing *et al.*, 2005) is an ongoing research effort.

Parameter estimation techniques based on the Ensemble Kalman Filter (EnKF) are a promising approach for the optimization of parameter inputs, based on the available observations of the system state. The EnKF formulation produces a low-rank estimate of the flow-dependent model state covariance matrix, which can be used to infer the state of unobserved state variables, such as the coupled thermospheric state, and of driver parameters. EnKF approaches have been shown to be effective in estimating the state of the coupled ionosphere-thermosphere system (Matsuo and Araujo-Pradere, 2011; Hsu *et al.*, 2014) and ionospheric forcing parameters (Matsuo *et al.*, 2013). An overview of parameter estimation strategies in data assimilation is presented in (Ruiz *et al.*, 2013) and additional parameter estimation strategies with Ensemble Kalman filters for non-global parameters are discussed in (Bellsky *et al.*, 2014).

Extreme events may present processes or features that are not well understood and may not be captured with existing model parameterizations in terms of magnitude or duration. For example, geomagnetic storms often have significant and persistent storm-time effects that modify the state of the ionosphere relative to quiet time conditions. See (Pedatella *et al.*, 2009) for a study of the long-lasting positive storm-time effects observed in the geomagnetic storm of 15 December 2006. (Chartier *et al.*, 2016) also discusses some storm-time electron density enhancements that are difficult to predict during the 10 September 2005 geomagnetic storm. The limited representation of storm-time effects provided by parameterizations may not capture the variability of ionospheric dynamics, which may introduce systematic model biases that can be difficult to remedy with further parameterization adjustments.

In this chapter, a strategy to compute spatially-varying corrections to compensate for model bias introduced from sub-optimal parameterized inputs during extreme events is proposed. It is assumed that solar and magnetospheric parameter configuration, which may be specified through some parameter estimation approach, is already in place. The proposed strategy treats the global distribution of model bias as a collection of bias correction parameters, which are estimated using an ensemble Kalman filter. The bias corrections are not used to adjust the model state vector directly. Instead, spatially-varying bias corrections are applied in the evaluation of the observation operator, with the intent of reducing model bias in its electron density predictions prior to the assimilation of observations. This methodology is proposed in (Baek *et al.*, 2006), where it is applied in observing system experiments with the Lorenz 40-variable model.

The application of spatially-varying bias corrections extends the degrees of freedom in the predictions made with a given parameterized configuration, which may be beneficial in capturing storm-time effects. The corrections are computed inde-

pendently of parameterized inputs, so this strategy may be applied together with direct parameter estimation. Furthermore, this approach helps provide more gentle state estimation updates due to the adjusted forecast predictions, which may help to avoid the introduction of spurious dynamical artifacts from drastic adjustments to the system state that may occur due to strong model bias presence during an extreme event. To the author's knowledge, the proposed bias correction strategy has not been applied previously to an ionospheric model with operational capabilities.

The proposed bias estimation strategy, which is designed for high-dimensional ionospheric systems that are sparsely observed, is applied in observing system experiments, where the global state of the ionosphere is estimated throughout the 26 September 2011 geomagnetic storm, in the presence of model bias resulting from sub-optimal parameterized solar and magnetospheric inputs. The local ensemble transform Kalman filter (LETKF) (Hunt *et al.*, 2007) is used to assimilate synthetic electron density vertical profiles into the Thermosphere-Ionosphere-Electrodynamics-Global-Circulation-Model (TIEGCM). The profile locations are the same as that of the Formosa/Constellation Observing System for Meteorology, Ionosphere, and Climate satellites 3 (COSMIC) (Rocken *et al.*, 2000), during 26-27 September 2011. The LETKF is a type of ensemble square root filter that computes its analysis using a low-rank estimate of the forecast covariance matrix. The analysis is computed independently grid point by grid point, by assimilating nearby observations simultaneously. The LETKF has been applied in the ionosphere with an idealized regional model (Durazo *et al.*, 2016) and for space-weather specification during an extreme events with the TIEGCM by (Durazo *et al.*, 2017). The LETKF has also been used with the Global Ionosphere-Thermosphere Model to estimate solar parameters during periods of low and high solar activity (Koller *et al.*, 2013; Godinez *et al.*, 2015).

The layout of this chapter is as follows. Section 6.2 describes the proposed bias

estimation methodology Sections 6.3 and 6.4 describe observing system experiments, in which the bias estimation approach is applied during a geomagnetic storm event. A discussion and conclusions of these results are given in Section 6.5.

6.2 Bias Estimation Methodology

Consider the application fo the LETKF in the context of a forecast model for which there is some systematic difference between its predicted state $\mathbf{u}_{n+1}^m$ of the ionosphere at time t_{n+1} and the "true" state. More precisely, the vector $\mathbf{u}_{n+1}^m$ contains all the dynamical variables associated with the model's depiction of ionospheric processes, such as electron density, thermospheric composition, etc., over the global model grid. The ionospheric model may be regarded as a collection of maps $\mathbf{M}_n$, $n = 1, 2, \ldots, N$, that yield global model state vectors of the form

$$\mathbf{u}_{n+1}^m = \mathbf{M}_n \left(\mathbf{u}_n^m \right) \tag{6.1}$$

for discrete times $t_1, t_2, \ldots, t_N$ over some interval of interest. The ionospheric model approximates a corresponding "true" set of state vectors $\mathbf{u}_n^t$ at each time and grid point. That is, the sequence

$$\mathbf{u}_{n+1}^t = \mathbf{F}_n \left(\mathbf{u}_n^t \right) \tag{6.2}$$

may be regarded as a finite-dimensional projection of the ionosphere's evolution to the grid points of the ionospheric forecast model at each time t_n. A perfect model would produce a sequence $\mathbf{u}_n^m$ that is identical to $\mathbf{u}_n^t$ for each n, provided that the initial state of the ionosphere and its drivers is the same. In practice, however, the initial ionosphere-thermosphere state is not known precisely. Additionally, the specification of solar and magnetospheric drivers, and the description of the ionosphere's reaction to their temporal variations, which produces each $\mathbf{u}_n^m$ is different than evolution operator that produces the true ionospheric states $\mathbf{u}_n^t$. Thus the vectors $\mathbf{u}_n^m$ and $\mathbf{u}_n^t$ differ, and

the objective is to develop a scheme to approximate these differences within the state space of the forecast model in the context of a data assimilation system.

The proposed approach is based on one suggested by (Baek *et al.*, 2006) (bias model II), in which it is assumed that the dynamics of the ionospheric model evolve on a different attractor from the "true" dynamics of the ionosphere. In such a case, substituting $\mathbf{u}_n^t$ (or a vector close to it) into the model $\mathbf{M}_n$ may excite spurious dynamics or inconsistencies between thermospheric and ionospheric states, possibly because the dynamical drivers represented by the model differ from those of the truth. For example, strong systematic biases may be introduced due differences in solar and magnetospheric drivers during extreme events.

To avoid this situation, the evolution of the state vector on the model attractor is regarded as a time-dependent translation of corresponding points on the true attractor. Rather than attempting to substitute the truth into the forecast model, the discrepancy between the truth and forecast state vectors is compensated for within the data assimilation system, whose objective now is to find the model state vector that yields the best forecast. For this purpose, define the model bias at each time t_n as

$$\mathbf{c}_{n+1} = \mathbf{F}_n(\mathbf{u}_n^t) - \mathbf{M}_n(\mathbf{u}_n^m) \tag{6.3}$$

where

$$\mathbf{u}_n^m = \mathbf{u}_n^t - \mathbf{c}_n. \tag{6.4}$$

Generally, the correction $\mathbf{c}_{n+1}$ depends on $\mathbf{c}_n$ as well as on $\mathbf{F}_n$ and $\mathbf{M}_n$. To make the scheme approach practicable within a data assimilation system, and following

the work of (Baek *et al.*, 2006), suppose that the corrections $\mathbf{c}_n$ evolve according to another dynamical system and implement the following augmented model:

$$\mathbf{u}^b_{n+1} = \mathbf{M}_n(\mathbf{u}^a_n) \tag{6.5}$$

$$\mathbf{c}^b_{n+1} = \mathbf{G}_n(\mathbf{u}^a_n, \mathbf{c}^a_n). \tag{6.6}$$

Here the subscript b refers to the "background" state vector (i.e., the forecast) and a to the "analyzed" state vector (i.e., the output of the data assimilation system).

The data assimilation procedure described in Section 3.4.1 proceeds as before, but it begins with an ensemble of augmented state vectors of the form $\left[\left(\mathbf{u}_n^{b(j)}\right)^{\mathrm{T}}\left(\mathbf{c}_n^{b(j)}\right)^{\mathrm{T}}\right]^{\mathrm{T}}$, where j indexes each ensemble member. In lieu of Equation (3.28), the observation operator at time t_n becomes

$$\mathbf{y}_L^{b(j)} = \mathbf{H}_L\left(\mathbf{u}_n^{b(j)} + \mathbf{c}_n^{b(j)}\right). \tag{6.7}$$

In other words, the model error is compensated for by adding a correction term to the model state vector whenever a comparison to observations is made, but the model state vector is not modified. The data assimilation procedure produces an augmented analysis ensemble of the form $\left[\left(\mathbf{u}_n^{a(j)}\right)^{\mathrm{T}}\left(\mathbf{c}_n^{a(j)}\right)^{\mathrm{T}}\right]^{\mathrm{T}}$, which is used as the initial condition in the subsequent forecasting step. Equations (6.5)-(6.6) are then applied to produce the augmented state, $\left[\left(\mathbf{u}_{n+1}^{b(j)}\right)^{\mathrm{T}}\left(\mathbf{c}_{n+1}^{b(j)}\right)^{\mathrm{T}}\right]^{\mathrm{T}}$, for the next analysis step.

The algorithm is complete once a choice of the bias evolution operators $\mathbf{G}_n$ is made. In this study, two simple choices are considered. Section 6.3 describes the results when $\mathbf{G}_n$ is persistence, i.e., $\mathbf{c}_{n+1} = \mathbf{c}_n$. Section 6.4 describes the results when $\mathbf{G}_n$ implements a exponential growth-relaxation correction term.

In principle, the bias correction procedure doubles the number of state variables that must be estimated from the same set of observations, which increases the variance

in the analyzed fields. On the other hand, not all model biases contribute equally to the forecast error, and it may suffice to apply the bias correction procedure only to some of the components of the model state vector; that is, the components of $\mathbf{c}_n^b$ are taken to be zero except for those corresponding to fields of greatest forecast interest. In the numerical experiments described in this chapter, the bias estimation procedure is applied only to the electron density component of the TIEGCM.

6.3 Numerical Experiment 1: Persistent Bias Evolution

In the following numerical experiments, the LETKF scheme is used to assimilate synthetic electron density vertical profiles into the TIEGCM. The observation network consists of synthetic electron density vertical profiles, whose locations are given by the COSMIC network for 26-27 September 2011. The implementation of this data assimilation system is the same as the one used in Chapter 5 and in (Durazo *et al.*, 2017).

The main source of ionization in the ionosphere, and thus one of its primary dynamical drivers, is the absorption of solar radiation in the thermosphere, primarily by O_1, O_2 and N_2 neutral gases. The TIEGCM represents effects of solar irradiance and its variability through auxiliary empirical models. The default solar input model, which is used in the TIEGCM configuration of this study, is the EUVAC irradiance model (Richards *et al.*, 1994). The EUVAC model specifies important ionospheric processes related to solar activity, such as ionization and dissociation rates, and heating of neutral gases, ions and electrons. The empirical representation of solar irradiance is parameterized with the $F_{10.7}$ index, which is a daily measurement radio flux at 10.7 cm wavelength, and its 81-day average. The TIEGCM offers the option to use measured spectral irradiance as the solar input (available at http://lasp.colorado.edu/see), but those options are not considered in this study. Solar proxy models parameterized with

the $F_{10.7}$ index have been widely used in other recent ionospheric data assimilation studies such as the ones in (Lee *et al.*, 2012) and (Matsuo and Araujo-Pradere, 2011).

Another key driver of ionospheric dynamics is geomagnetic activity. The ionosphere is strongly coupled to the magnetosphere in high-latitude regions but is typically approximated with auxiliary empirical models due its complexity. The default magnetospheric input, which is used in the configuration of the TIEGCM, is the Heelis model (Heelis *et al.*, 1982). The Heelis model specifies important high-latitude processes in the ionosphere, such as electric field patterns and auroral energy inputs, for different levels of geomagnetic disturbance. The TIEGCM allows for other magnetospheric inputs, such as the Weimer model (Weimer, 2005) and AMPERE model (Anderson *et al.*, 2014), but they are not considered in this study.

The Heelis model is parameterized with the Kp index, which is a widely used index derived from the horizontal component of geomagnetic field disturbances. Kp indices are provided every 3 hours and range in value from 0 to 9, to describe geomagnetic conditions ranging from quiet to extremely disturbed. In particular, the Kp index is used to calculate high-latitude auroral precipitation (Hp), which specifies high-latitude energy inputs, and the cross-tail potential (Cp), which specifies ion convection patterns in the polar regions. Historical records of $F_{10.7}$ and Kp indices are provided by the National Oceanic and Atmospheric Administration (NOAA).

6.3.1 Experiment Set Up

The numerical experiments presented in this paper simulate a scenario in which an ionospheric forecast is driven with sub-optimal specification of magnetospheric and solar inputs during a period of high geomagnetic disturbance. Geomagnetic indices are difficult to measure accurately during periods of magnetospheric disturbance, and solar irradiance patterns can deviate significantly from proxy models during extreme

solar events. Storm time effects can be difficult to model and predict in sign and magnitude, primarily due to uncertainties in these key inputs, as well as inherent limitations of the parameterized representations. For example, the study by (Pedatella *et al.*, 2009) discusses the significant enhancement in F-layer electron density peak and altitude during the geomagnetic storm of 15 December 2006. In particular, these storm time effects were observed to be long-lasting after the main phase of this geomagnetic storm. Observational analyses on the initial phase of this geomagnetic storm are presented by (Lei *et al.*, 2008b) and a study about observed traveling ionospheric disturbances are discussed in (Lei *et al.*, 2008a).

The goal in the following numerical experiments is to evaluate the benefit of applying the proposed bias correction strategy to reduce model bias in electron density predictions during a period of geomagnetic disturbance. these tests are presented with observing system experiments, in which the true state of the ionosphere-thermosphere system is taken to be given by a TIEGCM simulation. This "perfect model" assumption is made to isolate the performance of the bias estimation strategy from other issues that may arise in data assimilation problems, such as gross misrepresentation from model dynamics or complex error distribution in the assimilated observations. The focus of the experiments is on the 26 September 2011 geomagnetic storm event.

An ensemble of 40 forecasts is generated with normally distributed parameter values of $F_{10.7}$, Hp and Cp. Each parameter is centered around the respective index values published for 26-27 September 2011. The standard deviation for the $F_{10.7}$ distribution is its 21-day standard deviation during the spin-up period, which is taken to be to be from 5 to 25 September 2011, and the standard deviation for Kp is ± 1.0 units of Kp. The respective standard deviations for $F_{10.7}$, Cp and Hp are 13×10^{-22} W/m^{2}Hz, 8.4 kV and 7.2 GW, respectively. Figures 6.1(a)-(c) summarize the temporal evolution of the ensemble of forecast parameters (pink) and the ensemble

mean (red) for each respective parameter. The horizontal axis for each figure is time in hours, starting from 00:30 UTC on 26 September 2011 to 23:30 UTC on 27 September 2011.

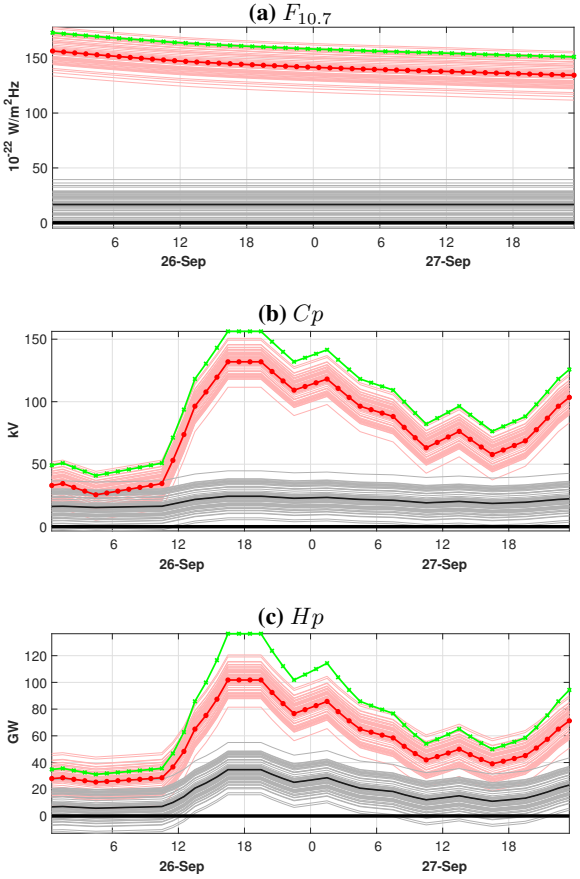

Figure 6.1: (a) Time series of $F_{10.7}$ indices in solar flux units $\left(10^{-22}\text{W}/(\text{m}^2\text{Hz})\right)$, (b) cross-tail potential (Cp) in kV, and (c) hemispheric power (Hp) in GW. The horizontal axis is in hours, starting at 00:30 UTC on 26 September and ending at 23:30 UTC on 27 September 2011. The parameters used to drive the truth are shown in green, the parameters used to drive the ensemble of forecasts are shown in pink and their respective ensemble means are shown in red. The deviation of each parameter from those used in the control simulation is shown in gray, and the ensemble mean of the deviations is shown in black.

The control simulation is driven with forcing parameters that have the same temporal evolution as the published indices but are shifted as shown by the green curves in Figure 6.1(a)-(c). This shift is introduced to simulate the scenario where the solar and magnetospheric inputs used to drive the forecast are a misspecification relative to

the "true" state of the ionosphere-thermosphere system. The imposed bias on $F_{10.7}$, shown in Figure 6.1(a) for the ensemble (gray) and the ensemble mean (black), is 10×10^{-22} W/m^{2}Hz and is kept constant throughout the simulation. The bias in the magnetospheric inputs is introduced by adding a shift of 1.0 units in the Kp index. The corresponding bias on Cp and Hp values is shown in Figures 6.1(b)-(c). Although the bias in the Kp index is fixed, the resulting bias in the Cp and Hp indices is temporally varying. Prior to the onset of the geomagnetic storm, which is about 12:30 UTC on 26 September, the bias in the magnetospheric inputs is relatively small and constant but increases considerably over the next 6 hours. The strongest bias occurs during the main phase of the geomagnetic storm, which takes place at about 16:30–19:30 UTC on 26 September. As geomagnetic conditions relax over the next 12 hours, the bias in the magnetospheric inputs also decreases accordingly.

6.3.2 Results

Model Bias Distribution

In the following section, it is tested how well the proposed bias estimation strategy, assuming that the bias evolution is persistence can capture the spatiotemporal evolution of model bias in the electron density field, resulting from the misspecification of the solar and magnetospheric model inputs. The misspecification of the parameterized input is assumed to be constant for solar inputs and time-varying for the magnetospheric inputs as shown in Figure 6.1.

At a given time t_n, denote the electron density component of the background state vectors, $\{\mathbf{u}_n^{b(j)}\}_{j=1}^k$, as $\{\mathbf{e}_n^{b(j)}\}_{j=1}^k$ and its ensemble mean as $\bar{\mathbf{e}}_n^{b(j)}$. Similarly, denote the "true "electron density component from the control simulation as $\mathbf{e}_n^t$. Define the

electron density bias to be given by the deviation of the forecast mean from the truth:

$$\mathbf{b}_n = \bar{\mathbf{e}}_n^b - \mathbf{e}_n^t. \tag{6.8}$$

Figure 6.2(a) shows the global distribution of electron density bias, in units of el/cm^3, at a fixed altitude of 375 km, which is about the altitude of the F-layer over the equatorial day-time ionosphere during this time period. The selected times show the distribution of electron density bias on 26 September 2011 before the major onset of geomagnetic disturbances (12:30 UTC), during the main phase of the geomagnetic storm (15:30 UTC, 18:30 UTC), and as geomagnetic disturbances begin to relax (21:30 UTC). The locations of the observed electron density profiles available for assimilation at each respective time are shown with the magenta markings.

Also shown in Figure 6.2(a) are 4 geographical regions over which the model bias is examined. The boundaries for the Northern and Southern high-latitude regions (R_1 and R_2) are located at the 30° magnetic co-latitudes, since processes associated with magnetospheric inputs are calculated explicitly by the TIEGCM below this co-latitude. Above the 20° magnetic co-latitudes, magnetospheric inputs are imposed directly with the Heelis model. Between these co-latitude bands, a linear combination of the imposed parametric and model solutions is used. See the TIEGCM model description for more information (http://www.hao.ucar.edu/modeling/tgcm/). The low-to-mid altitude regions are also separated in the day-time (R_3) from the night-time (R_4).

A positive (negative) sign indicates that the background electron density overestimates (underestimates) the true electron density. Prior to the main phase of the geomagnetic storm, the most notable bias structure is in the day-time mid-latitude regions (R_3), where there is a negative bias. As geomagnetic conditions become increasingly disturbed, the misspecification in the magnetospheric input grows, re-

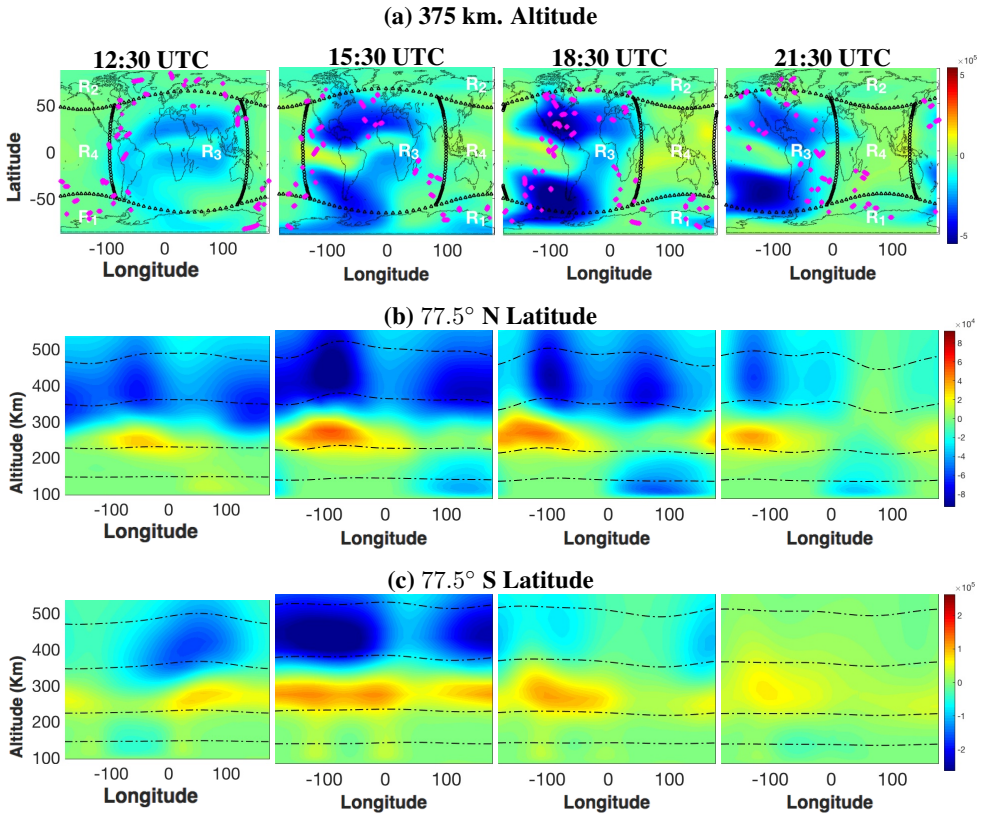

Figure 6.2: (a) Global maps of electron density bias in units of el/m^3, at a fixed 375 km altitude at the indicated times on 26 September 2011. The electron density bias is computed as deviation from the truth of the ensemble mean of the forecast ($\bar{\mathbf{e}}^b - \mathbf{e}^t$). The black curves denote the boundaries of the geographical regions, labeled in white, that partition the domain horizontally. The locations of the COSMIC vertical profiles at each time are shown with the magenta markings. (b) Vertical structure of electron density bias at a fixed 77.5°N latitude at 11:30, 15:30, 19:30 and 23:30 UTC on 26 September 2011. The vertical partitioning of the domain is denoted with the black curves, which correspond to pressure levels -2.5, 0, 2.5 and 5. (c) Analogous plots of the electron density bias vertical structure at a fixed 77.5°S latitude.

sulting in the formation of a significant negative bias structure over the Southern Polar region (R_1). The negative bias structure continues to grow over regions R_1 and R_3 over the next few hours throughout the rest of the main storm phase. There is also a formation of negative bias in the Northern polar region (R_2) and a positive bias forming over the night-time low-to-mid latitude region (R_4). After the main phase (21:30 UTC), the negative bias in the polar regions is largely diminished but

the bias over the day-time mid-latitudes remains relatively constant. This suggests that the misspecification of $F_{10.7}$, which is held constant throughout the simulation has its largest effect over the low-to-mid latitude regions. Over this time period, the temporal evolution of electron density bias has a pronounced westward drift of approximately 15°/h, so the persistent dynamics of the bias evolution operator, $\mathbf{G}$, provide a reasonable representation of the model error.

Figure 6.2(b) shows the vertical structure of the electron density bias at the $77.5N^{\circ}$ geographical latitude at the same times. The black horizontal curves denote model pressure levels -2.5, 0, 2.5 and 5, which correspond to altitudes of about 145 km, 215 km, 325 km and 460 km, respectively. Throughout the main phase of the geomagnetic storm, the most notable bias is negative, occurring above 250 km and 280 km altitudes in the day- and night-time regions, respectively, which is about the altitude of the lower portion of the F-layer. Figure 6.2(c) shows the vertical structure of the electron density bias at the 77.5°S geographical latitude. Similarly, the most notable bias is observed above the 280 km altitude. However, there is pronounced positive bias at about a 250 km altitude, which is located slightly below the F-layer peak density.

Region-Averaged Bias Estimates

Following the procedure described in Section 6.2, the ensemble of global bias correction vectors at a given time t_n, $\{\mathbf{c}_n^{b(j)}\}_{j=1}^k$, is updated to form an ensemble of analyzed global bias correction vectors, $\{\mathbf{c}_n^{a(j)}\}_{j=1}^k$. Generally, each $\mathbf{c}_n^{b(j)}$ is a $md \times 1$ vector, where m is the number of grid points and d is the number of state variables being analyzed. As discussed at the end of Section 6.2, bias corrections are considered only for the electron density field. For simplicity in notation, regard each $\mathbf{c}_n^{b(j)}$ to be an $m \times 1$ vector corresponding to the electron density field corrections only.

The forecast/analysis cycle is initialized at time t_0 with electron density corrections bias corrections at each grid point is given by

$$\mathbf{c}_n^{b(j)} = \frac{1}{4}\left(\mathbf{e}_n^{b(j)} - \bar{\mathbf{e}}_n^{b}\right), \tag{6.9}$$

which has ensemble mean 0 and reflects the spatial correlations of the electron density field. The procedure described in Section 6.2 is then carried out for subsequent analysis steps to produce spatially- and time-varying bias estimates throughout the domain that are evolved according to $\mathbf{G}_n$, which in this section is taken to be the identity map. For each grid point L, variance inflation is applied to the bias correction component so that its ensemble variance is at least 20% of the electron density component at L.

The model state vector is analyzed without the bias corrections for the first few analysis cycles after the forecast spin-up, so that errors associated with the unadjusted thermospheric state are reduced and the electron density bias is primarily dependent on the misspecification of the solar and magnetospheric inputs. Eight analysis cycles are computed, from 16:30 to 23:30 UTC on 25 September 2011, before using the bias correction strategy for the first time at 00:30 UTC on 26 September 2011.

To evaluate the skill of the bias estimation strategy, the model bias estimates, averaged over the 4 geographical regions, shown in Figure 6.2(a) are examined. Consider a fixed bias region, R, where $R = R_1, \ldots, R_4$ as defined above. Denote the electron density component of the jth forecast, $\mathbf{e}_L^{b(j)}$, averaged over all grid points $L \in R$, as $D_R^{b(j)} = N_R^{-1} \sum_{L \in R} \mathbf{e}_L^{b(j)}$, where N_R is the number of grid points in R. Similarly, form the region-averaged electron density of the control simulation, $D_R^t = N_R^{-1} \sum_{L \in R} \mathbf{e}_L^t$. The region-averaged deviation from the truth for the jth forecast is given by

$$B_R^{b(j)} = D_R^{b(j)} - D_R^t, \quad j = 1, \ldots, k \tag{6.10}$$

which yields an ensemble of regionally-averaged model bias estimates, whose ensemble

mean, $\bar{B}_R^b = \frac{1}{k}\sum_{j=0}^{k} B_R^{b(j)}$ is the most likely state of the regionally averaged model bias.

Figure 6.3(a) summarizes the temporal evolution of the region-averaged electron density bias in the Southern polar region (R_1) averaged between 1.0 and 5.5 pressure levels, which correspond altitudes of about ∼260 km to ∼485 km. The horizontal axis is time in hours, starting at 00:30 UTC on 26 September, 2011 and ending at 23:30 UTC on 27 September 2011, and the vertical axis is electron density in units of el/cm^3. The ensemble of spatially-averaged background electron densities, $\{D_R^{b(j)}\}_{j=1}^{k}$ and their ensemble mean, $\bar{D}_R^b$, are denoted with thin pink curves, and thick red curve, respectively. The analogous region-averaged electron density for the control simulation, D_R^t, is given by the green curve. The ensemble of bias correction parameters, $\{C_R^{b(j)}\}_{j=1}^{k}$ and their ensemble mean, $\bar{C}_R^b$, is shown with the cyan and thick blue curves, respectively. For direct comparison, the negative of the region-averaged forecast deviations, $\{B_R^{b(j)}\}_{j=1}^{k}$, is also shown with the thin gray curves and their respective ensemble mean, $\bar{B}_R^b$, is given by the thick black curve.

The region-averaged electron density for the control simulation varies considerably within region R_1, particularly during the period of main geomagnetic disturbance, where the electron density increases sharply between 12:30 UTC and 16:30 UTC on 26 September. After the main storm phase, electron density content drops considerably and remains relatively constant throughout the rest of the simulation. The region-averaged electron density trajectories for the forecast evolve similarly. Prior to the onset of geomagnetic disturbances, the electron density bias is relatively small, but it rises sharply during the main phase, peaking at around 16:30 UTC. The electron density bias is maintained until about 19:30 UTC, where it begins to drop considerably and remains relatively constant after 23:30 UTC on 26 September, until there is a small resurgence towards the end of the simulation.

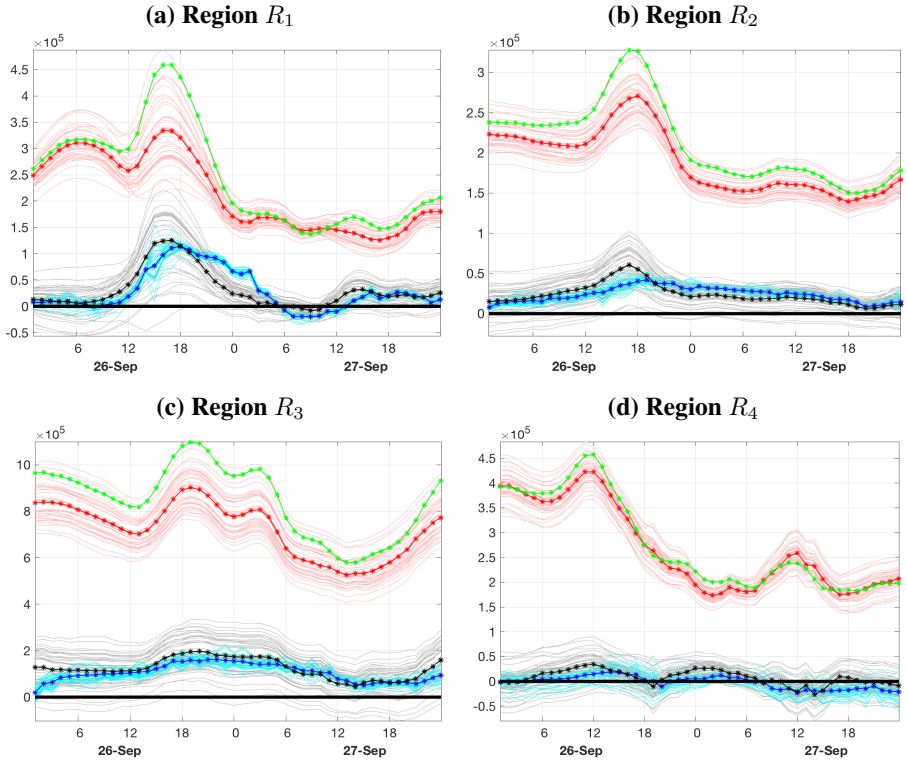

Figure 6.3: (a) Time series of the true electron density (green), in units of el/cm^3, averaged over region R_1 (as defined in Figure 6.2(a)), at the indicated altitudes. The electron density, averaged over the same region, for the background ensemble (pink) and its ensemble mean (red) are also shown. The deviation of the background ensemble and its ensemble mean from the truth are given by the thin gray and thick black curves, respectively. The time series of the ensemble of bias parameters and their ensemble mean is shown by the thin cyan and the thick blue curves, respectively. The horizontal axis is time in hours, starting at 00:30 UTC on 26 September 2011 and ending at 23:30 UTC on 27 September 2011. (b)-(d) Analogous time series of electron density, averaged over region R_2 (Northern polar region), region R_3 (day-time mid-latitudes), and region R_4 (night-time mid-latitudes), respectively.

Prior to the onset of geomagnetic disturbances (12:30 UTC), the bias strategy correctly detects the near zero positive bias and continually adjusts its estimates as model bias starts to increase during the transition into the main phase of the geomagnetic storm around 12:30 UTC. Although the temporal variations in the bias estimates evolve in a similar manner as the model bias, the temporal variations are not captured during this transitionary periods where the model bias increases and

decreases sharply during the initial and relaxation phases, respectively. Most notably, the bias corrections adjust too slowly and underestimate the model bias peak during the initial storm phase. After the main phase, the model bias decreases sharply and although the bias corrections adjust, they do so too slowly and overestimate the model considerably after 20:30 UTC for the next 8 hours.

The region-averaged bias over the day-time low-to-mid-latitude region (R_3), shown in Figure 6.3(c), exhibits a relatively larger bias compared to the high-latitude regions but remains relatively constant throughout the simulation, since the misspecification of $F_{10.7}$, which is held constant throughout the simulation, is the primary driver of bias over this region. The bias correction parameters approach the model bias relatively quickly and follow its temporal variations well, particularly during the relaxation phase where the model bias decreases gradually. The bias in the night-time region (R_4), shown in Figure 6.3(d), is considerably smaller than the bias in its surrounding regions. Its temporal evolution is relatively constant throughout the simulation and is well represented with the bias corrections.

Validation of Bias Correction

To validate the bias correction strategy, 1-hour predictions of electron density are compared before and after the bias corrections are applied. Since the bias corrections are applied only during the evaluation of the forward operator, the benefit of the bias corrections is evaluated at the observed locations. In particular, the prediction RMS error averaged over all vertical profiles is compared within each of the geographical regions:

$$RMSE = \sqrt{\frac{\sum_{i \in R} \left(\bar{\mathbf{y}}_i^b - \mathbf{y}_i^t \right)^2}{\ell_R}}, \tag{6.11}$$

where $\mathbf{y}_i^t$ and $\bar{\mathbf{y}}_i^b$ are the true and predicted electron densities at the ith observation location. The RMSE in Equation (6.11) is averaged over the set of ℓ_R observations located in the given region, R.

Figure 6.4(a) shows the RMSE time series of 1-hour predictions over region R_1, where the blue and red curves are computed with Equation (6.11), by taking $\bar{\mathbf{y}}_i^b$ to be the forecast predictions before and after the bias corrections are applied, respectively. Prior to the main phase of the geomagnetic storm (12:30 UTC), there is little benefit in applying the bias corrections due to the relatively small bias present during this time period. Considerable benefits in using the bias correction are observed between 15:30 UTC and 19:30 UTC, where the RMSE of uncorrected predictions reaches its peak. Due to the misrepresentation of the model bias over region R_1 during the relaxation phase, as seen in Figure 6.3(a), the bias corrections do not yield benefits in 1-h predictions and actually increase the forecast RMSE considerably throughout the relaxation storm phase. In the next section, it is shown that using a growth-relaxation model for the bias evolution significantly improves the temporal variations of the model bias during this time period.

Figure 6.4(b) shows the analogous RMSE time series for region R_2, which is similar to that of region R_1, in that there is a peak in RMSE during the main phase of the geomagnetic storm, although it is smaller in magnitude. The improvements prior to the onset of geomagnetic disturbances are relatively small, but there is significant improvement of about 30% during the main phase. During the relaxation phase, the RMSE is about the same with and without the bias correction, so the bias corrections do not yield much benefit during this time period.

Figure 6.4(c) shows analogous RMSE time series for the low-to-mid-latitude day-time region (R_3). The RMSE time series for this region has a less pronounced peak during the main phase and the improvement due to the applied bias corrections is

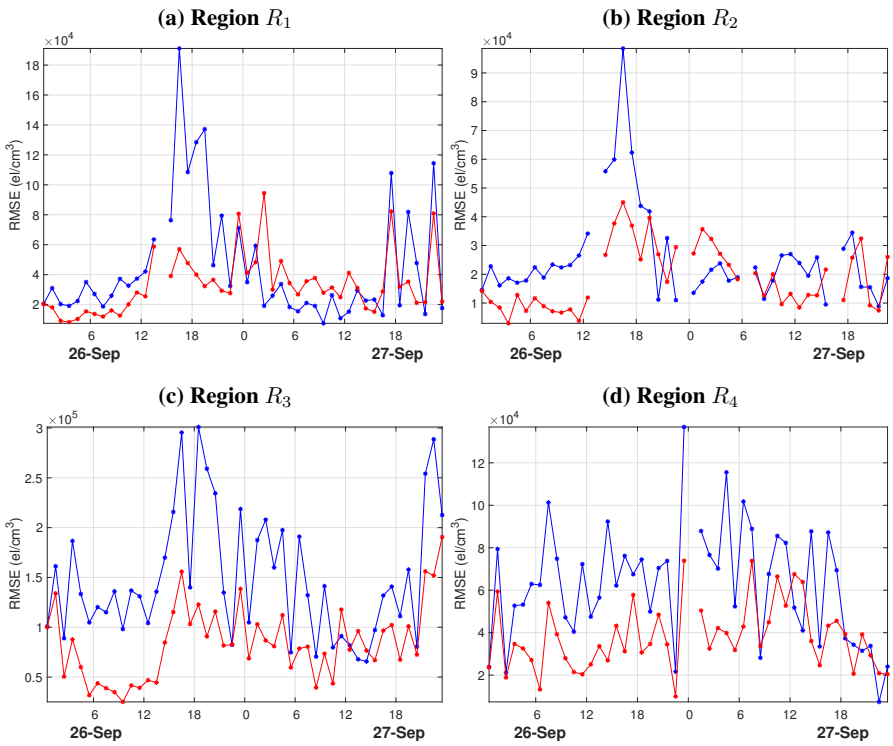

Figure 6.4: (a) Time series of the root mean squared error (RMSE) of 1-hour forecasted electron density predictions, in el/cm^3, averaged over all the observed locations in region R_1 (as defined in Figure 6.2). The RMSE values are shown for predictions before (blue) and after (red) the bias correction is applied. (b)-(d) Analogous time series of region-averaged RMSE values and ratios for regions R_2, R_3 and R_4. In all figures, missing values indicate that there are no observations in that region during that time. All time series begin at 00:30 UTC on 26 September 2011 and end at 00:30 UTC on 28 September 2011, with 1 hour intervals.

consistent throughout the simulation, although there are a few time periods where the bias corrections offer little to no improvement near the end of the simulation. The RMSE time series for the night-time region (R_4) is shown in Figure 6.4(d). In this region, the bias is also is relatively constant and the benefits of applying the bias corrections are consistent.

6.4 Numerical Experiment 2: Bias Estimation With Time-Varying Evolution Operator

The numerical experiments presented in Section 6.3 show that the bias correction strategy considerably improves 1-hour forecast forecasts throughout the domain, particularly during the main phase of the geomagnetic storm. However, some limitations of the strategy are observed during transitionary periods of the geomagnetic storm, where model bias undergoes relatively fast temporal variations. Most notably, the high-latitude bias corrections underestimate the sharp increase in model bias during the initial storm phase and overestimate the model bias as it sharply decreases during the relaxation storm phase. The misrepresentation of model bias during these time periods is primarily due to the choice of bias evolution operator, $\mathbf{G}_n$, which assumes the model bias remains constant during each forecasting step. Consequently, the predicted bias corrections partially diverge from the true state and yield inadequate background bias correction estimates over high-latitude regions. Due to the sparsity of observations, the bias estimates computed during each analysis step are not sufficiently adjusted to capture the temporal variations of the model bias.

To improve the estimation of the temporal model bias variations, a growth-relaxation model is implemented for $\mathbf{G}_n$ to propagate the model bias during the transitionary periods of the geomagnetic storm. In the following section, the numerical experiments presented in Section 6.3 are repeated with a growth-relaxation model for $\mathbf{G}$, as is now described.

6.4.1 Bias Propagation

The proposed growth-relaxation model for the bias evolution is of the form

$$\mathbf{G}_n(\mathbf{c}_n^a) = e^{\lambda_n}\mathbf{c}_n^a \tag{6.12}$$

to describe the temporal evolution of model bias estimates at each time t_n, such that the bias estimate for the jth forecast is given by $\mathbf{c}_{n+1}^{b(j)} = \mathbf{G}_n(\mathbf{c}_n^{a(j)})$. The parameter λ_n may vary with time t_n to describe the rate of model bias growth or decay, depending on the storm phase. With this choice of $\mathbf{G}_n$, the model bias estimate is evolved autonomously, in that it only depends only on its current state, $\mathbf{c}_n^{a(j)}$, and not on the state of the model, $\mathbf{u}_n^{b(j)}$.

Due to the different mechanisms driving the model bias in the ionosphere, suitable values for λ_n may change depending on location and altitude in the domain. To provide the necessary flexibility in the representation of the model bias evolution, λ_n is allowed to vary spatially, and denote its value at the Lth grid point as $\lambda_{n(L)}$. The component of $\mathbf{G}_n(\mathbf{c}_n^a)$ at the grid point L is given by $e^{\lambda_{n(L)}} c_{n(L)}^a$, where $c_{n(L)}^a$ is the Lth component of the bias vector $\mathbf{c}_n^a$.

It may be impractical to find optimal values of $\lambda_{n(L)}$ for each grid point due to the possibly complex mechanisms that drive the bias evolution. However, within suitably chosen geographical regions, the model bias evolution may have similar spatiotemporal variations that may be adequately represented with a simple model. To this end consider a time-varying parameter $\lambda_{n(R)}$, which governs the change in model bias from one assimilation time point to the next, within each of the geographic regions R, that were defined in Section 6.3.2. The value $\lambda_{n(L)} = \lambda_{n(R)}$ is set for all the grid points $L \in R$ at each fixed pressure level on the model grid, to account for the vertical structure of the electron density bias.

Specific values for $\lambda_{n(R)}$ are obtained as follows. Figure 6.5(a) shows a time series (blue dots) of electron density RMSE averaged over all grid points $L \in R_1$ at pressure level 2.0 ($\sim$375 km altitude). To obtain suitable values for $\lambda_{n(R)}$ at this pressure level during the initial phase of the storm, an exponential curve is fit (using the *fit* function in MATLAB) to the RMSE time series from 10:30 UTC to 13:30 UTC on

26 September, as shown by the red curves in Figure 6.5(a). The growth factor used in this exponential fit is used as the value for $\lambda_{n(R_1)}$ at pressure level 2.0 for the duration indicated by the curve (10:30 UTC to 13:30 UTC). Similarly, an exponential decay curve is fit to the RMSE time series from 19:30 UTC on 26 September to 06:30 UTC on 27 September to obtain $\lambda_{n(R_1)}$ used during the relaxation storm phase. During all other times, there is no red curve, indicating that $\lambda_{n(R_1)} = 0$, so that persistent dynamics are used for those times.

A similar approach is taken to compute $\lambda_{n(R_1)}$ at each pressure level above 0.5, which corresponds to the lower portion of the F-layer. Below this pressure level, the value $\lambda_{n(R_1)} = 0$ is set, since the model bias generally does not display a strong growth or decay at these pressure levels during transitionary periods of the storm and is also much weaker in magnitude relative to the topside ionosphere.

Figure 6.5(b). shows the vertical structure of $\lambda_{n(R_1)}$ values used during the initial (left) and relaxation (right) storm periods for region R_1 above pressure levels 1.0 and 0.5, respectively. During the initial storm phase, the largest rate of growth is at about pressure level 3.0 ($\sim$350 km altitude), which is about the top portion of the F-layer. Similarly, the greatest rates of decay are seen between pressure levels 3.0 and 4.0, which correspond to about 350 km and 400 km altitudes. Growth-relaxation rates are computed similarly for regions R_2 and R_3, and are shown in Figure 6.6. The growth-relaxation bias evolution model is not used in the night-time region (R_4) since the bias is not observed to have a pronounced growth or decay component as the other regions.

6.4.2 Results

The time series of region-averaged bias estimates, in the case where the growth-relaxation model described in Section 6.4.1 is used for $\mathbf{G}$, are shown for regions

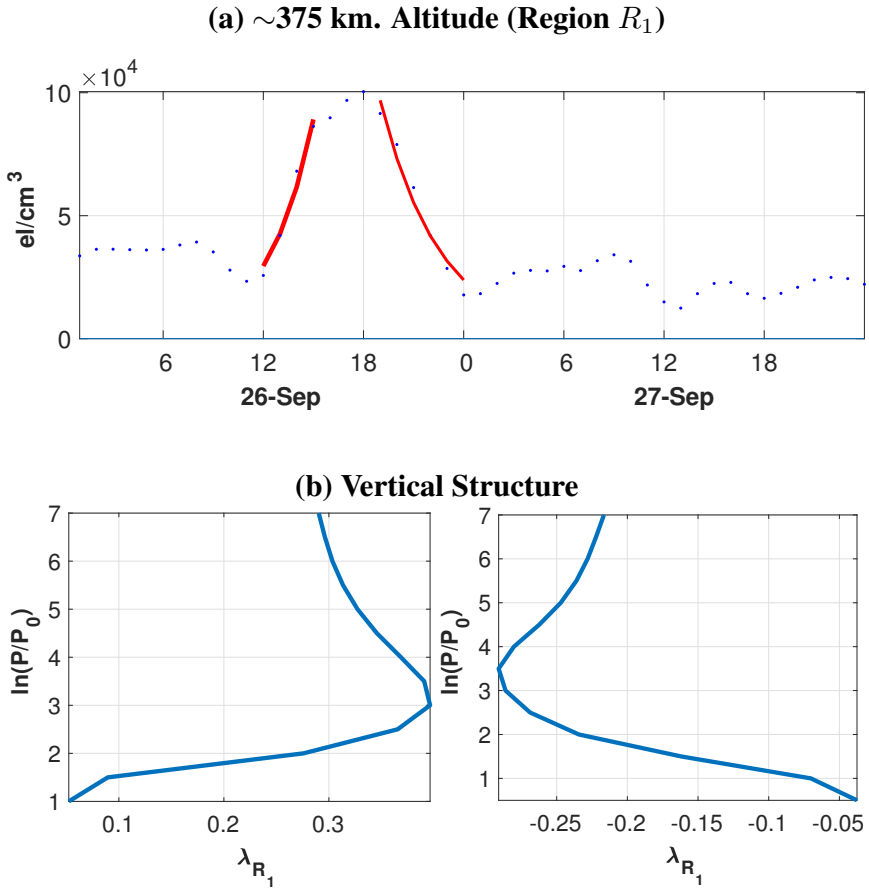

Figure 6.5: (a) RMSE time series of electron density bias in units of el/cm^3, at pressure level 3.5 (∼375 km), averaged over all grid points in region R_1. The horizontal axis is time in hours. The growth and decay factors used in the bias evolution operator, λ_{R_1}, are shown by the red curves and are applied only during the times that these curves cover. (b) Vertical structure of each λ_{R_1}, during the initial (left) and relaxation (right) phases of the geomagnetic storm. The vertical axis is in pressure levels and the horizontal axis is the value of λ_{R_1}.

R_1 and R_2 in Figures 6.7(a)-(b). The quantities in these figures are computed as described in Section 6.3.2 and are analogous to Figures 6.3(a)-(b). Regions R_1 and R_2 are the high-latitude regions over which rapid temporal variations occur during the transitionary periods of the geomagnetic storm. Comparison with Figure 6.3(a)-(b) reveals that the growth-relaxation model for $\mathbf{G}$ provides a considerably improved representation in the initial and relaxation phases of model bias throughout the main

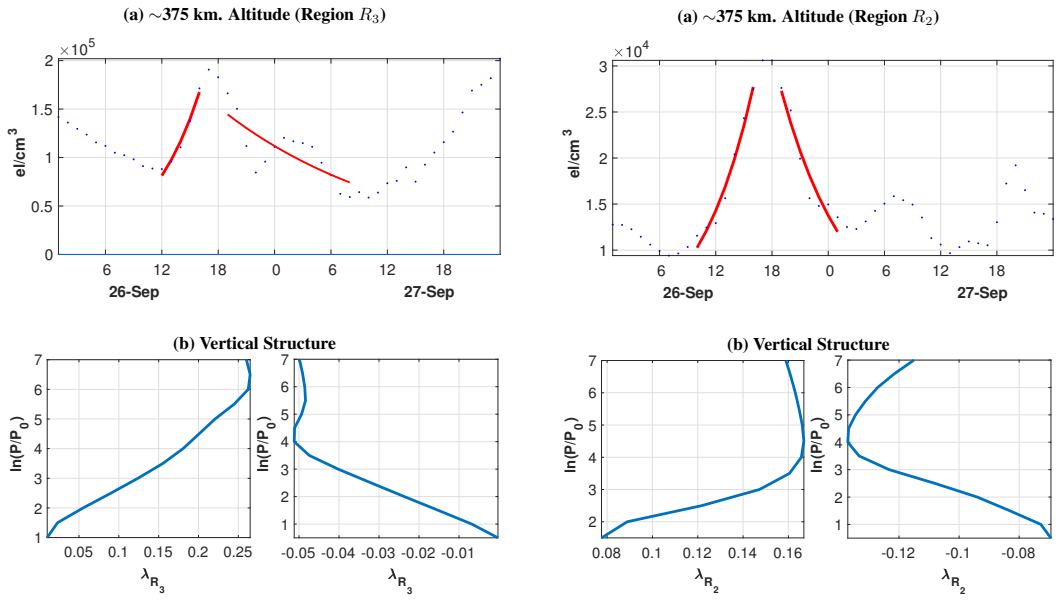

Figure 6.6: (a) Analogous plots of λ values to the ones shown in Figure 6.5, for bias regions R_2 and R_3, respectively.

phase of the geomagnetic storm. The bias estimates still slightly underestimate the rapid growth of the model bias in region R_1 during the beginning of the initial phase between 11:30 and 13:30 UTC, but the application of **G** allows the bias estimates to quickly grow and adequately estimate the peak of the model bias at 16:30 UTC and also follow the sharp decrease in model bias following the main storm phase. The bias evolution in region R_2 is similarly well represented throughout the simulation with the growth-relaxation model of **G**.

Figures 6.7(c)-(d) show the corresponding RMSE time series of 1-h electron density predictions over regions R_1 and R_2 in the case where the growth-relaxation model is used for **G**. These RMSE time series are computed as described in Section 6.3.2 and are analogous to those shown in Figure 6.4(a)-(b). An overall improvement is seen throughout the duration of the geomagnetic storm over regions R_1 and R_2. The most notable benefits of the bias corrections occur during the relaxation phase, where there is improvement of about 40% and 20% in regions R_1 and R_2, respectively, compared

to when using persistent bias evolution model.

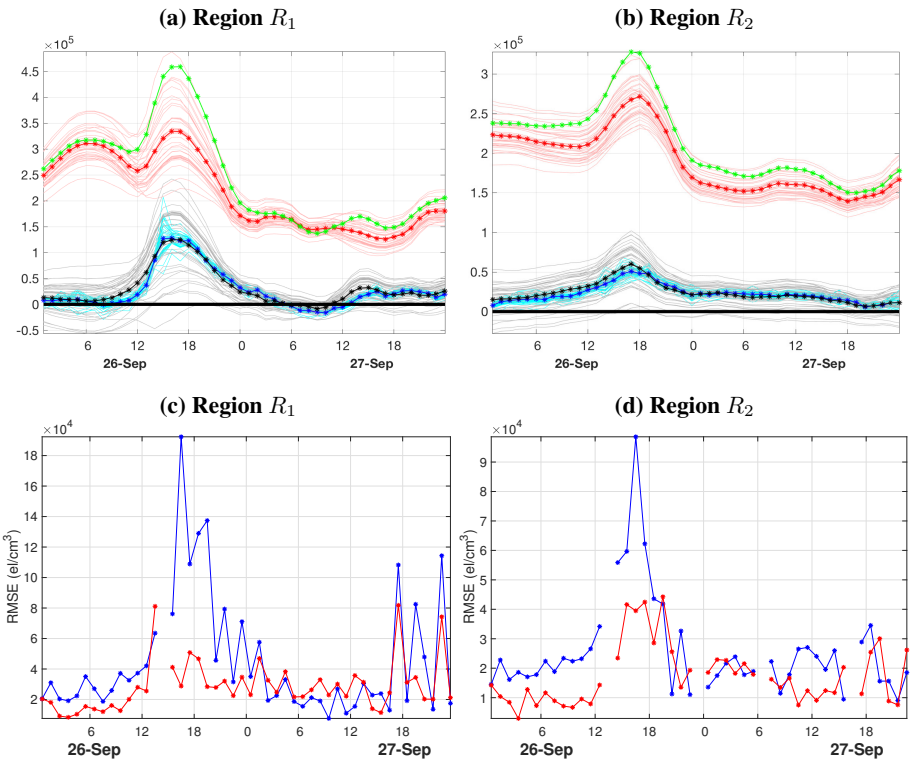

Figure 6.7: (a)-(b) Time series of bias correction parameters averaged over regions R_1 and R_2 respectively. The bias correction values in this figure are calculated in the same manner as Figure 6.3(a)-(b), but in the case where the bias evolution operator described in Section 5 is used. (c)-(d) Time series of the root mean squared error (RMSE) of 1-hour forecasted electron density predictions, in el/cm^3 in regions R_1 and R_2, respectively. The RMSE values in this figure are calculated in the same manner as Figure 6.4(a)-(b), but in the case where the bias evolution operator described in Section 6.4.1 is used.

Figure 6.8 summarizes the spatial distribution of the bias corrections at same representative times of the geomagnetic storm. For comparison, Figure 6.8(a) shows the same background electron density bias from Figure 6.2(a), computed as $\bar{\mathbf{e}}^b - \mathbf{e}^t$. This quantity yields the model error in 1-h electron density predictions at each of the indicated altitudes and times. Figure 6.8(b) shows the global distribution of background bias corrections, $\bar{\mathbf{c}}^b$, at the same altitude and times as Figure 6.2(a). Locations where there the bias corrections are red (blue) correspond to regions where the bias cor-

rections are positive (negative). Comparison with Figure 6.8(a) demonstrates that the spatial structure of bias corrections provides a reasonable estimate of the electron density bias, particularly over the day-time mid-latitude regions. Figure 6.8(c) shows the difference between Figure 6.8(b) and Figure 6.8(a), which indicate what the remaining model error would be after the bias corrections are added at each grid point. Locations where the bias corrections overestimate (underestimate) the model bias are shown in red (blue).

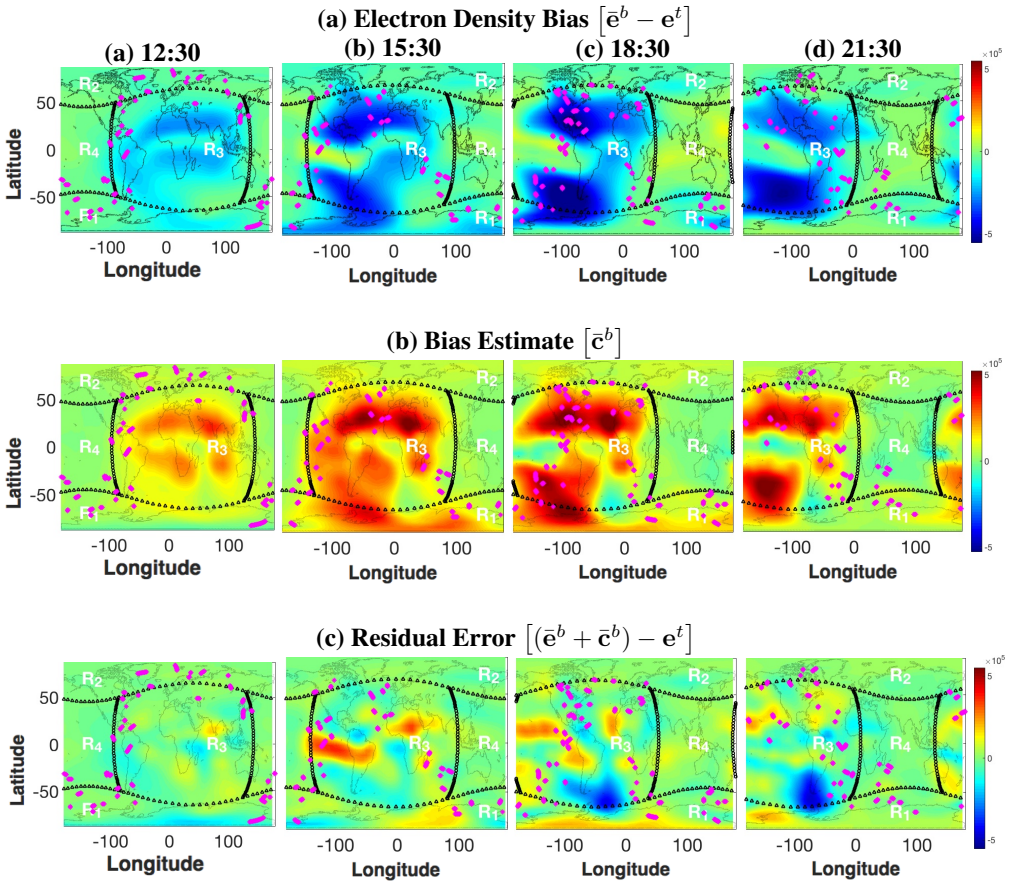

Figure 6.8: (a) Same global maps of electron density bias as shown in Figure 6.2(a). (b) Spatial structure of the bias correction estimates, $\bar{\mathbf{c}}_L^b$, at the same times and altitudes.. (c) Analogous global maps of electron density bias after the field of correction estimates is added.

.

Although there is considerable improvement in the overall bias distribution, there are a few locations where the bias correction misrepresents the model bias, mainly in regions of sparse data coverage. The misrepresentation in some of these locations can be relatively large, which is why these corrections are not used to directly modify the state of the electron density. Overall, the background prediction of electron density is improved at the observed locations, so that the data assimilation procedure may produce a more gentle update during the assimilation of the observations.

Figure 6.9(a) shows 1-h predictions of electron density averaged over all vertical profiles in region R_1 at 12:30, 15:30, 18:30 and 21:30 UTC on 26 September 2011. The horizontal axis is electron density in el/cm^3 and the vertical axis is altitude in km. The region-averaged observations are shown in green and the analogously averaged electron density is shown for the background ensemble mean before (blue) and after (red) the bias corrections are applied. The vertical structure of the applied bias corrections is given by the black curves. The vertical structure of the bias corrections varies similarly to that of the electron density profiles, although the maximum corrections are applied above the F-layer and there are negative corrections applied at altitudes slightly below the F-layer at certain assimilation times. The applied bias corrections improve state estimates of the maximum electron density in the F2-layer (NmF2) and its altitude (hmF2) considerably, particularly during the main phase of the storm (15:30 UTC and 19:30 UTC). Analogous plots of bias corrections applied to regions R_2 and R_3 are shown in Figures 6.9(b)-(c), respectively. Similar improvements in peak density and altitude are observed over these regions. The bias corrections for the most part correctly increase the electron density peak and its altitudes, but there are times where the electron density peak is correctly for improved agreement with the truth.

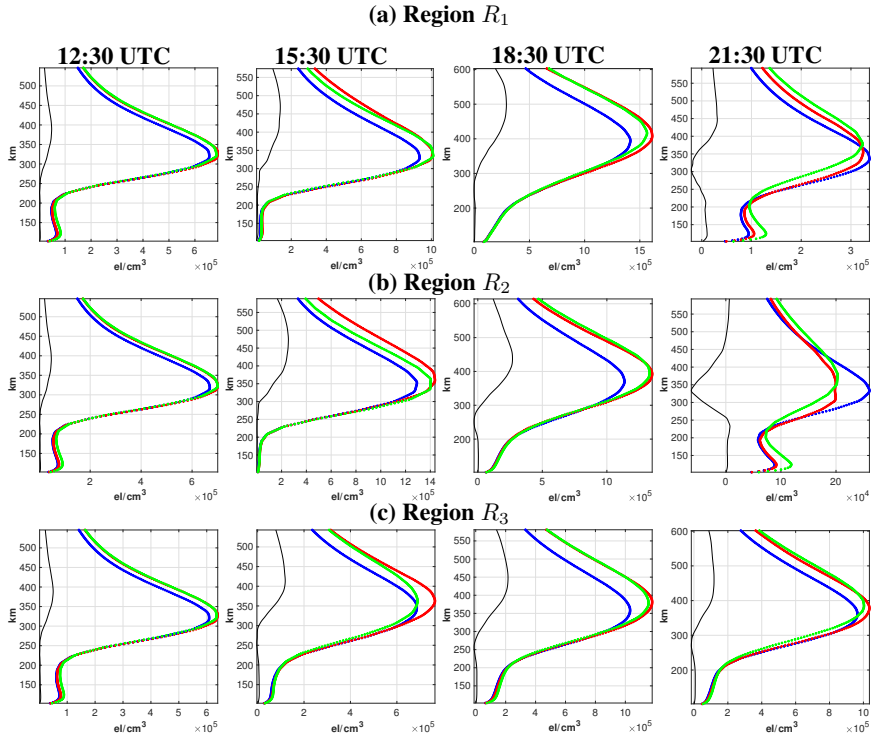

Figure 6.9: (a) Comparison of 1-hour electron density vertical profile predictions at the indicated times on 26 September 2011. The green profile in each figure is the average of all COSMIC electron density vertical profiles located in region R_1 (as defined in Figure 6.2(a)). The forecasted electron density is also shown before (blue) and after (red) the bias correction, shown by the black curve, is applied. (b)-(c) Analogous comparison of electron density predictions over regions R_2 and R_3 at the same times. In all figures, the vertical axis is altitude in km and the horizontal axis is electron density in el/cm^3.

6.5 Discussion and Conclusions

A strategy for model bias correction, within a data assimilation system for complex and sparsely observed dynamical systems is presented. The proposed strategy is applied in observing system experiments, in which the geomagnetic storm of 26 September 2011 is simulated with the TIEGCM. Synthetic electron density vertical profiles, whose locations are given by the COSMIC network during the time period of the storm, are assimilated using the LETKF. Systematic model bias in electron density predictions is simulated through the misspecification of parameterized so-

lar and magnetospheric inputs to the TIEGCM, which specify key dynamics in the ionosphere. The spatial distribution of electron density bias is estimated through a state-augmentation approach. The bias estimates are applied in the evaluation of the forward operator, with the intent to reduce the effect of model errors in the predicted electron densities prior to the assimilation of observations. This methodology permits spatially varying estimation of model bias, which may be useful during extreme events, to account for storm time effects that are not well represented with the parameterized representation of the solar and magnetospheric drivers alone.

Results show that the bias correction strategy yields reasonable estimates of model bias and improves 1-hour electron density forecasts overall. However, during transitionary periods of the geomagnetic storm, where there are sudden large-scale temporal variations in electron density over high-latitude regions, the bias correction strategy does not adjust its bias estimates quickly enough to capture the temporal variability of the model bias. In particular, the bias adjustments underestimate the sudden increase of model bias during the initial storm phase and overestimate the model bias during the relaxation storm phase. This is primarily due to the choice of bias evolution model, $\mathbf{G}$, which is initially taken to be persistence. During these transitionary periods, the model bias undergoes significant variations, while this choice of $\mathbf{G}$ holds the bias estimates constant during each forecasting step, causing some divergence of the bias estimates from the model bias during each forecasting step. Additionally, the observing network is relatively sparse over these regions during this time period, which further slows the adjustment of the bias estimates. Nevertheless, the bias corrections yield considerable improvements before the onset of the geomagnetic storm and during the initial and main phases, but yield inadequate adjustments during the relaxation phase.

To improve estimates of the temporal variations in the model bias during transitionary periods, a growth-relaxation model for $\mathbf{G}$, is implemented at certain pressure levels to approximate the estimated model bias. Over high-latitude regions, the large-scale increase in model bias during the initial storm phase is modeled with exponential growth, while the decrease in model bias that occurs during the relaxation phase is modeled with exponential decay. The rates of growth or decay used in the growth-relaxation model vary by geographical region and altitude.

The benefits of using the growth-relaxation model for $\mathbf{G}$ are primarily seen 1-h electron density predictions over high-latitude regions during transitionary periods of the geomagnetic storm, where the temporal variations in model bias are most prominent. Model bias corrections over the Southern polar region (R_1) yield on average about a 45% and 35% improvement in 1-h electron density predictions during the main and relaxation phases of the storm, respectively. The Northern polar region (R_2 displays a similar peak in model bias evolution and bias corrections yield an improvement of about 45% and 20% during the main and relaxation phases, respectively. Over day-time low-latitude regions, a steady improvement of about 25% is observed throughout the simulation. 1-h predictions of the global distribution of model bias are approximated reasonably well with the bias correction strategy overall. The bias corrections also yields adequate estimates of the vertical structure of the model bias and considerably improves 1-hour electron density predictions of the observed vertical profiles, including the adjustment of the F-layer peak and its altitude.

This work proposes a flexible methodology to extend the capabilities of existing parameterized inputs to drive ionospheric global circulation models. The choice of bias evolution operator is an important component of this methodology when there are significant temporal variations in model bias and the observing network is sparse. This chapter demonstrates the benefits that a simple growth-relaxation model for bias

evolution provides for ionospheric weather forecasting during a geomagnetic storm. Of course, there may be benefit in the development of more sophisticated bias evolution models for $\mathbf{G}$ through statistical studies of other geomagnetic storms, although there may be potential difficulties in determining adequate inputs for $\mathbf{G}$. On the other hand, it may be more suitable to use simple models for $\mathbf{G}$, such as the growth-relaxation model used in this paper, and use re-analysis studies of different geomagnetic storm events to formulate appropriate growth and relaxation rates to describe temporal variations of model bias for different extreme events, depending on annual season or geomagnetic conditions. There may also be benefits in tuning the region configuration over which $\mathbf{G}$ is applied, depending on the specific storm-time effects of interest or biases that a particular model may be known to have. These extensions offer significant potential and merit future consideration.

Chapter 7

CONCLUSION

In this dissertation, the local ensemble transform Kalman filter (LETKF) is proposed as a data assimilation scheme for ionospheric space-weather forecasting. Due to recent advancements in observational coverage in the ionosphere, data assimilation is becoming an increasingly important component for specification and prediction of space weather in the ionosphere. Thus, innovative data assimilation approaches are in high demand. The purpose of this dissertation is to present research done to apply the LETKF for the purposes of space-weather prediction in the ionosphere.

One of the main attractions of the LETKF scheme is its computational efficiency. and model independence. The LETKF is highly amenable to efficient implementation on parallel computers, which is an important consideration due to time constraints that may be involved in space-weather forecasting operations. Due to its localized approach, the operations involved in the LETKF analysis calculation are relatively small and this efficiency scales well to larger systems, since the dimensions involved in the calculations depend on the number of observations and ensemble size, but not state size. These considerations suggest that the LETKF merits consideration as an ionospheric data assimilation scheme. This dissertation is comprised of three main studies where the LETKF is applied to the ionosphere, as discuss below.

The work shown in Chapter 4 is based on (Durazo *et al.*, 2016), which is published in *Physica Scripta*. In this study, the LETKF is applied to an idealized ionospheric model that focuses the interaction of an ion density field and a gravity wave, and the subsequent electromagnetic forcing caused by the movement of charged particles. The results found in this study suggest that the LETKF is effective even when the

observational density is sparse, at least in this perfect-model scenarios. Good results are also obtained even when using large noise levels in the generation of synthetic observations. The results also suggest that the LETKF can perform well with sparse observational networks, provided that the density is reasonably uniform and that the local assimilation regions are chosen to be sufficiently large.

The work shown in Chapter 5 is based on (Durazo *et al.*, 2017), which is published in the *Journal of Geophysical Research: Space Physics*. In this study, the LETKF is applied to the TIEGCM, which is a global circulation model of the ionosphere with operational capabilities. Furthermore, the distribution of the assimilated synthetic observations taken from a realistic satellite network.

A targeted observation strategy based on the influence matrix diagnostic is proposed for ionospheric data assimilation with the LETKF. The targeted observation strategy can be used to optimally select locations for additional observations to target errors in specific state variables. The numerical experiments presented showed results in the case where electron density and zonal component of the neutral winds are targeted. The ability to target errors in this manner can help mitigate ionospheric forecast errors during geomagnetic storms. The quantification of the contribution in each observation is computed simultaneously with the analysis computation, using quantities that are already created, so this targeting strategy can be employed efficiently.

This work is novel in that it formulates the observation influence for the LETKF, and can readily be generalized to other ensemble kalman filters. This work provides a diagnostic tool to identify regions that should be observed in future extreme events, or where future observing infrastructure should be developed as the ionospheric observing network expands. The strategy may also be applied to target errors in other state variables and parameterized solar/magnetospheric inputs, which is an important

since these inputs drive many of the dynamics in the ionosphere. Results demonstrate that, in principle, this strategy can be be used to improve space-weather specification during extreme events, and lays a foundation that can be extended to more realistic operational settings.

The work shown in Chapter 6 is in revision for publication in the *Journal of Geophysical Research: Space Physics*. The work presented in this study describes a methodology to estimate the effects of model bias on 1-hour electron density forecasts. Systematic model bias in electron density predictions is simulated through the misspecification of parameterized solar and magnetospheric inputs to the TIEGCM, which specify key dynamics in the ionosphere. The model bias effects are effectively estimated and used to reduce error in the predicted electron densities.

The work shown in Chapter 6 is novel in that it proposes a bias estimation strategy, which has not previously been applied to complex models with operational capabilities. This work proposes a flexible methodology to extend the capabilities of existing parameterized inputs to drive ionospheric global circulation models. This approach permits spatially varying estimation of model bias, which may be useful during extreme events, to account for storm time effects that are not well represented with the parameterized representation of the solar and magnetospheric drivers alone.

The bias estimation methodology also allows for the prescription of a dynamical model to describe the temporal evolution of the model bias. This work demonstrates the potential benefits that a simple growth-relaxation model for bias evolution can provide for ionospheric weather forecasting during a geomagnetic storm. The approach shown in this work serves as a foundation to improve bias estimation in future extreme events. For example, re-analysis studies of different geomagnetic storm events can be used to formulate adequate growth and relaxation rates to describe temporal variations of model bias for different geomagnetic storms, that depend on

annual season or geomagnetic conditions. These extensions offer significant potential that in principle can greatly improve ionospheric weather prediction and merit future consideration for use in operational settings.

Anderson, B. J., H. Korth, C. L. Waters, D. L. Green, V. G. Merkin, R. J. Barnes and L. P. Dyrud, "Development of largescale birkeland currents determined from the active magnetosphere and planetary electrodynamics response experiment", Geophysical Research Letters **41**, 9, 3017–3025, URL `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014GL059941` (2014).

Anderson, J., "An ensemble adjustment Kalman filter for data assimilation", Monthly Weather Review **129**, 2884–2903 (2001).

Anderson, J., "An adaptive covariance inflation error correction algorithm for ensemble filters", Tellus **59**, 2, 210–224 (2007).

Anderson, J., "Spatially and temporally varying adaptive covariance inflation for ensemble filters", Tellus **61A**, 72–83 (2009).

Baek, S.-j., B. R. Hunt, E. Kalnay, E. Ott and I. Szunyogh, "Local ensemble kalman filtering in the presence of model bias", Tellus A **58**, 3, 293–306, URL `http://dx.doi.org/10.1111/j.1600-0870.2006.00178.x` (2006).

Baker, D. N., X. Li, A. Pulkkinen, C. M. Ngwira, M. L. Mays, A. B. Galvin and K. D. C. Simunac, "A major solar eruptive event in July 2012: Defining extreme space weather scenarios", Space Weather **11**, 585–591 (2013).

Bellsky, T., J. Berwald and L. Mitchell, "Non-global parameter estimation using local ensemble Kalman filtering", Monthly Weather Review **142**, 2150–2164 (2014).

Bilitza, D. and B. Reinisch, "International reference ionosphere 2007: improvements and new parameters", Advances in Space Research **42**, 599–609 (2007).

Bishop, C., B. Etherton and co authors, "Adaptive sampling with the ensemble transform Kalman filter. part 1: Theoretical aspects", Monthly Weather Review **129**, 420–436 (2001).

Bolduc, L., "GIC observations and studies in the Hydro-Quebec power system", Atmospheric and Solar-Terrestrial Physics **64**, 1793–1802 (2002).

Buensanto, M., "Ionospheric storms - a review", Space Science Review **88**, 563–601 (1999).

Burgers, G., P. Jan Van Leeuwen and co authors, "Analysis scheme in the ensemble Kalman filter", Monthly Weather Review **126**, 1719–1724 (1998).

Bust, G. and C. Mitchell, "History, current state, and future directions of ionospheric imaging", Reviews of Geophysics **46**, RG1003 (2008).

Cardinali, C., S. Pezzulli and A. Andersson, "Influence-matrix diagnostic of a data assimilation system", Quarterly Journal of the Royal Meteorological Society **130**, 2767–2786 (2004).

Chartier, A., D. Jackson and C. Mitchell, "A comparison of the effects of initializing different thermosphere-ionosphere model fields on storm time density forecasts", Journal of Geophysical Research: Space Physics **118**, 7329–7337 (2013).

Chartier, A. T., T. Matsuo, J. L. Anderson, N. Collins, T. J. Hoar, L. Gang, C. N. Mitchell, A. J. Coster, L. J. Paxton and G. S. Bust, "Ionospheric data assimilation and forecasting during storms", Journal of Geophysical Research: Space Physics **121**, 1, 764–778, URL `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014JA020799` (2016).

Chen, C., C. Lin, T. Matsuo, I. Lee, J. Liu, J. Lin and C. Hsu, "Ionospheric data assimilation with thermosphere-ionosphere-electrodynamics general circulation model and GPS-TEC during geomagnetic storm conditions", Journal of Geophysical Research: Space Physics **121**, doi:10.1002/ 2015JA021787 (2016).

Coster, A., J. Foster and P. Erickson, "Monitoring the ionosphere with GPS", GPS World **3**, 42–49 (2003).

Daley, R., "Atmospheric data analysis", Cambridge University Press **2**, 50–65 (1992).

Du, D., W. Xu, M. Zhao, B. Chen, J. Lu and G. Yang, "A sensitive geomagnetic activity index for space weather operation", Space Weather **8**, 12 (2010).

Dubey, S., R. Wahi and A. Gwal, "Ionospheric effects on GPS positioning", Advances in Space Research **38**, 2478–2484 (2006).

Durazo, J., E. Kostelich, A. Mahalov and W. Tang, "Observing system experiments with an ionospheric electrodynamics model", Physica Scripta **91**, 4, 044001 (2016).

Durazo, J. A., E. J. Kostelich and A. Mahalov, "Local ensemble transform kalman filter for ionospheric data assimilation: Observation influence analysis during a geomagnetic storm event", Journal of Geophysical Research: Space Physics **122**, 9, 9652–9669, URL `http://dx.doi.org/10.1002/2017JA024274`, 2017JA024274 (2017).

Edwin, S., A. Maute and A. Richmond, "Simulating magnetosphere-ionosphere coupling in tie-gcm", (2011).

Epstein, E., "Stochastic dynamic prediction", Tellus **47**, 692892 (1969).

Evensen, G., "Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics", Journal of Geophysical Research **99**, 10143–10162 (1994).

Evensen, G., *Data Assimilation* (Springer-Verlag, New York, 2009).

Ghil, M. and P. Malanotte-Rizzoli, "Data assimilation in metereology and oceanography", Advances in geophysics **33**, 141–160 (1991).

Godinez, H. C., E. Lawrence, D. Higdon, A. Ridley, J. Koller and A. Klimenko, *Specification of the Ionosphere-Thermosphere Using the Ensemble Kalman Filter*, pp. 274–283 (Springer International Publishing, Cham, 2015).

Hamill, T. and C. Snyder, "A hybrid ensemble Kalman filter-3d variational analysis scheme", American Metereological Society **128**, 2905–2919 (2000).

Hamill, T. M., J. S. Whitaker and C. Snyder, "Distance-dependent filtering of background error covariance estimates in an ensemble kalman filter", Monthly Weather Review **129**, 11, 2776–2790 (2001).

Heelis, R. A., J. K. Lowell and R. W. Spiro, "A model of the high-latitude ionospheric convection pattern", Journal of Geophysical Research: Space Physics **87**, A8, 6339–6345, URL `http://onlinelibrary.wiley.com/doi/10.1029/JA087iA08p06339/abstract` (1982).

Houtekamer, P. and H. Mitchell, "Data assimilation using an ensemble Kalman filter technique", Monthly Weather Review **126**, 796–811 (1997).

Hsu, C.-T., T. Matsuo, W. Wang and J.-Y. Liu, "Effects of inferring unobserved thermospheric and ionospheric state variables by using an ensemble Kalman filter on global ionospheric specification and forecasting", Journal of Geophysical Research: Space Physics **119**, 9256–9267 (2014).

Hunt, B. R., E. J. Kostelich and I. Szunyogh, "Efficient data assimilation for spatiotemporal chaos: A local ensemble Kalman filter", Physica D **230**, 112–126 (2007).

Kalnay, E., *Atmospheric Modeling, Data Assimilation, and Predictability* (Cambridge University Press, 2003).

Kelley, M., *The Earth's ionosphere: plasma physics and electrodynamics* (Elsevier, 2009).

Klimenko, V., M, V. V. Klimenko, F. Bessarab, K. Ratovsky, I. E. Zakharenkova, I. A. Nosikov and O. I. Yagodkina, "Influence of geomagnetic storms of September 26-30, 2011, on the ionosphere and hf radiowave propagation. i. ionospheric effects", Geomagnetism and Aeronomy **55**, 744–762 (2015).

Koller, J., S. Brennan, H. Godinez, D. Higdon and co authors, "Integrated modeling of perturbations in atmospheres for conjunction tracking", AMOS conference: LA-UR-13-26931 (2013).

Lee, I., T. Matsuo, A. Richmond and co authors, "Assimilation of FORMOSAT-3/COSMIC electron density profiles into thermosphere-ionosphere coupling model by using ensemble Kalman filter", Journal of Geophysical Research **117**, A10318 (2012).

Lei, J., A. G. Burns, T. Tsugawa, W. Wang, S. C. Solomon and M. Wiltberger, "Observations and simulations of quasiperiodic ionospheric oscillations and large?scale traveling ionospheric disturbances during the december 2006 geomagnetic storm", Journal of Geophysical Research: Space Physics **113**, A6, URL `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2008JA013090` (2008a).

Lei, J., W. Wang, A. G. Burns, S. C. Solomon, A. D. Richmond, M. Wiltberger, L. P. Goncharenko, A. Coster and B. W. Reinisch, "Observations and simulations of the ionospheric and thermospheric response to the december 2006 geomagnetic storm: Initial phase", Journal of Geophysical Research: Space Physics **113**, A1, URL `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2007JA012807` (2008b).

Leith, C., "Theoretical skill of Monte Carlo forecasts", Monthly Weather Review **102**, 409–418 (1974).

Lewis, J. M., S. Lakshmivarahan and S. Dhall, *Dynamic Data Assimilation: A Least Squares Approach*, vol. 104 of *Encyclopedia of Mathematics and its Applications* (Cambridge University Press, 2006).

Li, H., E. Kalnay and T. Miyoshi, "Simultaneous estimation of covariance inflation and observation errors within an ensemble kalman filter", Quarterly Journal of the Royal Metereological Society **135**, 523–533 (2009).

Liu, J., C. Lin, C. Lin, H. Tsai, S. Solomon, Y. Sun, I. Lee, W. Schreiner and Y. Kuo, "Artificial plasma cave in the low-latitude ionosphere results from the radio occultation inversion of the FORMOSAT-3/COSMIC", Journal of Geophysical Research **115**, doi:10.1029/2009JA015079 (2010).

Matsuo, T. and E. Araujo-Pradere, "Role of thermosphere-ionosphere coupling in a global ionospheric specification", Radio Science **46**, RS0D23 (2011).

Matsuo, T., I.-T. Lee and J. Anderson, "Upper atmosphere data assimilation with an ensemble Kalman filter", AGU p. RS0D23 (2013).

Miyoshi, T., "The Gaussian approach to adaptive covariance inflation and its implementation with the local ensemble transform Kalman filter", Monthly Weather Review **139**, 1519–1531 (2011).

Miyoshi, T., "The 10,240-member ensemble Kalman filtering with an intermediate AGCM", AGCM, Geophys. Res. Lett **41**, doi:10.1002/2014GL060863 (2014).

Ngwira, C., A. Pulkkinen, M. Leila Mays, M. Kuznetsova, A. Galvin, K. Simunac, D. Baker, Z. Li, X., Y. and A. Glocer, "Simulation of the 23 July 2012 extreme space weather event: What if this extremely rare CME was Earth directed?", Space Weather **11**, 671–679 (2013).

Ott, E., B. Hunt and co authors, "Exploiting local low dimensionality of the atmospheric dynamics for efficient ensemble Kalman filtering", No. Physics **126**, 0203058 (2002).

Ott, E., B. Hunt and co authors, "A local ensemble Kalman filter for atmospheric data assimilation", Tellus **56**, 415–428 (2004).

Patil, D. J., B. R. Hunt, E. Kalnay, J. A. Yorke and E. Ott, "Local low dimensionality of atmospheric dynamics", Physical Review Letters **86**, 5878–5881 (2001).

Pedatella, N. M., J. Lei, K. M. Larson and J. M. Forbes, "Observations of the ionospheric response to the 15 december 2006 geomagnetic storm: Long?duration positive storm effect", Journal of Geophysical Research: Space Physics **114**, A12, URL `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2009JA014568` (2009).

Petrie, R. and S. Dance, "Ensemble-based data assimilation and the localization problem", Weather **65**, 65–69 (2010).

Prolss, G. W. and M. K. Bird, *Physics of the Earth's space environment: an introduction* (Springer, 2004).

Richards, P. G., J. A. Fennelly and D. G. Torr, "Euvac: A solar euv flux model for aeronomic calculations", Journal of Geophysical Research: Space Physics **99**, A5, 8981–8992, URL `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/94JA00518` (1994).

Rocken, C., Y.-H. Kuo, W. Schreiner and co authors, "COSMIC system description", Terr. Atmos. Oceanic Sci. **11**, 21–52 (2000).

Ruiz, J. J., M. Pulido and T. Miyoshi, "Estimating model parameters with ensemble-based data assimilation: A review", Journal of the Meteorological Society of Japan. Ser. II **91**, 2, 79–99 (2013).

Scherliess, L., R. W. Schunk, J. J. Sojka and D. Thompson, "Global assimilation of ionospheric measurements (GAIM)", Radio Science **39**, RS1S02 (2004).

Schunk, R., L. Sherliess and J. Sojka, "Recent approaches to modeling ionospheric weather", Advances in Space Research **31**, 819–828 (2003).

Siscoe, G. and S. Solomon, "Aspects of data assimilation peculiar to space weather forecasting", Space Weather **4**, S04002 (2006).

Stauning, P., "Power grid disturbances and polar cap index during geomagnetic storms", Journal of Space Weather and Space Climate **3**, A22 (2013).

Szunyogh, I., E. Kostelich, G. Gyarmati, D. Patil and co authors, "Assessing a local ensemble Kalman filter: perfect model experiments with the National Centers for Environmental Prediction global model", Tellus **57A**, 528–545 (2005).

Talagrand, O., "Assimilation of observations, an introduction", Journal of Metereological Society of Japan **75**, 191–209 (1997).

Tang, W. and A. Mahalov, "Stochastic Lagrangian dynamics for charged flows in the *E-F* regions of ionosphere", Physics of Plasmas **20**, 032305 (2013).

Taylor, J. R., *Numerical simulations of the stratified oceanic bottom boundary layer.*, Ph.D. thesis, University of California, San Diego (2008).

Tippet, M., J. Anderson and co authors, "Ensemble square root filters", Monthly Weather Review **131**, 1485–1490 (2003).

Tsurutani, B., O. Verkhoglyadova, A. Mannucci, G. Lakhina, G. Li and G. Zank, "A brief review of solar flare effects on the ionosphere", Radio Science **44**, 1 (2009).

Wang, C., J. Hajj, B. Wilson and co authors, "Development of the Global Assimilative Ionosphere Model", Radio Science **39**, RS1S06 (2004).

Wang, X., C. Snyder and T. Hill, "On the theoretical equivalence of differently proposed ensemble-3dvar hybrid analysis schemes", Monthly Weather Review **135**, 222–226 (2006).

Weaver, M., "Halloween space weather storms of 2003", US Department of Commerce, National Oceanic and Atmospheric Administration, Oceanic and Atmospheric Research Laboratories, Space Environment Center **12**, 12–15 (2004).

Webber, W., F. McDonald, J. Lockwood and B. Heikkila, "The effect of the July 14, 2000 "Bastille Day" solar flare event on $>$ 70 MeV galactic cosmic rays observed at V1 and V2 in the distant heliosphere", Geophysical Research Letters **29**, 12–15 (2002).

Weimer, D. R., "Improved ionospheric electrodynamic models and application to calculating joule heating rates", Journal of Geophysical Research: Space Physics **110**, A5, URL `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2004JA010884` (2005).

Whitaker, J. and T. Hamill, "Ensemble data assimilation without perturbed observations", Monthly Weather Review **130**, 1913–1924 (2001).

Wing, S., J. Johnson, J. Jen, C.-I. Meng, D. Sibeck, K. Bechtold, J. Freeman, K. Costello, M. Balikhin and K. Takahashi, "Kp forecast models", Journal of Geophysical Research: Space Physics **110**, A4 (2005).

Xu, W., "Uncertainty in magnetic activity indices", Science in China Series E: Technological Sciences **51**, 10, 1659–1664 (2008).

Yue, X., W. Schreiner, Y.-H. Kuo and co authors, "Global 3-D ionospheric electron density reanalysis based on multi source data assimilation", Journal of Geophysical Research **117**, A09325 (2012).

Yue, X., W. Schreiner, J. Lei, S. Sokolovskiy, C. Rocken, D. Hunt and K. Y., "Error analysis of Abel retrieved electron density profiles from radio occultation measurements", Annales Geophysicae: Atmospheres, Hydrospheres and Space Sciences **28**, 217–222 (2010).

Appendix A

LETKF ALGORITHM PSEUDOCODE 1: TARGETED OBSERVATION STRATEGY

A.1 Observation Influence Algorithm Pseudocode

The following algorithm demonstrates the analysis procedure used to compute the analysis at an arbitrary grid point L, and how the observation influence is calculated. We assume there are d components in the state vector and that the ensemble is of size k. The quantities below are related to the grid point L and its associated local region, where we assume there are ℓ_C and ℓ_A observations from the COSMIC and augmented networks, respectively. The total number of observations is $\ell = \ell_C + \ell_A$. The computations below may be performed independently (and in parallel) at each grid point L.

1. **Analysis computation**. A detailed description of the analysis calculation and the associated computational requirements can be found in (Hunt *et al.*, 2007).
   (a) Compute the $d \times k$ and $\ell \times k$ matrices, $\mathbf{X}_L^b$ and $\mathbf{Y}_L^b$ as described in Section .3.4.1
   (b) Form the $k \times \ell$ matrix $\mathbf{G}_L^F = (\mathbf{Y}_L^F)^{\mathrm{T}}(\mathbf{R}_L^F)^{-1}$.
   (c) Compute the $k \times k$ analysis covariance matrix $\widetilde{\mathbf{P}}_L^{a(F)} = \left[(k-1)\mathbf{I}/\alpha_L + \mathbf{G}_L^F\mathbf{Y}_L^F\right]^{-1}$.
   (d) The analysis ensemble is constructed in terms of the "weight matrix", $\mathbf{W}_L^{a(F)}$, which is computed in two steps.
      i. Compute the $k \times k$ symmetric square root of the analysis covariance matrix $\widehat{\mathbf{W}}_L^{a(F)} = [(k-1)\widetilde{\mathbf{P}}_L^{a(F)}]^{1/2}$.
      ii. Define the k-vector $\bar{\mathbf{w}}^{a(F)} = \widetilde{\mathbf{P}}_L^{a(F)}\mathbf{G}_L^F(\mathbf{y}^F - \bar{\mathbf{y}}_L^F)$ and add it to each column of $\widehat{\mathbf{W}}_L^{a(F)}$ to form the $k \times k$ weight matrix $\mathbf{W}_L^{a(F)}$.
   (e) Compute the analysis perturbation matrix $\mathbf{X}_L^{a(F)} = \mathbf{X}_L^b\mathbf{W}_L^{a(F)}$.
   (f) The jth column of the analysis ensemble is formed by adding the background ensemble mean, $\bar{\mathbf{x}}_L^b$, to the jth column of $\mathbf{X}_L^{a(F)}$, $j = 1, 2, \ldots, k$.

2. **Observation influence on $\bar{\mathbf{x}}_L^{a(F)}$**
   (a) Compute the $k \times \ell_C$ and $k \times \ell_A$ matrices $\mathbf{G}_L^C = (\mathbf{Y}_L^C)^{\mathrm{T}}(\mathbf{R}_L^C)^{-1}$ and $\mathbf{G}_L^A = (\mathbf{Y}_L^A)^{\mathrm{T}}(\mathbf{R}_L^A)^{-1}$, respectively.
   (b) Compute the $d \times \ell_C$ matrix of influence from local COSMIC observations $\mathbf{S}_L^{XC} = \mathbf{X}_L^b\widetilde{\mathbf{P}}_L^{a(F)}\mathbf{G}_L^C$.
   (c) Compute the $d \times \ell_A$ matrix of influence from local augmented observations $\mathbf{S}_L^{XA} = \mathbf{X}_L^b\widetilde{\mathbf{P}}_L^{a(F)}\mathbf{G}_L^A$.
   (d) The sum of the columns of $\mathbf{S}_L^{XC}$ and $\mathbf{S}_L^{XA}$ respectively yields the d-vectors of total observation influence from the COSMIC and augmented networks on each component of $\bar{\mathbf{x}}_L^{a(F)}$.

3. **Compute analysis adjustment, $\bar{\mathbf{x}}_L^{a(F)} - \bar{\mathbf{x}}_L^{a(C)}$, due to the augmented observations.**

(a) Compute the $\ell_A \times \ell_A$ matrix $\left[\mathbf{I} - \mathbf{S}_L^{AA}\right]^{-1}$. If $k < \ell_A$, use the more efficient formulation $\left[\mathbf{I} - \mathbf{S}_L^{AA}\right]^{-1} = \mathbf{I} + (\mathbf{R}_L^A)^{-1}\mathbf{Y}_L^A\widetilde{\mathbf{P}}_L^{a(C)}(\mathbf{Y}_L^A)^{\mathrm{T}}$, where $\widetilde{\mathbf{P}}_L^{a(C)} = \left[(k-1)\mathbf{I}/\alpha_L + \mathbf{G}_L^C\mathbf{Y}_L^C\right]^{-1}$.

(b) Compute the ℓ_A-vector of analysis residuals at the augmented observation locations, $\mathbf{r}_L^A = \mathbf{y}_L^A - \mathbf{H}_L^A\bar{\mathbf{x}}_L^{a(F)}$

(c) Compute the d-vector of analysis adjustments $\bar{\mathbf{x}}_L^{a(F)} - \bar{\mathbf{x}}_L^{a(C)} = \mathbf{S}_L^{XA}\left[\mathbf{I} - \mathbf{S}_L^{AA}\right]^{-1}\mathbf{r}_L^A$. The ith entry contains the adjustment on the ith analyzed component of $\bar{\mathbf{x}}_L^{a(F)}$.

Appendix B

LETKF ALGORITHM PSEUDOCODE 2: BIAS ESTIMATION STRATEGY

B.1 Bias Estimation Algorithm Pseudocode

The following algorithm demonstrates the procedure used to compute the analysis at an arbitrary grid point L, and how the model bias corrections are applied and updated. We assume there are d components in the local state vector, $\mathbf{x}_L^b$, including the bias correction parameter, and that the ensemble is of size k. The quantities below are related to the grid point L and its associated local region, where we assume there are ℓ observations. The computations below may be performed independently (and in parallel) at each grid point L.

1. **Analysis computation**. A detailed description of the analysis calculation and the associated computational complexity is provided in (Hunt *et al.*, 2007). In the following suppose the forecast and bias estimates have been propagated to a time t_n, so that the global state vector is given by $\mathbf{u}_n^{b(j)} = \mathbf{M}_n(\mathbf{u}_{n-1}^{a(j)})$ and the global bias correction vector is given by $\mathbf{c}_n^b = \mathbf{G}_{n-1}(\mathbf{c}_{n-1}^a)$. Now consider the analysis update of a local state vector, $\mathbf{x}_L^{b(j)}$ at a fixed grid point L.

   (a) Construct the $d \times k$ matrix $\mathbf{X}_L^b$, corresponding to the augmented local state vector $\left[\left(\mathbf{x}_L^{b(j)}\right)^{\mathrm{T}} \left(\mathbf{c}_L^{b(j)}\right)^{\mathrm{T}}\right]^{\mathrm{T}}$

   (b) Construct the $\ell \times k$ matrix, $\mathbf{Y}_L^b = \mathbf{H}_L(\mathbf{u}^{b(j)} + \mathbf{c}^{b(j)})$ as described in Section 6.2.

   (c) Form the $k \times \ell$ matrix $\mathbf{G}_L = (\mathbf{Y}_L)^{\mathrm{T}}(\mathbf{R}_L)^{-1}$.

   (d) Compute the $k \times k$ analysis covariance matrix $\widetilde{\mathbf{P}}_L^a = \left[(k-1)\mathbf{I}/\alpha_L + \mathbf{G}_L\mathbf{Y}_L\right]^{-1}$.

   (e) The analysis ensemble is constructed in terms of the "weight matrix", $\mathbf{W}_L^a$, which is computed in two steps.

      i. Compute the $k \times k$ symmetric square root of the analysis covariance matrix $\widehat{\mathbf{W}}_L^a = [(k-1)\widetilde{\mathbf{P}}_L^a]^{1/2}$.

      ii. Define the k-vector $\bar{\mathbf{w}}^a = \widetilde{\mathbf{P}}_L^a\mathbf{G}_L(\mathbf{y}^o - \bar{\mathbf{y}}_L)$ and add it to each column of $\widehat{\mathbf{W}}_L^a$ to form the $k \times k$ weight matrix $\mathbf{W}_L^a$.

   (f) Compute the analysis perturbation matrix $\mathbf{X}_L^a = \mathbf{X}_L^b\mathbf{W}_L^a$.

   (g) The jth column of the analysis ensemble is formed by adding the background ensemble mean, $\bar{\mathbf{x}}_L^b$, to the jth column of $\mathbf{X}_L^a$, $j = 1, 2, \ldots, k$.

2. **Propagate ionospheric and model bias state estimates**

   (a) Apply forecast model $\mathbf{M}_n$ (given by the TIEGCM). Computes $\mathbf{u}_{n+1}^{b(j)} = \mathbf{M}_{n+1}(\mathbf{u}_n^{a(j)})$

   (b) Apply bias evolution operator $\mathbf{G}$, to compute $\mathbf{c}_{n+1}^b = \mathbf{G}_{n+1}(\mathbf{c}_n^a)$

Appendix C

DERIVATION OF THE ANALYSIS ADJUSTMENT DUE TO AUGMENTED OBSERVATIONS

C.0.1 Derivation of Analysis Adjustment

The discussion below describes the derivation of the analysis adjustment, as described in Section 5.5. Consider the weighted regression system given in Equation (5.1) of the main text at a fixed grid point L, which consists of the full set of observations and background forecast at L. Due to the partitioning of the full network of observations into the COSMIC and augmented networks discussed in Section 4, the system of regressions for the full system may also be partitioned as $\mathbf{Z}_L^F = \left[(\mathbf{H}_L^C)^{\mathrm{T}}\ (\mathbf{H}_L^A)^{\mathrm{T}}\ \mathbf{I}_d\right]^{\mathrm{T}}$ and the associated predictants as $\mathbf{z}_L^F = \left[(\mathbf{y}_L^C)^{\mathrm{T}}\ (\mathbf{y}_L^A)^{\mathrm{T}}\ (\bar{\mathbf{x}}_L^b)^{\mathrm{T}}\right]^{\mathrm{T}}$. The superscripts C and A refer to quantities associated with the COSMIC and augmented observation networks, respectively, and the b denotes the background forecast. The full system covariance matrix is given by

$$\mathbf{\Omega}_L^F = \begin{bmatrix} \begin{bmatrix} \mathbf{R}_L^C & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_L^A \end{bmatrix} & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_L^b \end{bmatrix}. \tag{C.1}$$

The vector of regression parameters for the full network is given by

$$\boldsymbol{\beta}_L^F = \left[(\mathbf{Z}_L^F)^{\mathrm{T}}(\mathbf{\Omega}_L^F)^{-1}\mathbf{Z}_L^F\right]^{-1}(\mathbf{Z}_L^F)^{\mathrm{T}}(\mathbf{\Omega}_L^F)^{-1}\mathbf{z}_L^F, \tag{C.2}$$

which is equivalent to equation (12) in the main text. The solution of system (C.2) for the full observing network is then given by

$$\widehat{\mathbf{z}}_L^F = \mathbf{Z}_L^F\boldsymbol{\beta}_L^F = \left[(\mathbf{H}_L^C\boldsymbol{\beta}_L^F)^{\mathrm{T}}\ (\mathbf{H}_L^A\boldsymbol{\beta}_L^F)^{\mathrm{T}}\ (\boldsymbol{\beta}_L^F)^{\mathrm{T}}\right]^{\mathrm{T}}. \tag{C.3}$$

The analyzed components of the state vector at L obtained with the full observation network regression are thus given by $\boldsymbol{\beta}_L^F$.

We now consider the solution of the analogous weighted regression system, in which all augmented observations are not included in the observation vector. Suppose the augmented observations system consists of regressions, predictants and covariance matrix given by $\mathbf{Z}_L^A = \mathbf{H}_L^A$, $\mathbf{z}_L^A = \mathbf{y}_L^A$ and $\mathbf{\Omega}_L^A = \mathbf{R}_L^A$, respectively. Due to the partitioning of the full observation network, the resulting system has only quantities associated with COSMIC observations and the background forecast. We denote these quantities with the superscripts C and b respectively. The matrix of regressions and the vector of predictants associated with resulting system are given by $\mathbf{Z}_L^C = \left[(\mathbf{H}_L^C)^{\mathrm{T}}\ \mathbf{I}_d\right]^{\mathrm{T}}$ and $\mathbf{z}_L^C = \left[(\mathbf{y}_L^C)^{\mathrm{T}}\ (\bar{\mathbf{x}}_L^b)^{\mathrm{T}}\right]^{\mathrm{T}}$, and the system covariance matrix is given by $\mathbf{\Omega}_L^C = \begin{bmatrix} \mathbf{R}^C & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_L^b \end{bmatrix}$. The vector of regression parameters given only the COSMIC observing system is

$$\boldsymbol{\beta}_L^C = \left[(\mathbf{Z}_L^C)^{\mathrm{T}}(\mathbf{\Omega}_L^C)^{-1}\mathbf{Z}_L^C\right]^{-1}(\mathbf{Z}_L^C)^{\mathrm{T}}(\mathbf{\Omega}_L^C)^{-1}\mathbf{z}_L^C. \tag{C.4}$$

The COSMIC-only system solution is given by

$$\widehat{\mathbf{z}}_L^C = \mathbf{Z}_L^C\boldsymbol{\beta}_L^C = \left[(\mathbf{H}_L^C\boldsymbol{\beta}_L^C)^{\mathrm{T}}\ (\boldsymbol{\beta}_L^C)^{\mathrm{T}}\right]^{\mathrm{T}}. \tag{C.5}$$

Similarly to the full network regression, the analyzed components of the state vector at L obtained with the COSMIC-only observation network regression are given by $\boldsymbol{\beta}_L^C$. The analysis adjustment due to augmented observations is thus quantified by computing $\boldsymbol{\beta}_L^F - \boldsymbol{\beta}_L^C$. To do so, we make use of the identities

$$\begin{aligned}
\left(\mathbf{P}_L^{a(F)}\right)^{-1} &= (\mathbf{Z}_L^F)^{\mathrm{T}}(\boldsymbol{\Omega}_L^F)^{-1}\mathbf{Z}_L^F && \text{(C.6)}\\
(\mathbf{Z}_L^F)^{\mathrm{T}}(\boldsymbol{\Omega}_L^F)^{-1}\mathbf{Z}_L^F &= (\mathbf{Z}_L^C)^{\mathrm{T}}(\boldsymbol{\Omega}_L^C)^{-1}\mathbf{Z}_L^C + (\mathbf{Z}_L^A)^{\mathrm{T}}(\boldsymbol{\Omega}_L^A)^{-1}\mathbf{Z}_L^A && \text{(C.7)}\\
(\mathbf{Z}_L^F)^{\mathrm{T}}(\boldsymbol{\Omega}_L^F)^{-1}\mathbf{z}_L^F &= (\mathbf{Z}_L^C)^{\mathrm{T}}(\boldsymbol{\Omega}_L^C)^{-1}\mathbf{z}_L^C + (\mathbf{Z}_L^A)^{\mathrm{T}}(\boldsymbol{\Omega}_L^A)^{-1}\mathbf{z}_L^A, && \text{(C.8)}
\end{aligned}$$

which can be verified by simply expanding both sides of each equation. Applying identities (C.6) and (C.7) to equation (C.4) yields

$$\boldsymbol{\beta}_L^C = \left[(\mathbf{P}_L^{a(F)})^{-1} - (\mathbf{Z}_L^A)^{\mathrm{T}}(\boldsymbol{\Omega}_L^A)^{-1}\mathbf{Z}_L^A\right]^{-1}(\mathbf{Z}_L^C)^{\mathrm{T}}(\boldsymbol{\Omega}_L^C)^{-1}\mathbf{z}_L^C$$

Applying the Woodbury matrix identity on the inverse factor yields

$$\begin{aligned}
&\left[(\mathbf{P}_L^{a(F)})^{-1} - (\mathbf{Z}_L^A)^{\mathrm{T}}(\boldsymbol{\Omega}_L^A)^{-1}\mathbf{Z}_L^A\right]^{-1}\\
&= \mathbf{P}_L^{a(F)} + \mathbf{P}_L^{a(F)}(\mathbf{Z}_L^A)^{\mathrm{T}}\left[\boldsymbol{\Omega}_L^A - (\mathbf{Z}_L^A)^{\mathrm{T}}\mathbf{P}_L^{a(F)}(\mathbf{Z}_L^A)^{\mathrm{T}}\right]^{-1}\mathbf{Z}_L^A\mathbf{P}_L^{a(F)}\\
&= \mathbf{P}_L^{a(F)} + \mathbf{P}_L^{a(F)}(\mathbf{Z}_L^A)^{\mathrm{T}}(\boldsymbol{\Omega}_L^A)^{-1}\left[\mathbf{I} - (\mathbf{Z}_L^A)^{\mathrm{T}}\mathbf{P}_L^{a(F)}(\mathbf{Z}_L^A)^{\mathrm{T}}(\boldsymbol{\Omega}_L^A)^{-1}\right]^{-1}\mathbf{Z}_L^A\mathbf{P}_L^{a(F)}\\
&= \mathbf{P}_L^{a(F)}\left[\mathbf{I} + (\mathbf{Z}_L^A)^{\mathrm{T}}(\boldsymbol{\Omega}_L^A)^{-1}\left[\mathbf{I} - \mathbf{S}_L^{AA}\right]^{-1}\mathbf{Z}_L^A\mathbf{P}_L^{a(F)}\right],
\end{aligned}$$

where we used $\mathbf{P}_L^{a(F)} = \mathbf{X}_L^b\widetilde{\mathbf{P}}_L^{a(F)}(\mathbf{X}_L^b)^{\mathrm{T}}$ (equation (12) of the main text) to obtain $\mathbf{S}_L^{AA}$. The matrix $\mathbf{S}_L^{AA}$ is the influence matrix of augmented observations on the analyzed augmented observations in equation (11) of the main text. The vector of regression parameters for the COSMIC system becomes

$$\boldsymbol{\beta}_L^C = \mathbf{P}_L^{a(F)}\left[\mathbf{I} + (\mathbf{Z}_L^A)^{\mathrm{T}}(\boldsymbol{\Omega}_L^A)^{-1}\left[\mathbf{I} - \mathbf{S}_L^{AA}\right]^{-1}\mathbf{Z}_L^A\mathbf{P}_L^{a(F)}\right](\mathbf{Z}_L^C)^{\mathrm{T}}(\boldsymbol{\Omega}_L^C)^{-1}\mathbf{z}_L^C.$$

Making use of identities (C.6), we rewrite equation (C.2) as $\boldsymbol{\beta}_L^F = \mathbf{P}_L^{a(F)}(\mathbf{Z}_L^F)^{\mathrm{T}}(\boldsymbol{\Omega}_L^F)^{-1}\mathbf{z}_L^F$. and then we apply identity (C.8). The difference between the regression parameters for the full and COSMIC-only systems is given by

$$\begin{aligned}
\boldsymbol{\beta}_L^F - \boldsymbol{\beta}_L^C &= \mathbf{P}_L^{a(F)}(\mathbf{Z}_L^A)^{\mathrm{T}}(\boldsymbol{\Omega}_L^A)^{-1}\left[\mathbf{z}_L^A - \left[\mathbf{I} - \mathbf{S}_L^{AA}\right]^{-1}\mathbf{Z}_L^A\mathbf{P}_L^{a(F)}(\mathbf{Z}_L^C)^{\mathrm{T}}(\boldsymbol{\Omega}_L^C)^{-1}\mathbf{z}_L^C\right]\\
&= \mathbf{S}_L^{XA}\left[\mathbf{z}_L^A - \left[\mathbf{I} - \mathbf{S}_L^{AA}\right]^{-1}\mathbf{Z}_L^A\mathbf{P}_L^{a(F)}(\mathbf{Z}_L^C)^{\mathrm{T}}(\boldsymbol{\Omega}_L^C)^{-1}\mathbf{z}_L^C\right]\\
&= \mathbf{S}_L^{XA}\left[\mathbf{I} - \mathbf{S}_L^{AA}\right]^{-1}\left[\left[\mathbf{I} - \mathbf{S}_L^{AA}\right]\mathbf{z}_L^A - \mathbf{Z}_L^A\mathbf{P}_L^{a(F)}(\mathbf{Z}_L^C)^{\mathrm{T}}(\boldsymbol{\Omega}_L^C)^{-1}\mathbf{z}_L^C\right],
\end{aligned}$$

where $\mathbf{S}_L^{XA}$ is the influence matrix of augmented observations on the analyzed components of the state vector at the grid point L, given in equation (10) of the main text. Making use of identity (C.7) again to rewrite $(\mathbf{Z}_L^C)^{\mathrm{T}}(\mathbf{\Omega}_L^C)^{-1}\mathbf{z}_L^C$, we obtain (after cancellation)

$$\begin{aligned}
\boldsymbol{\beta}_L^F - \boldsymbol{\beta}_L^C &= \mathbf{S}_L^{XA}\left[\mathbf{I} - \mathbf{S}_L^{AA}\right]^{-1}\left[\mathbf{z}_L^A - \mathbf{Z}_L^A\mathbf{P}_L^{a(F)}(\mathbf{Z}_L^F)^{\mathrm{T}}(\mathbf{\Omega}_L^F)^{-1}\mathbf{z}_L^F\right] \\
&= \mathbf{S}_L^{XA}\left[\mathbf{I} - \mathbf{S}_L^{AA}\right]^{-1}\left[\mathbf{z}_L^A - \mathbf{Z}_L^A\boldsymbol{\beta}_L^F\right] \\
&= \mathbf{S}_L^{XA}\left[\mathbf{I} - \mathbf{S}_L^{AA}\right]^{-1}\mathbf{r}_L^A, \qquad \text{(C.9)}
\end{aligned}$$

where $\mathbf{r}_L^A$ is the vector of analysis residuals for the analyzed augmented observations.

C.0.2 Efficient Computation of the Analysis Adjustment

The computation of the analysis adjustment as shown in equation (C.9) requires the inversion of the $\ell_A \times \ell_A$ matrix, $\left[\mathbf{I} - \mathbf{S}_L^{AA}\right]^{-1} = \left[\mathbf{I} - (\mathbf{R}_L^A)^{-1}\mathbf{Y}_L^A\widetilde{\mathbf{P}}^{a(F)}(\mathbf{Y}_L^A)^{\mathrm{T}}\right]^{-1}$, where ℓ_A is the number of augmented observations. This operation is repeated for all grid points that are influenced by augmented observations, so its calculation may be computationally expensive if ℓ_A is large. In this section, we present an equivalent formulation that may be more computationally efficient. We begin by writing

$$\left[\mathbf{I} - (\mathbf{R}_L^A)^{-1}\mathbf{Y}_L^A\widetilde{\mathbf{P}}^{a(F)}(\mathbf{Y}_L^A)^{\mathrm{T}}\right]^{-1} = \left[\mathbf{R}_L^A - \mathbf{Y}_L^A\widetilde{\mathbf{P}}^{a(F)}(\mathbf{Y}_L^A)^{\mathrm{T}}\right]^{-1}\mathbf{R}_L^A. \qquad \text{(C.10)}$$

The Woodbury matrix identity yields

$$\begin{aligned}
&\left[\mathbf{R}_L^A - \mathbf{Y}_L^A\widetilde{\mathbf{P}}_L^{a(F)}(\mathbf{Y}_L^A)^{\mathrm{T}}\right]^{-1} \\
&= (\mathbf{R}_L^A)^{-1} + (\mathbf{R}_L^A)^{-1}\mathbf{Y}_L^A\left[\left(\widetilde{\mathbf{P}}_L^{a(F)}\right)^{-1} - (\mathbf{Y}_L^A)^{\mathrm{T}}(\mathbf{R}_L^A)^{-1}\mathbf{Y}_L^A\right]^{-1}(\mathbf{Y}_L^A)^{\mathrm{T}}(\mathbf{R}_L^A)^{-1} \\
&= (\mathbf{R}_L^A)^{-1} + (\mathbf{R}_L^A)^{-1}\mathbf{Y}_L^A\left[\alpha_L^{-1}(k-1)\mathbf{I} + (\mathbf{Y}_L^C)^{\mathrm{T}}(\mathbf{R}_L^C)^{-1}\mathbf{Y}_L^C\right]^{-1}(\mathbf{Y}_L^A)^{\mathrm{T}}(\mathbf{R}_L^A)^{-1} \\
&= (\mathbf{R}_L^A)^{-1} + (\mathbf{R}_L^A)^{-1}\mathbf{Y}_L^A\widetilde{\mathbf{P}}_L^{a(C)}(\mathbf{Y}_L^A)^{\mathrm{T}}(\mathbf{R}_L^A)^{-1},
\end{aligned}$$

where we make use of the identity
$(\mathbf{Y}_L^F)^{\mathrm{T}}(\mathbf{R}_L^F)^{-1}\mathbf{Y}_L^F = (\mathbf{Y}_L^C)^{\mathrm{T}}(\mathbf{R}_L^C)^{-1}\mathbf{Y}_L^C + (\mathbf{Y}_L^A)^{\mathrm{T}}(\mathbf{R}_L^A)^{-1}\mathbf{Y}_L^A$. Thus equation (C.10) becomes

$$\left[\mathbf{I} - \mathbf{S}^{AA}\right]^{-1} = \mathbf{I} + (\mathbf{R}_L^A)^{-1}\mathbf{Y}_L^A\widetilde{\mathbf{P}}_L^{a(C)}(\mathbf{Y}_L^A)^{\mathrm{T}}. \qquad \text{(C.11)}$$

$\widetilde{\mathbf{P}}_L^{a(C)}$ is constructed exactly as the covariance matrix in equation (4) in the main text, but using only the ℓ_C observations from the COSMIC network. Computing $\widetilde{\mathbf{P}}_L^{a(C)}$ requires the inversion of a $k \times k$ matrix, where k is the ensemble. This formulation yields a more efficient computation when the ensemble size is smaller than the number of augmented observations, which can be expected in an operational setting. In the

case that there are only augmented observations in the analysis calculation at a grid point L, then $\widetilde{\mathbf{P}}_L^{a(C)} = \alpha_L(k-1)^{-1}\mathbf{I}$, and

$$\left[\mathbf{I} - \mathbf{S}^{AA}\right]^{-1} = \mathbf{I} + \alpha_L(k-1)^{-1}(\mathbf{R}_L^A)^{-1}\mathbf{Y}_L^A(\mathbf{Y}_L^A)^{\mathrm{T}}, \tag{C.12}$$

which does not require any matrix inversions. In this case, the analysis adjustment given by Equation (C.9) is simply the difference between the analysis and the background at L.

Appendix D

INITIAL APPLICATION OF THE LETKF TO AN IONOSPHERIC GLOBAL CIRCULATION MODEL

D.1 Initial Application of the LETKF to an Ionospheric Global Circulation Model

This section presents initial numerical experiments, where the local ensemble transform Kalman filter (LETKF) is used to assimilate synthetic observations into a global circulation model of the ionosphere called the thermosphere-ionosphere-electrodynamics general circulation model (TIEGCM). The assimilated observations are of synthetic electron density observations, whose locations are given by a realistic satellite-derived observing network, into ionospheric forecasts made with the TIEGCM. We employ a "perfect" model assumption in which the "true" state of the I-T system is computed with a TIEGCM simulation. The ionospheric state of this control simulation is then "observed" at the observation locations to estimate the unknown spatiotemporal evolution of the three dimensional "true" state of the ionosphere and its dynamical drivers.

The system used to carry out the observing system experiments is comprised of the forecast model and data assimilation scheme, which are the TIEGCM and LETKF, respectively. We refer to this data assimilation system as the TIEGCM-LETKF system. The primary goal of this section is to describe the TIEGCM-LETKF data assimilation system that will be used in the forecast experiments, give an overview of the general experiment set up, and present the initial results. The same data assimilation system is used in the remaining chapters of this thesis.

The observing system experiments are carried out with synthetic observations that are generated from a TIEGCM simulation, which we regard as the "true" state of the ionosphere. With this perfect model assumption, we test the LETKF in its skill to adjust the forecast towards some unknown "true" state trajectory, and errors that arise are solely due to the performance of the data assimilation system. In reality, data assimilation systems must be able to cope with errors in the model as well as in the data as they can contribute significantly to forecast uncertainty, but these perfect model experiments allow us to separate the two issues for this initial investigation.

In the synthetic experiments we make a direct comparison of the analysis with the truth to obtain a global distribution of the error structures. In particular, we are interested in the approximation of important ionospheric descriptors, such as the electron density peak in the F_2 (NmF2) layer and its elevation (HmF2), and also quantify the improvement of ionospheric drivers, including multiple unobserved thermospheric state variables, and the solar and geomagnetic parameterizations.

D.1.1 Ionospheric Forecast Model

The model used in this study is the Thermosphere-Ionosphere-Electrodynamics Global Circulation Model (TIEGCM), which is a first principles, three dimensional, non-linear representation of the coupled ionosphere-thermosphere (I-T) system. The TIEGCM self-consistently solves the fully coupled, non-linear, hydrodynamic, thermodynamic, and continuity equations of neutral gas, the ion and electron energy and momentum equations, the ion continuity equation, and solves for neutral wind dynamo. The coordinate system is spherical and is fixed with respect to the rotating Earth, with longitude and latitude as the horizontal coordinates and pressure surface as the vertical coordinate. The pressure surfaces are defined as $z = \ln(P_0/P)$, where P_0 is a reference pressure at 5×10^{-5} Pa. The model has standard spatial resolution of

$5^\circ \times 5^\circ$ in the horizontal direction, covering -87.5° to 87.5° in latitude and -180° to 180° in longitude. In the vertical direction there are 29 pressure surfaces ranging from -7 to 7 that extend to altitudes of about about 97 km to about 600 km depending on solar activity. The standard time step is of 2 minutes. An overview of the main equations solved is provided in Appendix E, and a comprehensive model description can be found at (http://www.hao.ucar.edu/modeling/tgcm/).

D.2 TIEGCM-LETKF Data Assimilation System

Our TIEGCM-LETKF data assimilation system runs as a forecasting step, which estimates the state of the ionosphere-thermosphere (I-T) system, using the TIEGCM, up to a time $t_n + \Delta t$, followed by an analysis step, in which observations in the time window $[t_n - \Delta t, t_n + \Delta t]$ are assimilated using the LETKF. In all data assimilation experiments presented in this thesis, the radius of the analysis window is taken to be $\Delta t = 0.5$ hrs, as has been done in other I-T data assimilation systems (Lee *et al.*, 2012) (Matsuo *et al.*, 2013). The result of the analysis step is an updated I-T state estimate at the time t_n that serves as the initial condition for the subsequent forecasting step. This forecast/analysis sequence is repeated at 1-hr intervals for the duration of the simulation. The benefits of employing forecast/analysis time steps of less than 1 h have been suggested by (Chen *et al.*, 2016), but we do not consider them for our data assimilation experiments. The specific details regarding the dates of the simulations will be provided in the subsections that describe the forecast simulations. We describe the details of the TIEGCM-LETKF system below.

D.2.1 Forecast Initialization

The initialization phase of the data assimilation system consists of a simulation that produces a "spin-up" of a forecast, which is used as the background estimate for the first analysis step. In the TIEGCM-LETKF system, the initial condition for the spin-up is obtained from a climatology simulation of the ionosphere-thermosphere (I-T) system that is created with the TIEGCM by the High Altitude Observatory (HAO). The climatology simulations are a representation of the average conditions of the I-T system throughout its seasonal cycle. To account for year-to-year variabilities, solar and geomagnetic indices published by NOAA are used to drive the simulation towards a state that is representative of the solar and geomagnetic conditions of the period of interest. This approach is taken to generate the control simulation that is regarded as the "true" state of the I-T system.

This approach is also taken to generate the ensemble of k background forecasts used in the data assimilation cycle. Each of the k forecasts begin from the same initial condition, as given by the climatological simulations discussed above, and is integrated with k normally distributed solar and magnetospheric parameterized inputs to the TIEGCM. The center of the parameter distribution for the solar and magnetospheric parameters are taken to be those published by NOAA (also used for the control simulation) during the time period of the simulations, but we also consider centering the parameter distribution on other values. The standard deviation for the distribution of each parameter is taken to be the standard deviation of each parameter over the duration of the spin-up period. Specific values for the distribution mean and standard deviation for each parameter are given during the discussion of individual

numerical experiments.

This approach yields an ensemble of k statistically equivalent forecasts, whose variability covers a wide range of representative I-T states that for the time period of interest. A similar approach for ensemble generation is used in wide variety of ionospheric data assimilation experiments. See for example, (Lee *et al.*, 2012; Matsuo and Araujo-Pradere, 2011; Matsuo *et al.*, 2013; Hsu *et al.*, 2014; Chen *et al.*, 2016).

D.2.2 Observation Operator

The observation operator is used to relate quantities defined on the model grid to the observed quantities during each analysis calculation. In the context of the LETKF, the input of the observation operator is a vector of the relevant state variables defined at the grid points located within the local region centered around a given grid point, and the output consists of model predictions for the observed quantities at their corresponding spatiotemporal location. In our TIEGCM-LETKF system, the only observed variable is electron density, which is also a state variable, so the observation operator consists of an interpolation procedure to the locations and times of the observed electron density vertical profiles. The location of the observed electron density vertical profiles are given horizontally in terms of longitude and latitude, and the vertical coordinate is in terms of kilometers.

The horizontal coordinates of the TIEGCM model grid are also defined in terms of longitude and latitude. However, the vertical coordinate is defined in terms of pressure levels, as discussed in Section D.1.1. The corresponding altitude for each grid point has dependence on several state variables (i.e temperature), so it varies with location, time and is different for each forecast in the ensemble. Figure D.1(a) shows the horizontal structure of the geometric altitudes at the fixed pressure level 3.0, as described in Section D.1.1 at 11:30 UTC on 26 September 2011. Generally, the day-time ionosphere is located at higher altitude by about 50 km for fixed pressure level. Figure D.1(b) shows the vertical structure of the geometric altitudes at 17.5°N latitude. Also shown are level curves corresponding to fixed altitudes of 200 km, 350 km and 500 km.

The interpolation of electron density to the observation locations is done as follows. For a given vertical profile, find the model longitude and latitude coordinates which contain it. Within these coordinates, interpolate the model altitudes to a fixed altitude grid for each ensemble member, with the altitude chosen to be the minimum among the ensemble members (to avoid extrapolation at the top of the domain), and discard all observations located above. The variability of altitude among ensemble members is generally about 1%. From this fixed-altitude grid, electron density is interpolated to the observation locations using trilinear interpolation in space. Generally, interpolation in time is also needed, since all observations taken within a time window of the forecast time are assimilated. This is accomplished by applying the vertical interpolation procedure described above at multiple model times, and then linearly interpolating to the observation times. The same procedure is repeated for each of the observed vertical profile.

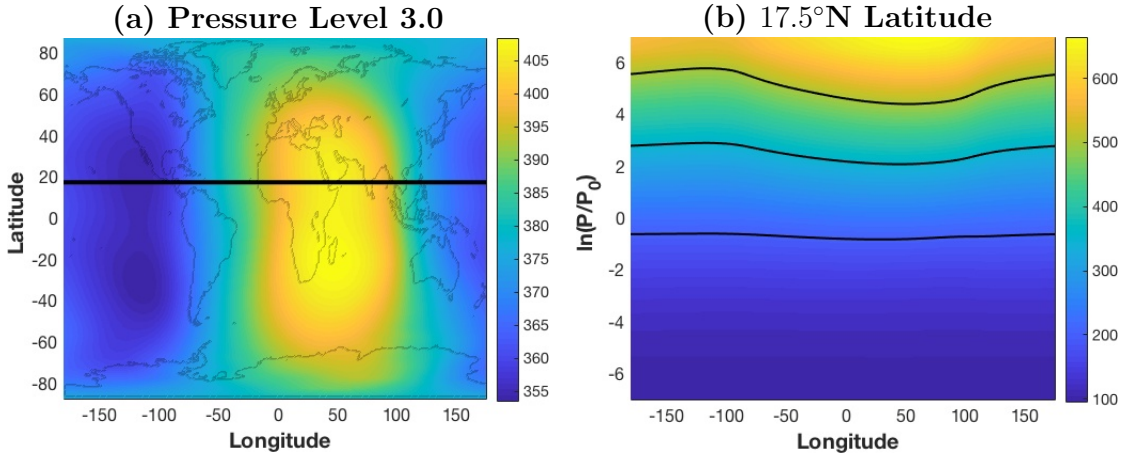

Figure D.1: (a) Global map of geometric altitude in km at model pressure level 3.0, as defined in Section D.1.1, at 11:30 UTC on 26 September 2011. (b) Vertical structure of geometric altitude at 17.5°N latitude (black curve shown in Figure (a)) at the same time. The black curves denote 200 km, 350 km, and 500 km altitudes.

D.2.3 LETKF State Vector Augmentation

The ionosphere is strongly coupled with the thermosphere, as described in Section 2.1. This is a feature that must be accounted for in forecasting of ionospheric weather forecasting. The TIEGCM is a model of the coupled I-T system, so the dynamics involving their coupling are accounted for during the forecasting stage of the data assimilation cycle. However, the benefits of incorporating the I-T coupling during the analysis calculation, by updating thermospheric state variables in addition to electron density, have been shown (Matsuo and Araujo-Pradere, 2011).

The inference of unobserved state variables is readily accomplished with the LETKF (and other ensemble-based Kalman filters), through state augmentation, as discussed in Section 3.4.3. Through the state augmentation approach, the same linear combination of local forecasts used to compute analyzed electron density is used to compute the analysis for the augmented unobserved thermospheric state variables and model parameters. Thus observations of electron density can be used to infer an analyzed state of thermospheric state variable based on the observations that are assimilated. The accuracy in the state augmentation approach depends on the electron density having strong correlation with the thermospheric variables being inferred and also this correlation being well represented with the ensemble of forecasts.

Each integration of the TIEGCM yields a 1-hr forecast of the I-T state vector. During each analysis step, a subset of the components are *analyzed*, i.e., updated using the assimilated observations. We refer to this subset as the *LETKF state vector*. The components of the TIEGCM state vector that are not included in the LETKF state vector are left unchanged during the analysis step. In our numerical experiments, the LETKF state vector is composed of electron density (N_e), neutral temperature (T_n), zonal (U_n) and meridional (V_n) components of neutral winds, as well as atomic (O_1) and molecular (O_2) mass mixing ratios. This choice of LETKF state vector is done

following the work of (Matsuo *et al.*, 2013),(Lee *et al.*, 2012). The importance of also updating ionospheric drivers, such as solar and geomagnetic parameters, during the analysis steps has been shown by (Matsuo and Araujo-Pradere, 2011), (Hsu *et al.*, 2014). The estimation of these parameterized drivers may be done using the same state augmentation approach described in Section 3.4.3, and is shown later in this section.

D.3 Initial Synthetic Observation Experiments

We generate the 'true' ionosphere state trajectory, $\{\mathbf{u}_i^t\}$, which is a TIEGCM simulation containing temporal evolution of all TIEGCM state variables at each time step t_i. The dynamical trajectory of the true system state is driven with a set of solar and magnetospheric forcing parameters, which we denote as $\{\mathbf{p}_i^t, i = 1, 2, \ldots, N\}$. For each time t_i, $\mathbf{p_i^t}$ is vector containing the F_{107}, Cp and Hp values used to drive the control simulation. A description of these parameters and their effect on ionospheric modeling is described in Section 2.1.2. Similarly, denote the set of k parameters used to drive the ensemble of background forecasts as $\{\mathbf{p}_i^{b(j)}, j = 1, 2, \ldots, k\}$, where $\mathbf{p}_i^{b(j)}$ contains the normally distributed values of $\mathrm{f}_{10.7}$, Cp and Hp for the jth forecast at the time t_i.

Synthetic observations are generated by sampling the electron density component of $\mathbf{u}_i^t$ at the times and locations of the COSMIC observing network (Rocken *et al.*, 2000), during 26 September 2011. The COSMIC electron density vertical profiles typically vary from about 80 km to 800 km in altitude, with a vertical resolution of about 10 km. About 85 profiles are available for assimilation during each hour. The COSMIC electron density profile data set is available at the COSMIC Data Analysis and Archive Center (CDAAC) at UCAR (http://cdaac-www.cosmic.ucar.edu).

Synthetic observations are generated with additive Gaussian noise to represent observation processing noise: $\mathbf{y}_i^o = \mathbf{H}_i\mathbf{u}_i^t + \boldsymbol{\epsilon}_i$, where $\mathbf{H}_i$ interpolates the electron density component of $\mathbf{u}_i^t$ to the respective observation locations and times during the ith analysis step and $\boldsymbol{\epsilon}_i$ is a Gaussian random vector with zero mean and covariance matrix $\mathbf{R}_i$. Observation errors are assumed to be independent so $\mathbf{R}_i$, is diagonal. The standard deviation of $\boldsymbol{\epsilon}$ is assumed to scale as 10% of the of the electron density component of $\mathbf{u}^t$ at the observation locations. Minimum and maximum thresholds of 1×10^3 and 1×10^4 are applied to the standard deviation in the generation of observations, following the error description provided in the CDAAC website.

D.3.1 Results: Synthetic Observation Experiments

The first row of Figure D.2 shows global maps of electron density maps in el/cm^3 at 12:30 UTC, 16:30 UTC and 20:30 UTC on 26 September 2011. The maps are averaged over 250 km to 450 km altitudes, so they cover the the F-layer. The selected times show the electron density at the onset of geomagnetic disturbance (12:30 UTC) during the main phase of the geomagnetic storm (16:30 UTC) and as geomagnetic conditions begin to relax (20:30 UTC). The second row of Figure D.2 shows maps of electron density at the same times after the observations, whose locations are shown by the magenta markings, are assimilated. The second and third rows of Figure D.2 show the deviation from the truth for the background and analyzed electron density,

respectively. Regions where the deviation from the truth are positive (negative) are shown in red(blue). Comparison of the second and third rows demonstrate that the electron density error is significantly reduced after the assimilation of observations.

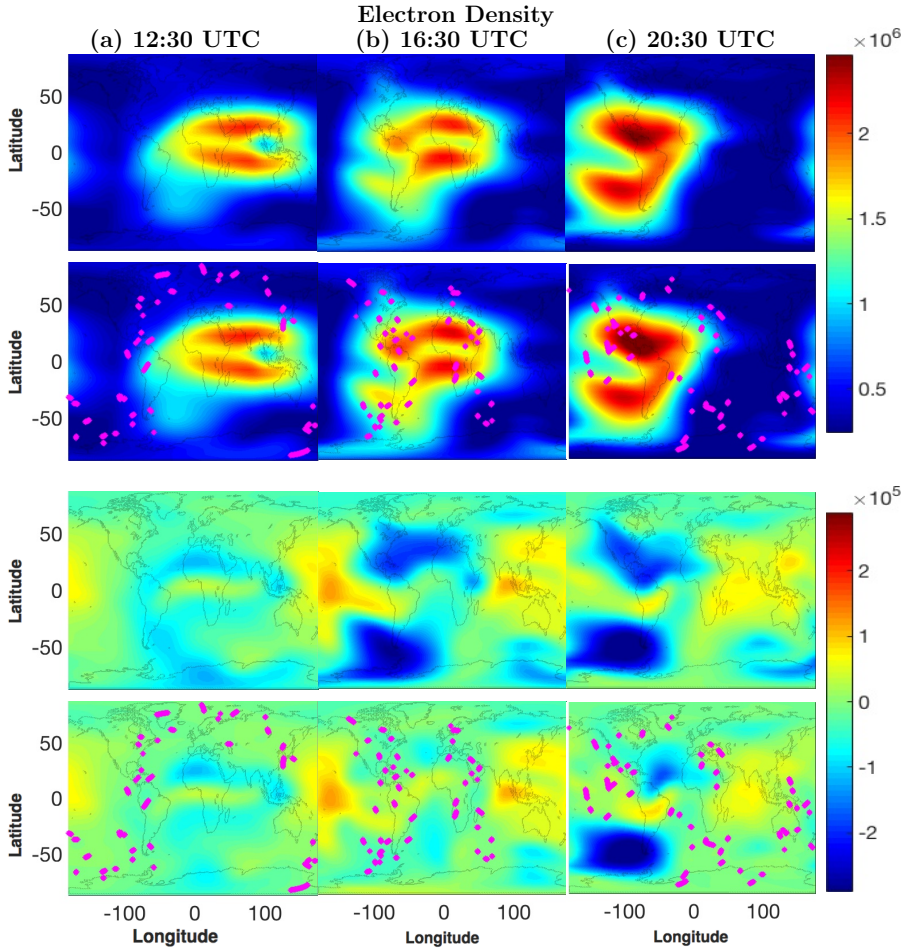

Figure D.2: Global maps of background (first row) and analyzed (second row) estimates of electron density, in el/cm^3, averaged from 200 km to 500 km altitudes at 12:30, 16:30 and 20:30 UTC on 26 September 2011. Rows 2-3: Analogous global maps of electron density deviations from the truth for the background and analyzed electron density at the same times. The analyzed electron density (second row) and its deviation from the truth (fourth row) are obtained by assimilating only synthetic COSMIC observations, whose locations are shown by magenta markings. The color scale in the first and second rows differs from the third and fourth rows.

Figure D.3 shows analogous global maps of the zonal component of neutral winds U_n in units of cm/s. These maps are shown at the same times and are averaged over the same altitudes as Figure D.2. Although only observations of electron density are used, the estimates of U_n are significantly improved. Similarly, the assimilation of electron density observations also improve the composition of atomic oxygen at the same times and altitudes.

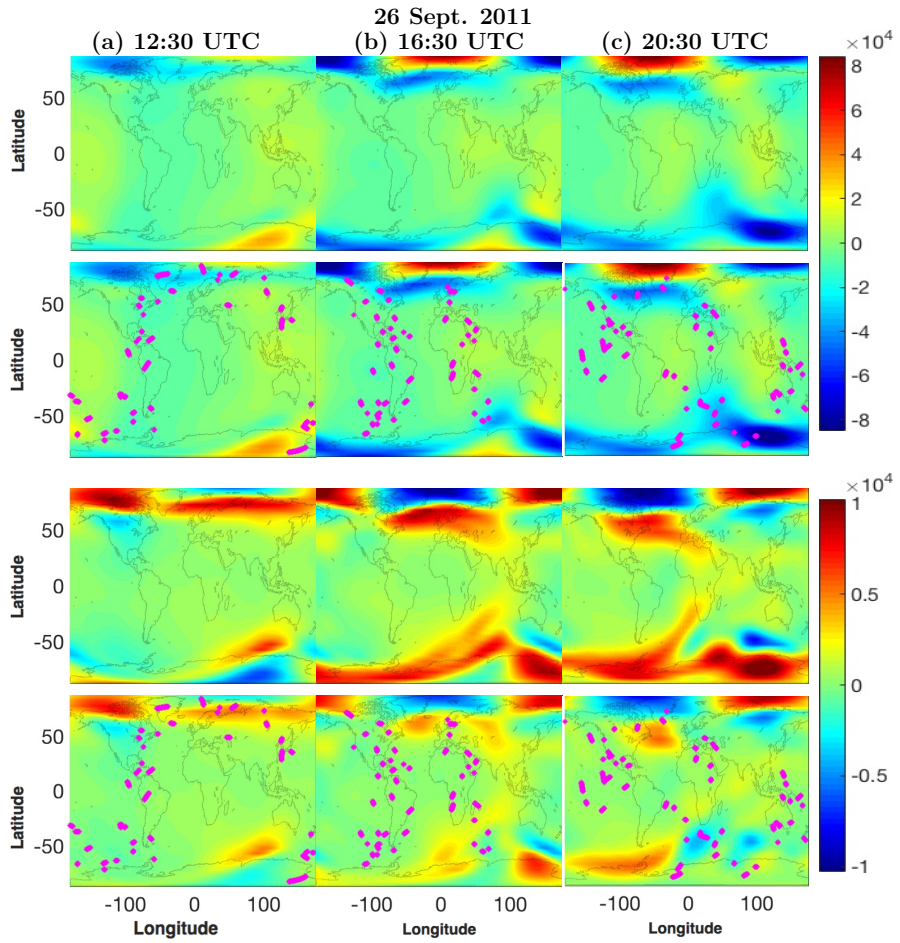

Figure D.3: Global maps of background and analyzed estimates of neutral winds (U_n), in cm/s. The times and locations are the same as the maps shown in Figure 1.

Figure D.5(a) shows time series of root mean square error of the electron density estimates, averaged over the Southern polar region, which we define to be all grid points south of 60°S latitude. The time series begins at 00:30 UTC on 26 September 2011 and ends at 23:30 UTC on 27 September 2011. The thin and thick red curves correspond to the background and analyzed electron density estimates respectively, when analyzing electron density and the other neutral state variables described in Section D.2.3. Similarly, the blue curves show the background and analyzed electron density when only electron density is analyzed. For comparison, a free run, where no observations are assimilated, is also shown in black. Results show a pronounced reduction in RMSE for the analyzed electron density estimates compared to the free run. However, during the main phase of the geomagnetic storm, between 12:30 UTC and 23:30 UTC, the background electron density estimates are worse than the free run, suggesting that some ionospheric drivers are not adjusted adequately during this time period. Figure D.5(b) shows analogous RMSE time series plots. Similar improvements in the analyzed electron density are seen throughout the simulation. The

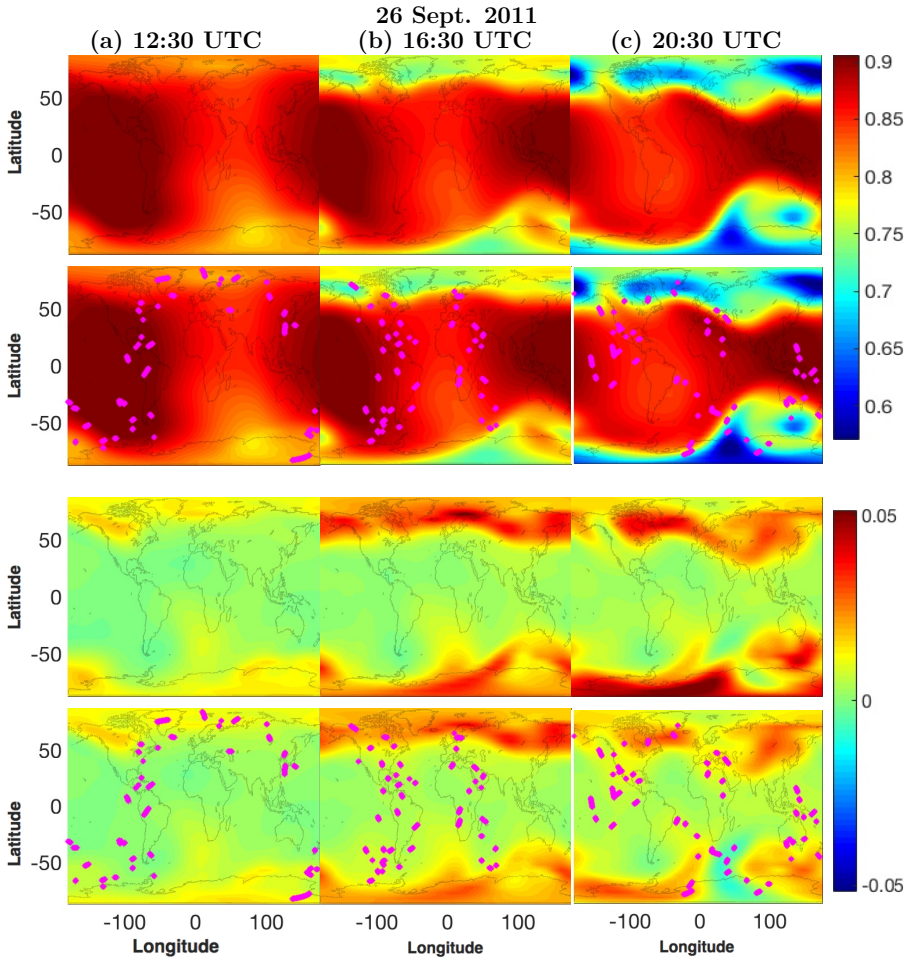

Figure D.4: Global maps of background and analyzed estimates of atomic oxygen composition, in units of mass mixing ratio (mmr). The times and locations are the same as the maps shown in Figure D.2.

background also shows significant improvement, implying that there is improvement in 1-hour predictions of electron density.

Analogous time series of RMSE for the zonal components of neutral winds (U_n) and atomic oxygen composition are shown in Figures D.6 and D.7. In both cases there are signifcant improvements in the background and analyzed estimated of each variable over both polar regions.

An important aspect of ionospheric forecasting is the correct specification of solar and magnetospheric inputs. In the TIEGCM, these inputs are specified through auxiliary parameters. As previously described, these parameters may be estimated through state augmentation approaches. Figure D.8 show the time evolution of each ionospheric forcing parameters. The "true" values correspond to the indices published by NOAA during 26-27 September 2011 and are shown with the green curves. The ensemble of forcing parameters for each respective parameter are shown in red and the ensemble means are shown in black. The parameter estimates follow the overall

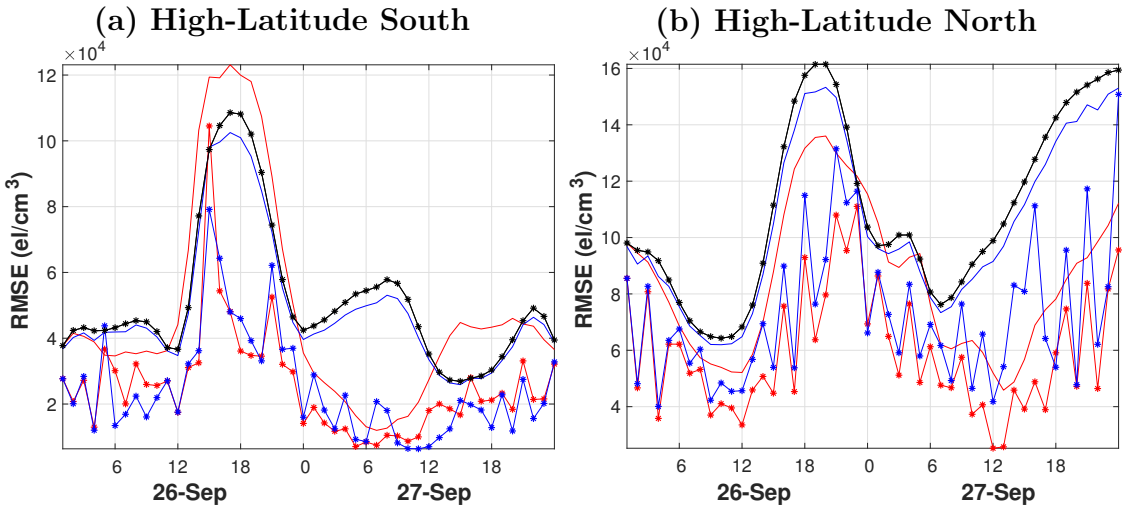

Figure D.5: (a) Time series of the root mean squared error (RMSE) for electron density estimates in units of el/cm^3 averaged over the Southern polar regions. The thin and red curves correspond to the RMSE of the background and analyzed estimates, respectively, when analyzing electron density and the other neutral state variables given in Section D.2.3. Similarly, the thin and thick curves give the RMSE for background and analyzed electron density in the case where only electron density is analyzed. Also shown is a free run, where no variables are analyzed with observations (black). (b) Analogous RMSE plots averaged over the Northern polar region. The time series begin at 00:30 UTC on 26 September 2011 and end at 23:30 UTC on 27 September 2011, with 1-hour intervals.

trend of the temporal evolution, particularly during the geomagnetically disturbed periods, although these estimates are relatively noisy.

Discussion

The synthetic observation experiments presented in this section aimed at studying the skill of the LETKF system in forecasting ionospheric electron density in a perfect model scenario. This constrains forecasting errors to arise only from the data assimilation system itself. Although data assimilation systems must be able to cope with other issues, such as model bias and non-Gaussian observation errors with missing information, the perfect model experiments allows us to isolate the issues.

In this framework, the LETKF is able to skillfully estimate and forecast the state of the electron density field using a moderately size ensemble of 40 members. The assimilation of the synthetic satellite observations produced analysis estimates that were a significant improvement over the background estimates throughout the entire model domain. However the analysis estimates were not perfect, and small uncertainties in these initial conditions and in the forcing parameters caused uncertainties to grow over the subsequent forecasting step. The forecast deviation from the truth was continually corrected by assimilating more observations, and as a result, the errors in the forecasted states were always significantly smaller compared to the free run,

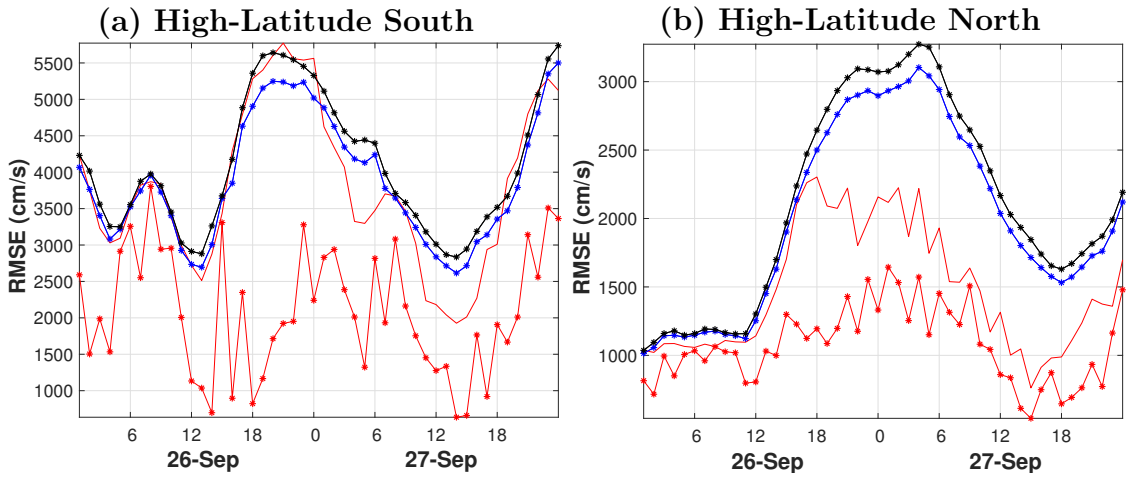

Figure D.6: (a)-(c) Analogous RMSE time series to the ones showed in Figure D.5, but for the zonal components of the neutral winds U_n in units of el/cm.

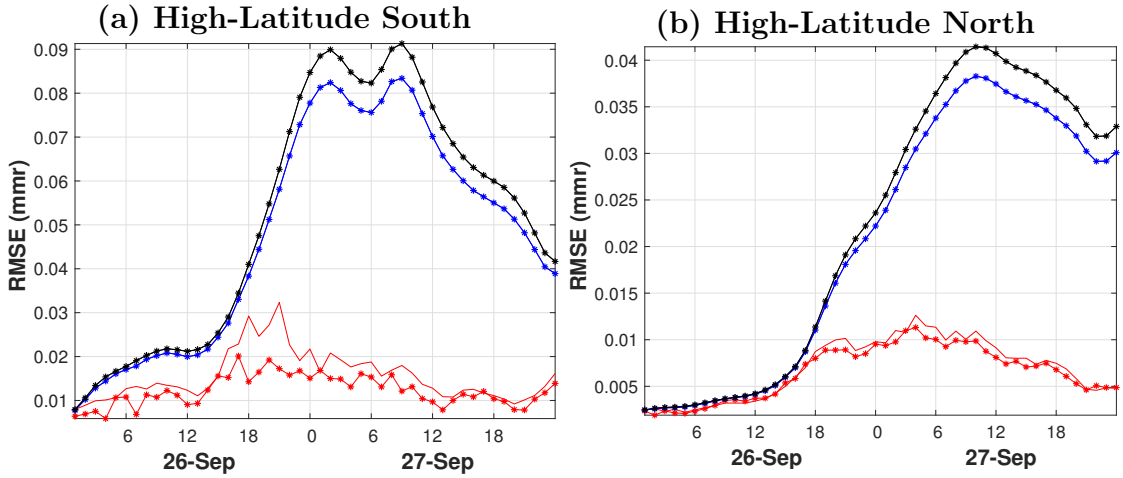

Figure D.7: (a)-(c) Analogous time series to the ones showed in Figure D.5, but for the atomic oxygen composition O_1 in units of mass mixing ratio (MMR).

where no observations were assimilated.

Although forecasting results remained stable and accurate even when using a noisy and sparse observing network, it must be noted that several challenges existing in real numerical weather forecasting applications were not considered in these experiments. For example, we did not consider model bias here, but it is a significant source of forecast uncertainty and non-trivial to overcome. Additionally, the observations were mutually independent and Gaussian as expected in the in the Kalman filter, and their representation within the LETKF was perfectly known. In practice, the statistical properties of the observations are not known accurately and are often not Gaussian. Additionally, they are often correlated and this is not captured by a diagonal obser-

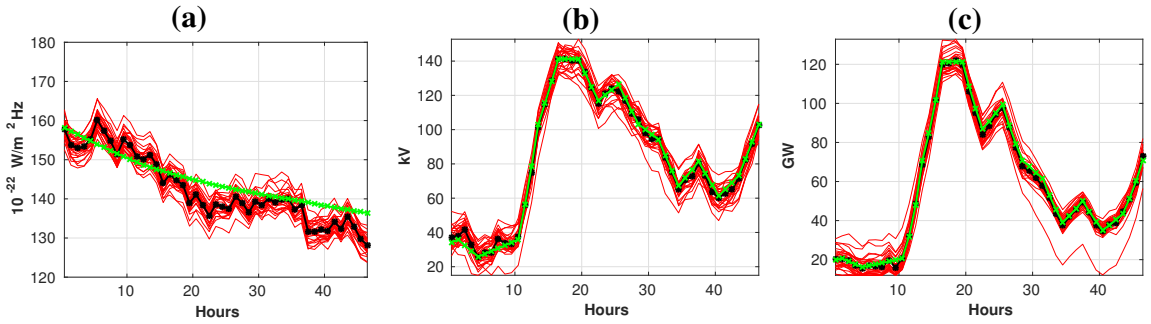

Figure D.8: Timeseries of analyzed $F_{10.7}$ (a), Cp (b) and Hp (c) values obtained during the simulation of the 26-27 September 2011 geomagnetic storm. The horizontal axis shows time in hours, starting from 00:30 UTC on 26 September 2011 and ending at 23:30 UTC on 27 September 2011. In each of the figure, the red curves correspond to the ensemble members and the black lines are the respective ensemble means. The parameters used to drive the "truth" are given by the green curves and correspond to the actual indices published by NOAA for this time period.

vation covariance matrix, **R**.

Appendix E

MAIN EQUATIONS SOLVED IN THE TIEGCM

E.1 TIEGCM: Main Equations

The TIEGCM self-consistently solves the fully coupled, non-linear, hydrodynamic, thermodynamic, and continuity equations of neutral gas, the ion and electron electron energy and momentum equations, the ion continuity equation, and solves for neutral wind dynamo. The full mathematical descriptions of the PDE's solved in the TIEGCM code are available in the model description document provide at High Altitude Observatory (HAO). A full description of the neutral dynamics in the model description available at the HAO website:
`http://www.hao.ucar.edu/modeling/tgcm/doc/description/model_description.pdf`

E.1.1 Horizontal Momentum Equations

The zonal and meridional winds are solved from the momentum equations.
The zonal momentum equations have the form

$$\begin{aligned}\frac{\partial u_n}{\partial t} =& \frac{ge^z}{p_0}\frac{\partial}{\partial Z}\left[\frac{\mu\partial u_n}{H\partial Z}\right] + f^{cor}v_n + \lambda_{xx}(v_{ExB,x} - u_n) + \lambda_{xy}(v_{ExB,y} - v_n) \\ & - v_n\cdot\nabla u_n + \frac{\partial u_n v_n}{R_E}\tan(\lambda) - \frac{1}{R_E\cos(\lambda)}\frac{\partial\Phi}{\phi} - W\frac{\partial u_n}{\partial Z} - hd_u\end{aligned} \tag{E.1}$$

The meridional momentum equations have the form

$$\begin{aligned}\frac{\partial v_n}{\partial t} =& \frac{ge^z}{p_0}\frac{\partial}{\partial Z}\left[\frac{\mu\partial v_n}{H\partial Z}\right] + f^{cor}v_n + \lambda_{yy}(v_{ExB,x} - v_n) + \lambda_{yx}(v_{ExB,y} - u_n) \\ & - v_n\cdot\nabla v_n + \frac{\partial u_n u_n}{R_E}\tan(\lambda) - \frac{1}{R_E\cos(\lambda)}\frac{\partial\Phi}{\phi} - W\frac{\partial v_n}{\partial Z} - hd_v\end{aligned} \tag{E.2}$$

The meaning of each terms for both equations are as follows:

$$\begin{aligned}\frac{\partial u_n}{\partial t} =& \text{ vertical viscosity } + \text{ coriolis force } + \text{ ion drag } + \text{ horizontal advection} \\ & + \text{momentum } + \text{ pressure gradient } + \text{ vertical advection } + \text{ horizontal diffusion}\end{aligned} \tag{E.3}$$

Leap frog is used for time stepping of momentum transport, geopotential gradient, Coriolis force, and ion drag, while an implicit scheme is used for eddy and molecular viscosity in the vertical direction. A Shapiro filter is also used for numerical stability.

E.1.2 Vertical Momentum Equations

The vertical velocity is calculated by solving the continuity equation of the thermospheric neutral gas. The continuity equation has the form

$$\frac{1}{R\cos\lambda}\frac{\partial}{\partial\lambda}(v_n\cos\lambda) + \frac{1}{R\cos\lambda}\frac{\partial u_n}{\partial\phi} + e^z\frac{\partial}{\partial Z}(e^{-z}W) = 0 \tag{E.4}$$

where the vertical velocity is given by $W = \frac{\partial Z}{\partial t}$. The vertical velocity relative to a pressure level is the obtained by integrating the continuity equation over Z to get W, and hen multiply W by the scale height H.

E.1.3 Thermodynamics Equation

Neutral temperature is obtained by solving the equation of thermodynamics:

$$\begin{aligned}\frac{\partial T_n}{\partial t} =& \frac{ge^z}{p_0 C_p}\frac{\partial}{\partial Z}\left\{\frac{K_T}{H}\frac{\partial T_n}{\partial Z} + K_E H^2 C_p \rho\left[\frac{g}{C_p} + \frac{1}{H}\frac{\partial T}{\partial Z}\right]\right\} - v_n \cdot \nabla T_n \\ & - W\left(\frac{\partial T_n}{\partial Z} + \frac{RT_n}{C_p m}\right) + \frac{Q^{exp} - e^z L^{exp}}{C_p} - L^{imp}\end{aligned} \tag{E.5}$$

Where C_p is the specific heat per unit mass, p_0 is the reference pressure, K_T is the molecular thermal conductivity, H is the pressure scale height, K_E is the eddy diffusion coefficient, ρ is atmospheric mass density, V_n is the horizontal neutral velocity with the zonal and meridional components u_n and v_n, W is the dimensionless vertical velocity given by $W = \frac{\partial Z}{\partial t}$, R is the universal gas constant, m is the mean atmospheric mass, and Q and L are the other heating and cooling terms.

Major species continuity equations
These equations have the form

$$\frac{\partial \Psi_O}{\partial t} or \frac{\partial \Psi_{O_2}}{\partial t} = \text{vertical molecular diffusion} + \text{vertical eddy diffusion} + \text{horizontal/vertical advection} + \text{chemical production loss} + \Psi_{N_2} = 1 - \Psi_O - \Psi_{O_2}$$

This equation is solved to obtain the species Ψ_O, Ψ_{O_2} and Ψ_{N_2} .

Minor species continuity equations
These equations have the form

$$\frac{\partial \psi_O}{\partial t} = \text{vertical molecular diffusion} + \text{vertical eddy diffusion} + \text{horizontal/vertical advection} + \text{chemical production loss,} + \text{chemical production}$$

This equation is solved to obtain the species $\psi_{N(4S)}, \psi_{NO}$

Thermodynamic equation (energy equation)
This equation has the form

$$\frac{\partial T_n}{\partial t} = \text{vertical molecular and eddy heat conduction} + \text{horizontal/vertical advection} + \text{heating (solar, geomagnetic and collisions)} + \text{radiative cooling}$$

This equation is solved to obtain neutral temperature T_n

Electron energy equation assuming a thermal quasi-stable state
This equation has the form

$$\text{divergence of electron heat} + \text{heating (solar radiance, auroral particle precipitation)} = \text{cooling due to collisions with ion and neutrals}$$

This equation is solved to obtain electron temperature T_e

Ion energy equation assuming a thermal quasi-stable state
This equation has the form

$$\begin{aligned}&\text{Heating due to electron-ion collisions } + \text{ Joule heating}\\ &= \text{cooling due to collisions with ion and neutrals}\end{aligned}$$

This equation is solved to obtain ion temperature T_i

Steady state electro-dynamo equation
This equation has the form

$$\begin{aligned}&\text{horizontal divergence of height-integrated current density}-\\ &\text{field-aligned current from the magnetosphere}\end{aligned}$$

This equation is solved to obtain the electric potential Φ

Ion Continuity equation
This equation has the form

$$\begin{aligned}\frac{\partial n_O}{\partial t} &= \text{transport by neutral wind } + \text{ transport by ambipolar diffusion}\\ &+ \text{transport by} E \times B + \text{ chemical preduction } + \text{ chemical loss}\end{aligned}$$

This equation is solved to obtain the ion density n_O

Appendix F

REFORMULATION OF KALMAN FILTER ANALYSIS EQUATIONS

F.1 Kalman Filter Reformulation

This is the reformulation to go from equation 3.14 and 3.15 and :

$$\mathbf{P}^a = \left[(\mathbf{P}^b)^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\right]^{-1} \tag{F.1}$$

$$\mathbf{x}^a = \mathbf{P}^a\left[(\mathbf{P}^b)^{-1}\mathbf{x}^b + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{y}^o\right] \tag{F.2}$$

To equations 3.17 and 3.18.

$$\mathbf{P}^a = \left[\mathbf{I}_n + \mathbf{P}^b\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\right]^{-1}\mathbf{P}^b \tag{F.3}$$

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{P}^a\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\left(\mathbf{y}^o - \mathbf{H}\mathbf{x}^b\right) \tag{F.4}$$

Where $\mathbf{I}_n$ is the identity matrix of dimension $n \times n$ and n is the number of grid points. To obtain equation F.4, we solve equation F.1 for $(\mathbf{P}^b)^{-1}$ and performing some algebraic manipulations:

$$\begin{aligned}
\mathbf{x}^a &= \mathbf{P}^a\left[(\mathbf{P}^b)^{-1}\mathbf{x}^b + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{y}^o\right] \\
&= \mathbf{P}^a\left[\left[(\mathbf{P}^a)^{-1} - \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\right]\mathbf{x}^b + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{y}^o\right] \\
&= \left[\mathbf{I}_n - \mathbf{P}^a\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\right]\mathbf{x}^b + \mathbf{P}^a\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{y}^o \\
&= \mathbf{x}^b + \mathbf{P}^a\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\left(\mathbf{y}^o - \mathbf{H}\mathbf{x}^b\right)
\end{aligned}$$

Introducing the Kalman gain matrix, $\mathbf{K} = \mathbf{P}^a\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}$, we get the desired result. To obtain equation F.3, we right multiply equation F.1 by $(\mathbf{P}^b)^{-1}\mathbf{P}^b$ and perform some algebraic manipulations

$$\begin{aligned}
\mathbf{P}^a &= \left[(\mathbf{P}^b)^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\right]^{-1} \\
&= \left[(\mathbf{P}^b)^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\right]^{-1}(\mathbf{P}^b)^{-1}\mathbf{P}^b \\
&= \left[(\mathbf{P}^b)\left((\mathbf{P}^b)^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\right)\right]^{-1}\mathbf{P}^b \\
&= \left[\mathbf{I}_n - \mathbf{P}^b\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\right]^{-1}\mathbf{P}^b
\end{aligned}$$