CARBON DATA ASSIMILATION USING AN ENSEMBLE KALMAN FILTER

NAN MIAO

A THESIS SUBMITTED TO
THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF

MASTER OF SCIENCE

GRADUATE PROGRAM IN EARTH AND SPACE SCIENCE
YORK UNIVERSITY
TORONTO, ONTARIO

May 2014

Abstract

As a first step to build an ensemble data assimilation and source inversion system for atmospheric carbon, I implemented column-integrated carbon monoxide (CO) mixing ratio assimilation capability in an ensemble Kalman filter (EnKF) data assimilation system with the Weather Research and Forecasting model coupled with Chemistry (WRF-Chem). In spite of its global coverage, the CO retrievals from the Measurements Of Pollution In The Troposphere (MOPITT) instrument onboard the Terra satellite are available only once per day. There has been restricted use of these CO data for atmospheric chemistry forecasting. Data assimilation provides an effective way to guide the model in time.

This WRF-Chem/EnKF system has been tested for a real forest fire case in British Columbia in 2010. It has been observed that after assimilating MOPITT data, the model has been constrained closer toward the observations and the root-mean-square errors (RMSE) between the forecasts and the observations have been reduced. An inverse modeling of CO sources using parameter estimation with an EnKF was also performed. Comparisons of the assimilated CO profiles with optimal emissions to observations indicate that the assimilation leads to a considerable improvement of the model simulations as compared with a control run with no assimilation. Model biases in the simulation of background values are reduced and an improvement in the simulation of very high concentrations is observed.

Acknowledgements

This thesis would not have been possible without the guidance and the help of several individuals who in one way or another contributed and extended their valuable assistance in the preparation and completion of this study. I would like to extend my appreciation especially to the following.

Foremost, I would like to express my sincere gratitude to my advisor Dr. Yongsheng Chen for the continuous support of my study and research, for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me in all the time for research and writing of this thesis. I could not have imagined having a better advisor and mentor for my master study.

Besides my advisor, I would like to express my gratitude to the rest of my thesis committee: Dr. Robert McLaren, Dr. Tom McElroy and Dr. Jinliang Liu for their encouragement, insightful comments and constant support. In addition, I would like to offer my special thanks to Dr. Jack McConnell, my original supervisor, who was instrumental in giving my thoughts the right direction, and then sadly passed away before he could see the product of his guidance.

My sincere thanks also goes to Dr. Arthur Mizzi and Dr. Chris Snyder, for their constructive comments and warm encouragement when I was visiting National Center for Atmospheric Research (NCAR).

I thank my fellow lab mates: Geoffery Bell, Sopan Kurkute, Zhan Li, Jianyu Liang, and Zhongqi Yu for the stimulating discussions, and for all the fun we have had in the last several years.

I would also like to offer my special thanks to the NSERC CREATE Training Program for Integrating Atmospheric Chemistry and Physics from Earth to Space (IACPES) for their consistent financial support.

Last but not the least, I would like to thank my family for supporting me spiritually throughout my life.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Carbon, the building block of life, is stored primarily in rocks and sediments on Earth, with only a tiny fraction residing in the atmosphere, oceans, soils, and biosphere. The small fraction of carbon present in the atmosphere is of paramount importance in regulating the climate of the planet by controlling its abundance. Although, constantly being transferred between various reservoirs, carbon was in a state of dynamic equilibrium over many years until human activities altered the global carbon cycle significantly since the last 150 years. Understanding the consequences of these activities is critical for societal decisions about the management of the carbon cycle. Among all the compounds of carbon that exist in the atmosphere, carbon monoxide (CO) plays a central role in atmospheric chemistry. As a precursor of ozone (O_3) and an important sink of the hydroxyl radical (OH), it strongly influences the oxidizing capacity of the atmosphere. CO is also a good tracer of atmospheric pollution as its mid-range lifetime allows plumes

to be transported over long distances. It is emitted by the incomplete combustion of fossil fuels and biomass, and is also produced by the oxidation of methane (CH_4) and biogenic non-methane hydrocarbons (NMHCs). Hence, monitoring and predicting the atmospheric CO concentration are of great importance.

Computer modeling and simulation play pivotal roles in contemporary air chemistry forecast. The use of chemical transport models (CTMs) to produce air quality forecasts has become a new application area, providing important information to the public, decision makers and researchers. Although chemical transport models have improved substantially during the last several decades, their forecast ability still falls behind numerical weather models. Air quality predictions have large uncertainties associated with: incomplete and/or inaccurate emission information; lack of key measurements to impose initial and boundary conditions; missing science elements; and poorly parameterized processes. Improvements in the analysis capabilities of CTMs require them to be better constrained through the use of observational data. Fortunately, in the last decades several spaceborne instruments provided information on the distribution of chemical species in the troposphere, leading to an improved understanding of chemical and transport processes as well as emissions. In particular, tropospheric CO columns and profiles have been obtained from polar orbiting satellites, and Measurements Of Pollution In The Troposphere (MOPITT) is one of them. One way to incorporate these observations within the model simulations is through data assimilation (DA). DA was originally introduced to meteorology to provide more accurate initial conditions for numerical weather prediction (NWP).

Borrowing lessons learned from the evolution of NWP models, improving air quality predictions through the assimilation of chemical observation data holds significant promise.

Various assimilation techniques exist to assimilate chemical tracers in chemical transport models (CTMs). In this work a sequential ensemble Kalman filter (EnKF) data assimilation system with the Weather Research and Forecasting model coupled with Chemistry (WRF-Chem) is used to assimilate MOPITT CO retrievals. Although this system is designed to provide a global CO forecast, it will be tested in this first study at a regional scale. Assimilation techniques offer a powerful tool to propagate in space and time the information provided by the satellites and to constrain surface chemical tracer emissions models. These are also the two foci in this study. Firstly, we will investigate the effects of data assimilation on improving the performance of short-range air quality forecasts. Secondly, we focus on the capability of the system to estimate CO emissions through inverse modeling. The optimized emissions will also be applied into WRF-Chem to further improve the model performance.

The thesis is organized as follows: Chapter two briefly introduces the Earth's carbon cycle and chemical transport models that are used currently. Chapter three describes the data assimilation techniques with an emphasis on ensemble Kalman filtering. In Chapter four, the development of the analysis and forecast system is given and its performance tested in a case study is provided in Chapter five. The inverse modeling ability of this system is studied in Chapter six. Finally, summaries and conclusions are given in Chapter seven.

Chapter 2

Carbon cycle and chemical transport models

2.1 The carbon cycle

Carbon cycling between the atmosphere, the oceans, and the terrestrial biosphere makes a part of the global matter cycle. The three compartments all consist of reservoirs that store and exchange carbon at different rates and quantities, and with different lifetimes. According to Gruber et al. [2004], the largest amount of carbon, about 38,000 Pg C, is stored in the middle and deep ocean, which is relatively inert. Smaller, but still considerably large, amounts are found in the terrestrial biosphere (2,100 – 3,000 Pg C), the surface layer of the oceans (600 Pg C) and the atmosphere (700 – 800 Pg C), whose turnover

rates are relatively high. The principal matter carrier in the carbon cycle is carbon dioxide (CO_2). Photosynthesis and respiration are the primary processes facilitating carbon exchange between the biosphere and the atmosphere, with an estimated exchange rate at 120 Pg C $year^{-1}$ [Emerson and Hedges, 2008]. However, these influxes and effluxes are closely in balance, making the flux of CO_2 into the atmosphere from anthropogenic sources significant in perturbing the atmospheric CO_2 budget.

Beside CO_2, carbon monoxide (CO) is the next most important species in the carbon cycle. Although it is less abundant than CO_2, it plays a central role in atmospheric chemistry. Since it is the major sink of the hydroxyl radical (OH) (Equation (2.1)) and a major precursor of ozone (O_3) (Equation (2.7)) in the troposphere, two of the most important oxidizing species in the air, CO strongly influences the oxidizing capacity of the atmosphere and determines the lifetimes and abundances of most other atmospheric trace gases including methane (CH_4), non-methane hydrocarbons (NMHCs), hydrochlorofluocarbons (HCFCs), hydrogen sulfide (H_2S) and sulfur dioxide (SO_2) [Gupta et al., 1998; Logan et al., 1981; Thompson, 1992].

$$CO + OH \rightarrow CO_2 + H \tag{2.1}$$

In an environment rich in NO, ozone is formed by the oxidization of CO:

$$CO + OH \rightarrow CO_2 + H \tag{2.2}$$

$$H + O_2 + M \rightarrow HO_2 + M \tag{2.3}$$

$$HO_2 + NO \rightarrow OH + NO_2 \tag{2.4}$$

$$NO_2 + h\nu \rightarrow NO + O \tag{2.5}$$

$$+ \qquad O + O_2 \rightarrow O_3 \tag{2.6}$$

$$Net: \qquad CO + 2O_2 \rightarrow CO_2 + O_3 \tag{2.7}$$

where M is a third "body" with mass, primarily nitrogen or oxygen molecules in the atmosphere.

CO is a global pollutant with a variety of sources. It is produced by both natural emissions and human activities and it is formed primarily through natural atmospheric oxidation processes and incomplete combustion from burning fossil fuels and biomass. CO has a mean lifetime of about two months in the atmosphere and thus can serve as a regional and global tracer of the transportation of the pollution. However, due to the sparse distribution of CO observation networks, conventional estimations of tropospheric sources and sinks, and the global budget of CO, still have large uncertainties [Brasseur et al., 1999], and a closer study using a numerical model is desired.

2.2 Chemical transport models

2.2.1 Introduction

A chemical transport model (CTM) is a computer-based numerical model, which simulates the emission, transportation and evolution of atmospheric chemical species. Aiming to realistically present the processes that occur in the real atmosphere, CTMs are recognized as useful tools for accessing and evaluating air quality. A CTM solves the continuity equation for the concentration of certain constituents of interest either coupled offline or online with a meteorological model. In an offline CTM, the chemical processes are computed independently from the meteorological model, whose output feeds the transport of the chemicals. Since the output time frame of the meteorological model is much longer than the time scale of physical and chemical processes that affect the distribution of the chemical species, information is lost due to the decoupling of the two [Grell et al., 2005]. The Weather Research and Forecasting model coupled with Chemistry (WRF-Chem), an online CTM, is used in this study to reduce the potential information lost. Based on the WRF model, WRF-Chem has the capability to simulate the coupling between dynamics, radiation, and chemistry. Brief descriptions of the WRF and the WRF-Chem are provided in the next section.

2.2.2 Weather Research and Forecasting model coupled with Chemistry

The WRF model [Skamarock et al., 2008] is a numerical weather prediction system which has been used for both operational forecasting and atmospheric research ranging from local scale to global scale. The effort to develop WRF has been a collaborative partnership, principally among the National Center for Atmospheric Research (NCAR), the National Oceanic and Atmospheric Administration (NOAA), the National Center for Environmental Prediction (NCEP), the Forecast Systems Laboratory (FSL), the Air Force Weather Agency (AFWA), the Naval Research Laboratory (NRL), Oklahoma University, and the Federal Aviation Administration (FAA). The default dynamics solver, Advanced Research WRF (ARW) developed primarily at the NCAR, uses a time-split second or third order Runge-Kutta scheme for integration of the governing equations in the atmosphere. It is an Eulerian, non-hydrostatic and fully compressible model, which uses a terrain following vertical pressure coordinate system and staggered Arakawa C-grid horizontal grid structure. It includes full physics options for land-surface, planetary boundary layer, atmospheric and surface radiation, microphysics and cumulus convection, and is highly adaptable to very specific problems.

Coupling the chemistry module fully with the WRF gives the WRF-Chem model. The on-line WRF-Chem model version 3.4.1 [Fast et al., 2006; Grell et al., 2005] is used in this dissertation. The on-line approach has the advantage of using the same time steps (or a multiple of it), grid cells, physics schemes and advection schemes as WRF and is fully

consistent with the meteorological components. WRF-Chem includes an emissions driver, a photolysis driver, a dry deposition driver, parameterization for convective transport, a chemical mechanism driver, and an aerosol chemistry driver. The chemical mechanism driver includes aqueous phase chemistry coupled with aerosols and the microphysics parameterization of the meteorological model, as well as a gas-phase chemical reaction mechanism. The emissions driver includes anthropogenic emissions as well as calculations for approximate biogenic emissions. As the same with WRF, WRF-Chem also offers a choice of different physical and chemical options. The two hard-coded gas phase chemical mechanisms in WRF-Chem are the second generation Regional Acid Deposition Model mechanism (RADM2) [Stockwell et al., 1990], and the Carbon Bond Mechanism version Z (CBM-Z) [Zaveri and Peters, 1999]. The kinetic preprocessor (KPP) [Grell et al., 2011; Salzmann, 2008] is also available in WRF-Chem, which allows many additional gas phase chemical mechanisms to be used in WRF-Chem.

Chapter 3

Data assimilation

3.1 Introduction

Data assimilation (DA) can be defined as the incorporation of observations into a dynamical model to optimally estimate the state of a physical system. A good assimilation makes the model state more consistent with the observations. DA was originally introduced to meteorology to provide more accurate initial conditions for numerical weather prediction [Lynch, 2006], and its positive impact on the accuracy of weather forecasts is unquestionable [Simmons and Hollingsworth, 2002]. Beside that, DA can also be used to access the best available estimate of the atmospheric state [Compo et al., 2006; Uppala et al., 2005]; estimate the value of existing or hypothetical observations [Khare and Anderson, 2006; Zhang et al., 2004]; evaluate forecast models; guide model development

by estimating values for model parameters that are most consistent with observations [Aksoy et al., 2006; Houtekamer et al., 1996].

Inspired by the success of data assimilation in meteorology, it is now used also for chemical constituents analysis and forecast. Studies [Singh and Sandu, 2009; Zhang and Sandu, 2007] have illustrated the benefits of chemical data assimilation in improving initial and boundary conditions that contribute to better air quality forecasts. Unlike short-range weather forecasts whose results largely depend on initial conditions, air chemistry forecasts are also largely driven by emission and removal processes in addition to initial conditions and lateral boundary conditions in regional simulations. Therefore, to improve the analysis and forecast capabilities of CTMs, it is necessary to consider optimal emission estimation through data assimilation [Menut, 2003; Stewart, 1993]. The use of refined emission estimates from data assimilation has been demonstrated to improve the model forecast [Alexe and Sandu, 2011; Chai et al., 2009].

Variational and Ensemble Kalman filter (EnKF) methods are two approaches to data assimilation that are widely used in applications. Detailed descriptions of the two methods can be found in Kalnay [2003]. Relatively speaking, variational methods have been studied more comprehensively. The ability of three-dimensional variational (3D-Var) method and four-dimensional variational (4D-Var) method to improve chemical initial conditions [Bei et al., 2008; Dethof and Holm, 2004; Flemming et al., 2011; Jackson, 2007; Tang et al., 2004] and the potential of 4D-Var in emission inversion [Chai et al., 2009; Maki et al., 2011; Yumimoto and Uno, 2006] all have been validated. However, Kalman filters have been employed successfully in chemical data assimilation for only around a

decade [Clark et al., 2007; Khattatov et al., 2000; Lamarque et al., 2002; Pierce et al., 2007], and not until more recently that the ensemble Kalman filter (EnKF) [Evensen, 1994] has been examined in the context of chemical data assimilation [Constantinescu et al., 2007a,b]. Studies that compare the relative merits and performance of different approaches show KF is comparable to or even outperforms 4D-Var [Singh et al., 2011; Wu et al., 2008]. Hence, the capability of EnKF in chemical DA was explored in detail in this dissertation.

This chapter consists of a thorough review of the formulation of the Kalman Filter (KF). The principles of the extended Kalman filter (EKF) and ensemble Kalman filter (EnKF) are introduced. The ensemble adjustment Kalman filter (EAKF) is used in the experiments of this project and consequently will be presented and described. EnKF also has its limitations such as undersampling and filter divergence. Two most widely used ways to avoid these problems: localization and inflation will be briefly described at the end of this chapter.

3.2 Kalman filtering

3.2.1 The Kalman filter

The Kalman filter (KF) [Kalman, 1960; Kalman and Bucy, 1961] is a well established and widely used method of sequential data assimilation. It is a linear estimator that produces the minimum variance estimate in a least-squares sense under the assumption

of Gaussian noise. The problem is solved through two stages: the forecast stage and the analysis stage. First of all, a forecast state, or a background state, is computed by evolving a given linear model in time in the forecast stage. Then observations are assimilated to adjust the forecast estimate to more accurately reflect the true stat of the system.

Denote $\mathbf{x}_t^f$ and $\mathbf{x}_t^a$ as the forecast and analysis state estimate at time step t, respectively. Both of them have an associated error covariance matrix, $\mathbf{P}$, which are defined by

$$\mathbf{P}_t^f =< (\mathbf{x}_t^f - \mathbf{x}_t^{true})(\mathbf{x}_t^f - \mathbf{x}_t^{true})^T >, \tag{3.1}$$

$$\mathbf{P}_t^a =< (\mathbf{x}_t^a - \mathbf{x}_t^{true})(\mathbf{x}_t^a - \mathbf{x}_t^{true})^T > . \tag{3.2}$$

Here, $\mathbf{x}_t^{true}$ is the true state, which is normally unknown, and for a given quality s, $\langle$ s $\rangle$ denotes the expected value of s.

In the first stage; a forecast state estimate of the system at current time step t is produced by advancing the model in time using the best available state estimation at previous time step t-1. An important feature of the Kalman filter is that, as well as updating the state estimate, the error covariance matrix is also updated at each step. This is commonly referred to as the flow-dependent background errors. The update equations

for the state forecast and the error covariance forecast at time step t are:

$$\mathbf{x}_t^f = \mathbf{M}_t \mathbf{x}_{t-1}^a + \eta_t, \tag{3.3}$$

$$\mathbf{P}_t^f = \mathbf{M}_t \mathbf{P}_{t-1}^a \mathbf{M}_t^T + \mathbf{Q}_t. \tag{3.4}$$

$\mathbf{M}_t$ is a matrix defining the linear system dynamics, η_t and $\mathbf{Q}_t$ are the random model error covariance at time t. In general, this model error is unknown, it is either set to zero or estimated separately [Evensen, 2003]. In KF, this error is assumed to be unbiased (i.e. $\langle\ \eta_t\ \rangle = 0$) and following a normal distribution.

In order to compare model state with observations, a linear observation operator, $\mathbf{H}$, is applied to map the model space to observation space. Observations at a given time t, $\mathbf{y}_t$ satisfy

$$\mathbf{y}_t = \mathbf{H}\mathbf{x}_t^{true} + \epsilon_t. \tag{3.5}$$

The random observational error is given by ϵ_t and is assumed to be unbiased, $\langle\ \epsilon_t\ \rangle = 0$. The observational error covariance is computed using

$$\mathbf{R}_t = < \epsilon_t \epsilon_t^T > . \tag{3.6}$$

The second stage is the analysis stage. A weighting, also known as the Kalman gain matrix $\mathbf{K}_t$, is calculated using

$$\mathbf{K}_t = \mathbf{P}_t^f \mathbf{H}_t^T (\mathbf{H}_t \mathbf{P}_t^f \mathbf{H}_t^T + \mathbf{R}_t)^{-1}. \tag{3.7}$$

The Kalman gain matrix is a weighting to give to the observations according to the ratio between the forecast and observational error covariance. The best linear estimate of the state can be obtained using the following equation

$$\mathbf{x}_t^a = \mathbf{x}_t^f + \mathbf{K}_t(\mathbf{y}_t - \mathbf{H}_t\mathbf{x}_t^f). \tag{3.8}$$

The filter is optimal in the sense that the analysis defined by Equation (3.8) minimizes the cost function

$$\mathcal{J}(\mathbf{x}) = (\mathbf{x} - \mathbf{x}^f)^T(\mathbf{P}^f)^{-1}(\mathbf{x} - \mathbf{x}^f) + (\mathbf{y} - \mathbf{H}\mathbf{x})^T\mathbf{R}^{-1}(\mathbf{y} - \mathbf{H}\mathbf{x}). \tag{3.9}$$

This function is a weighted measure of the distance from the state $\mathbf{x}$ to the forecast $\mathbf{x}^f$ and the observation $\mathbf{y}$. The analysis represents a combination of both information sources of forecast and observation, with a larger weight given to the more certain components. Finally the error covariance is updated using the same Kalman gain.

$$\mathbf{P}_t^a = (\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)\mathbf{P}_t^f. \tag{3.10}$$

The major drawback of KF is that it is only valid for linear systems where CTMs are based on non-linear dynamical models. Moreover, it is computationally expensive. Operational global forecast models have 10^8 state variables and assimilate 10^5 - 10^6 observations per assimilation period and the number is continuously growing. This requires enormous storage to store state error covariance matrices of size $10^8 \times 10^8$. And the

calculation of the Kalman gain in Equation (3.7) involves the inversion of a matrix of size $10^5 \times 10^5$ or larger. It is not feasible given the current computing power.

3.2.2 The extended Kalman filter

To extend the Kalman filter to nonlinear models and observations, the extended Kalman filter (EKF) method was developed. The linear models of dynamics $\mathbf{M}$ in Equation (3.3) is replaced by the nonlinear models of dynamics $\mathcal{M}$. And a nonlinear forward operator $\mathcal{H}$ is used instead of the linear forward operator $\mathbf{H}$ in Equation (3.8). The error covariance update equations and the calculation of the Kalman gain remain the same with the proviso that $\mathbf{M}$ and $\mathbf{H}$ are now the tangent linear operators (Jacobins) of $\mathcal{M}$ and $\mathcal{H}$ respectively. The forecast and analysis steps of EKF can be written as:

$$\mathbf{x}_t^f = \mathcal{M}_t \mathbf{x}_{t-1}^a + \eta_t, \tag{3.11}$$

$$\mathbf{P}_t^f = \mathbf{M}_t \mathbf{P}_{t-1}^a \mathbf{M}_t^T + \mathbf{Q}_t, \tag{3.12}$$

$$\mathbf{K}_t = \mathbf{P}_t^f \mathbf{H}_t^T (\mathbf{H}_t \mathbf{P}_t^f \mathbf{H}_t^T + \mathbf{R}_t)^{-1}, \tag{3.13}$$

$$\mathbf{x}_t^a = \mathbf{x}_t^f + \mathbf{K}_t(\mathbf{y}_t - \mathcal{H}_t \mathbf{x}_t^f), \tag{3.14}$$

$$\mathbf{P}_t^a = (\mathbf{I} - \mathbf{K}_t \mathbf{H}_t)\mathbf{P}_t^f. \tag{3.15}$$

Since EKF relies on linear approximations to nonlinear functions, it works well if the system is only weakly nonlinear. Nevertheless, the update equations no longer preserve the unbiasedness, the error covariance update equations are no longer exact, and the filter

is no longer optimal in the sense of minimizing the cost function (3.9). The EKF also does not address the problem of handling huge covariance matrices.

3.2.3 The ensemble Kalman filter

The ensemble Kalman filter (EnKF) is introduced to reduce the amount of computation while still use nonlinear models in the formulation. It was originally developed by Evensen [1994], and had many subsequent developments [Burgers et al., 1998; Evensen, 2003; Houtekamer and Mitchell, 1998]. The key idea of EnKF is, instead of using a single state estimate to maintain a separate covariance matrix, an ensemble of state estimates are used to approximate the error covariance statistically. Since EKF represents nonlinearity using derivatives that only take into account behavior in an infinitesimal neighborhood of a point, the use of an ensemble will provide a better representation of the effects of nonlinearity.

Consider an initial ensemble of size N, the ensemble mean of all the members $\mathbf{x}_i$, i = 1, 2, ..., N, is

$$\bar{\mathbf{x}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i. \tag{3.16}$$

And the ensemble error covariance $\mathbf{P}$ can be written as

$$\mathbf{P} = \frac{1}{N-1}\sum_{i=1}^{N}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T. \tag{3.17}$$

The division over the ensemble by N-1 not N ensures that $\mathbf{P}$ is an unbiased estimate of the covariance [Barlow, 1989]. Define a state ensemble matrix as

$$\mathbf{X} = \frac{1}{\sqrt{N-1}}(\mathbf{x_1}\ \ \mathbf{x_2}\ \ ...\ \ \mathbf{x_N}), \tag{3.18}$$

where each column is the state estimate for an individual ensemble member. An ensemble perturbation matrix can be written as

$$\mathbf{X}' = \frac{1}{\sqrt{N-1}}(\mathbf{x_1} - \bar{\mathbf{x}}\ \ \mathbf{x_2} - \bar{\mathbf{x}}\ \ ...\ \ \mathbf{x_N} - \bar{\mathbf{x}}). \tag{3.19}$$

This allows us to rewrite the ensemble covariance matrix in Equation (3.17) as

$$\mathbf{P} = \mathbf{X}'\mathbf{X}'^T. \tag{3.20}$$

To include the measurement error covariance matrix $\mathbf{R}$ into the expression, the observations are treated as random variables which are perturbed by unbiased Gaussian errors with standard deviation of σ and covariance of $\mathbf{R}$:

$$\mathbf{y}_i(t) = \mathbf{y}(t) + \sigma_i, \tag{3.21}$$

and

$$\mathbf{R} = < \sigma\sigma^T > . \tag{3.22}$$

For each ensemble member, the forecast step evolves the nonlinear model dynamics forward in time:

$$\mathbf{x}_i^f(t) = \mathcal{M}(t)\mathbf{x}_i^a(t-1) + \eta_i(t), \tag{3.23}$$

and the corresponding error covariance is computed by:

$$\mathbf{P}_t^f = \mathbf{X}_t'^f(\mathbf{X}_t'^f)^T. \tag{3.24}$$

Then Kalman gain is calculated using the $\mathbf{P}_t^f$ from Equation (3.24):

$$\mathbf{K}_t = \mathbf{P}_t^f\mathbf{H}_t^T(\mathbf{H}_t\mathbf{P}_t^f\mathbf{H}_t^T + \mathbf{R}_t)^{-1}. \tag{3.25}$$

The terms involving $\mathbf{P}_t^f$ are approximated by sample error covariance matrices:

$$\mathbf{P}_t^f\mathbf{H}_t^T \approx \frac{1}{N-1}\sum_{i=1}^{N}(\mathbf{x}_{ti} - \overline{\mathbf{x_t}})[\mathcal{H}(\mathbf{x}_{ti}) - \overline{\mathcal{H}\mathbf{x_t}}]^T \tag{3.26}$$

$$\mathbf{H}_t\mathbf{P}_t^f\mathbf{H}_t^T \approx \frac{1}{N-1}\sum_{i=1}^{N}[\mathcal{H}(\mathbf{x}_{ti}) - \overline{\mathcal{H}\mathbf{x_t}}][\mathcal{H}(\mathbf{x}_{ti}) - \overline{\mathcal{H}\mathbf{x_t}}]^T \tag{3.27}$$

The analysis step for the EnKF consists of the following updates performed on each of the model state ensemble members

$$\mathbf{x}_i^a(t) = \mathbf{x}_i^f(t) + \mathbf{K}_t[\mathbf{y}_i(t) - \mathcal{H}_t\mathbf{x}_i^f(t)], \tag{3.28}$$

Finally, the error covariance of the analysis state is approximated by:

$$\mathbf{P}_t^a = \mathbf{X}_t'^a (\mathbf{X}_t'^a)^T. \tag{3.29}$$

It is easy to see from the formulation that EnKF differs from the KF and EKF in using an ensemble of state estimates instead of a single state estimate and not maintaining a separate error covariance matrix, which offers the advantages of reduced computational cost, better handling of nonlinearity, and greater ease of implementation.

3.2.4 The ensemble adjustment Kalman filter

Alternative ways to the perturbed observations approach as in EnKF are the ensemble square-root filters (EnSRF) that generate an analysis ensemble mean and covariance satisfying the Kalman filter equations. Different square-root filters are possible since the same analysis error covariance can be archived using different analysis ensemble perturbations. The EnSRF include ensemble adjustment Kalman filter (EAKF) [Anderson, 2001], serial EnSRF [Whitaker and Hamill, 2002], ensemble transform Kalman filter (ETKF) [Bishop et al., 2001], local ensemble Kalman filter (LEKF) [Ott et al., 2004], and local ensemble transform Kalman filter [Hunt et al., 2007]. Among them, EAKF has been implemented into the Data Assimilation Research Testbed (DART) infrastructure and has been applied to many geophysical problems and this study also used this algorithm to assimilate observations into a CTM.

EAKF is similar to the EnKF except it uses a different algorithm for updating the ensemble. In EAKF, a new ensemble that has the exact mean and covariance while maintaining as much as possible the higher moment structure of the prior distribution is generated directly. This is done by applying a linear operator, $\mathbf{A_d}$, to the prior ensemble to get the updated ensemble:

$$\mathbf{x}_i^a = \mathbf{A_d}^T(\mathbf{x}_i^f - \overline{\mathbf{x}^f}) + \overline{\mathbf{x}^a}, i = 1, 2, ..., N. \tag{3.30}$$

The prove of existence and computation of $\mathbf{A_d}$ can be found in Anderson [2001].

The EAKF performs well under the situation that the ensemble size is much smaller than the state space dimension and has the ability to assimilate observations with complex nonlinear relations to the state variables. It eliminates the noise introduced by EnKF, which uses a random sample of the observational error distribution.

3.3 Inflation and localization in data assimilation

Previous studies shows that EnKF has a decreasing ability to correct the ensemble state toward the observations at the end of the assimilation cycle. The progressive underestimation of the model error covariance magnitude during the integration finally lead to the filter divergence [Hamill, 2004; Houtekamer and Mitchell, 1998]. The filter becomes too confident in the model and ignores the observations in the analysis process. One solution is to increase the covariance of the ensemble artificially and therefore decrease the filter's

confidence in the model results. A spatially and temporally varying adaptive covariance inflation algorithm [Anderson, 2009] is used in this study. It associates a different inflation value with each model state vector component and avoids the problem that the ensemble variance be inappropriately small or large.

Good performance of EnKF with small ensemble size may require localizing the impact of an observation to state variables that are geographically close to the observation. Localization is viewed as a means to ameliorate sampling error when a small ensemble size is used to sample the statistical relation between an observation and a state variable.

3.4 Inverse modeling

Although the EnKF has generally been used for initial state estimation, parameter estimation can readily be included in the same framework by the means of state space augmentation [Anderson, 2001; Derber, 1989]. The principle is that the parameters could also be considered to be part of the model state alongside the conventional variables, and then the covariance sampled by the ensemble members can be used directly to update parameters in exactly the same manner as for the state variables. Consider the emission rates of the atmospheric chemical constitutions as model parameters; updating them during each data assimilation cycle will give an optimal estimation of the sources. This is the so called inverse modeling approach, which use measurements of atmospheric chemical mixing ratios to determine source distributions that lead to optimal agreements between model simulations and these observations. It is completely different from the traditional

up-scaling approach. The up-scaling approach interpolates and extrapolates all available existing information about source and sink processes, including results of local field experiments and statistics on regional or national levels, to obtain a global picture of the source distributions.

Chapter 4

Analysis and forecast system development

4.1 Introduction

The analysis and forecast system mainly consists of two components: the EnKF module that updates the state variables using observations and the WRF-Chem model that forwards state variables in time. The EnKF module used in this study is the Data Assimilation Research Testbed (DART) developed at NCAR. DART has already included the entire data assimilation algorithm and significantly simplifies our work in building the analysis and forecast system. The only work we have to do is to build an interface between DART and WRF-Chem, a foreward operator for MOPITT column integrated mixing ratio of CO and prepare the observations that can be read by the DART program.

A brief overview of DART will be provided in the next section and followed by a thorough description of the system development.

4.2 Data assimilation research testbed

The Data Assimilation Research Testbed, released in 2009 from the Data Assimilation Research Section (DAReS) at NCAR, is an ensemble data assimilation facility. It is a software package that consists of various ensemble methods for data assimilation. In this thesis, EAKF is applied to relax the underlying models toward a state that is more consistent with the observations using a modular programming approach, which is customizable and easy to implement and use.

Figure 4.1 [Anderson et al., 2009] is a schematic flow diagram of ensemble data assimilation coupling DART with a numerical model. Start at the top and work clockwise. The assimilation cycle starts by reading in a namelist, initial states for the ensemble, and the observations. First of all, DART determines the types of observations that are used and then the corresponding forward operators appropriate for the particular types of observations are applied to each ensemble member to generate the model's estimation of the observations. Then all the data are passed to the main executable of DART, named "filter". This program updates the state variables according to the differences between the real observations and model equivalent observations, namely the observational innovation. The updated state vectors are generated by adding a weighted difference between the observation and model estimate to the initial state vectors. After that, the updated

DART state vectors are converted to the format required by the underlying model, so that the model can advance to the next analysis time when new observations are ready to be assimilated. Converting the new model forecast output back into DART form closes the cycle. We are now back at the beginning and the cycle continues as long as there are observations to be assimilated or it is terminated by the control information in the Fortran namelist. By the end of the entire data assimilation process, a diagnostic file containing the statistical information of the ensemble is written for the exploration of the data assimilation performance.

One advantage of DART is that the algorithms are designed so that customizing to new observation types and new models requires minimal coding, and it does not require modifications of the existing model code. In this study, a new observation type, MOPITT CO total column, is defined in order for DART to read MOPITT CO total column retrievals. And two modules are mandatory to incorporate DART with the WRF-Chem model. One is the model module that informs DART the layout of the state variables and the other is a forward operator module that maps model space to observation space. They will be discussed separately in the following section.

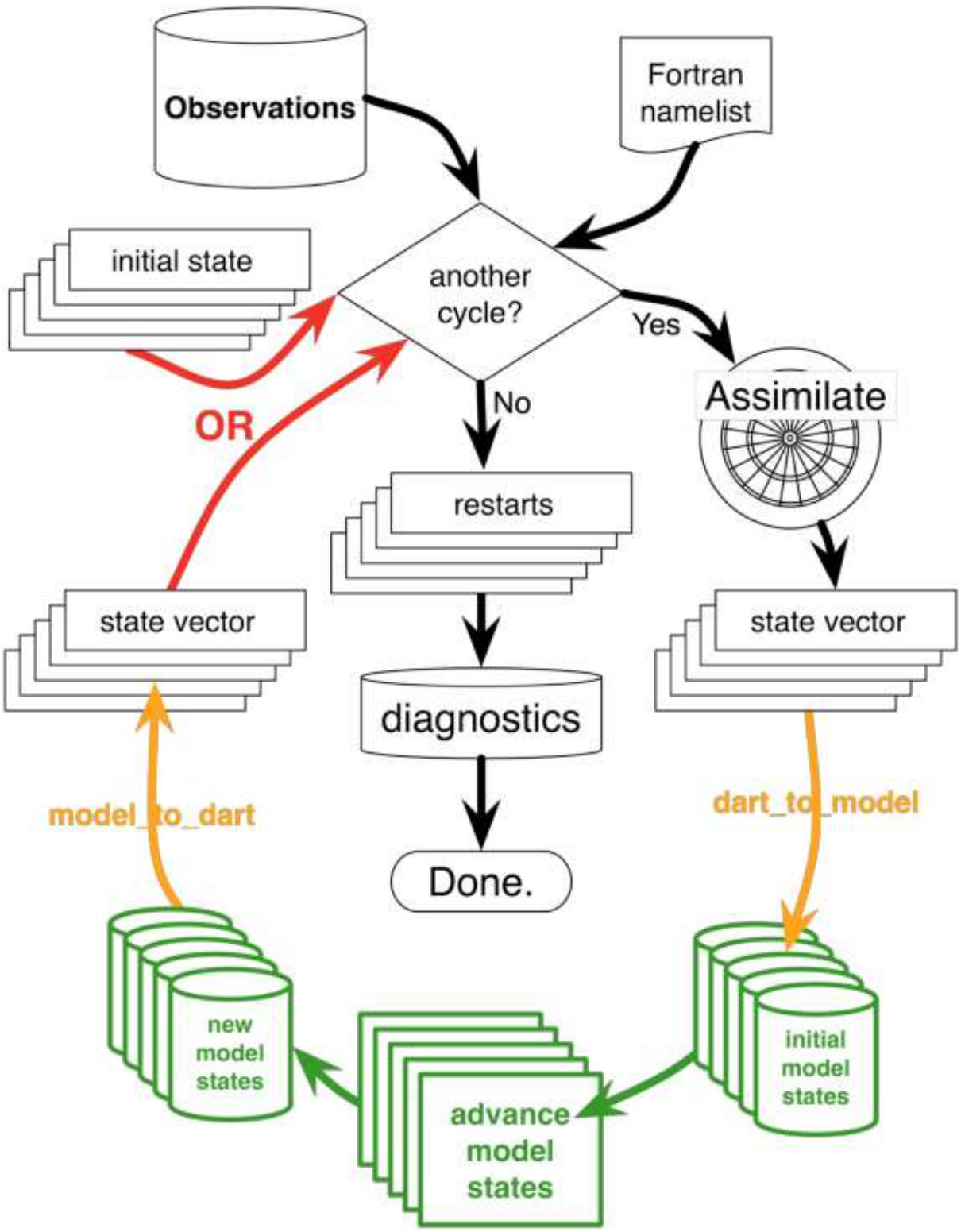

FIGURE 4.1: Schematic flow diagram of ensemble data assimilation [Anderson et al., 2009]

4.3 System development

4.3.1 MOPITT observation

The Measurements of Pollution In The Troposphere (MOPITT) experiment was launched on December 18, 1999, onboard the National Aeronautics and Space Administration (NASA) Terra satellite. It is a nadir viewing infrared radiometer, which targets measurements of carbon monoxide and methane. The primary objective of MOPITT is to enhance knowledge of the lower atmosphere system through monitoring the distribution, transport, sources and sinks of CO and CH_4 in the troposphere. A more complete description of the MOPITT instrument can be found in the MOPITT mission description document [Drummond, 1996].

The MOPITT instrument makes measurements of radiation emerging from the atmosphere in two spectral bands for the retrieval of CO total columns and vertical profiles. The first band targets the infrared radiation from either the thermal emission or absorption of CO at the wavelength of 4.7 μm, and the second focuses on reflected solar radiation at 2.2 - 2.4 μm. The operational MOPITT retrieval uses a non-linear optimal estimation algorithm [Pan et al., 1998; Rodgers, 2000] to invert the measured signals to determine tropospheric mixing ratios of CO. The CO retrieval algorithm used for MOPITT exploits the maximum a posteriori (MAP) solution which is a specific type of optimal estimation technique [Rodgers, 2000]. MOPITT retrieves CO volume mixing ratio (VMR) profile on a ten-level pressure grid (surface, 900 hPa, 800 hPa, 700 hPa, 600

hPa, 500 hPa, 400 hPa, 300 hPa, 200 hPa, 100 hPa) with a horizontal resolution of 22 km $\times$ 22 km. The CO profiles are retrieved using an optimal estimate of the maximum likelihood solution. With this technique, the retrieved CO profiles depend not only on MOPITT radiance measurements ($\mathbf{Z_o}$), but also on a priori CO profile. The a priori is used as the first guess to the retrieved field. For MOPITT CO retrievals, Model for OZone and Related chemical Tracers version 4 (MOZART-4) CO monthly mean climatology over 1997 - 2004 was used as the a priori ($\mathbf{Z_a}$). A Newtonian iterative form of the maximum a posteriori solution is found which combines the actual measurements and the a priori state vector, inversely weighted by their respective covariances. The retrieved CO profile ($\mathbf{Z_r}$) is calculated using

$$\begin{aligned} \mathbf{Z_r} &= \mathbf{Z_a} + \mathbf{A}(\mathbf{Z_o} - \mathbf{Z_a}) \\ &= \mathbf{A}\mathbf{Z_o} + (\mathbf{I} - \mathbf{A})\mathbf{Z_a} \end{aligned} \tag{4.1}$$

where $\mathbf{A}$ is the averaging kernel matrix and $\mathbf{I}$ is the identity matrix . The averaging kernels indicate the sensitivity of the retrieval to the atmospheric state and provide the relative weighting between the true and a priori profile. The ideal situation would be one where $\mathbf{A}$ equals the identity matrix $\mathbf{I}$ then from Equation 4.1 we see that

$$\mathbf{Z_r} = \mathbf{Z_o} \tag{4.2}$$

and the retrieved profile reflects the true profile. Generally though this is not the case, the averaging kernel gives some indication of the vertical resolution and sensitivity of the

retrieval and the influence of the a priori. Uncertainties for MOPITT retrieval includes both instrumental noise and geophysical noise, i.e., random errors in the calibrated radiances resulting from the combined effects of field of view motion and fine-scale spatial variability in surface radiative properties [Deeter et al., 2011].

In this study, MOPITT CO Version 5 (V5) Level 2 (L2) data retrieved using Equation 4.1 were used and some preprocessing of raw MOPITT data was performed before data assimilation. DART processes the observations through an observation sequence text file. In this file, all observations are ordered according to the observation time. And for a single MOPITT observation, the observed CO total column is given first followed by the observation location. The number of vertical layers, normally ten, but possibly smaller if the surface pressure is less than 900 hPa, is also given as an indicator of the dimensions of the a priori and the averaging kernel. Instead of using the actual CO VMR, the $log_{10}(VMR)$ is used throughout the entire retrieval process. Hence the a priori given in the raw MOPITT data are in log-normal scale and the averaging kernels describe the sensitivity of retrieved $log_{10}(VMR)$ to actual $log_{10}(VMR)$. We will use the same log-normal scale in the observation sequence file. Notice that, from Equation 4.1, the final retrieval result is a function of $(\mathbf{I} - \mathbf{A})\mathbf{Z_a}$. Therefore, instead of writing the a priori, the weighted a priori $(\mathbf{I} - \mathbf{A})\mathbf{Z_a}$ is written to the observation sequence file, followed by the log-normal averaging kernel. The observation time is also given for reference. The file ends with the observed CO total column error variance. This kind of observation file is generated at each assimilation cycle with a time window of ± 3 hours, i.e. observations taken three hours ahead or after the assimilation time will be gathered in the same

observation sequence file and assimilated at once.

4.3.2 WRF-Chem and DART interface

The modifications to DART are necessary to link with the WRF-Chem model and build an ensemble data assimilation and forecasting system. Transformation programs are necessary to transform the WRF-Chem restart files into the required state vector and then back to WRF-Chem files before/after each analysis and forecast step respectively (i.e. the yellow dart_to_model/model_to_dart programs in Figure 4.1). DART also needs information about the WRF-Chem variable names, grid specifications and time stepping in order to assign suited assimilation windows and to ensure the right interpolation of observations in the vertical and horizontal directions.

In this study, the default EAKF is used to assimilate the CO total column as well as the conventional observations including surface, radiosonde, aircraft, and satellite measurements. Consequently, DART needs at least the following variables in a state vector. They are: temperature, pressure, wind, geopotential height, water vapour mixing ratio and carbon monoxide. In addition, surface temperature, pressure, wind and some chemical species such as O_3 and NO_x are also included for analysis. In DART, a desired generic 'kind' (like KIND_PRESSURE, KIND_TEMPERATURE, KIND_SPECIFIC_HUMIDITY, KIND_CO) is assigned to each entry of the state vector as an identifier. Two separate programs transfer the relevant variables (T, P, U, V, W, H, Q, CO plus optional variables) from the model space to the state vector (model_to_dart) and, after the state vector is

updated by DART, back (dart_to_model) to the model grid. The WRF-Chem/DART interface reads the state vector, determines the model time, model grid size and assimilation window. For this study, the assimilation window is six hours, that is observations within ±3 hours of the analysis time are used to update the state variables. This step will be discussed in detail in the next section.

4.3.3 Forward operator

As described in Chapter 3, forward operators needed to map the state space to the observation space for calculating observational innovations. For conventional meteorological observations including temperature, wind and humidity, which are directly available in the model, the forward operator is a simple spatial interpolation, which interpolates the gridded model value to the observation location. Whereas a viable satellite data assimilation scheme requires a more complicated forward operator. The forward operator for MOPITT satellite CO total column observations is developed as part of this study. It uses WRF-Chem model variables to calculate the CO total column at the observation locations.

In the first step, WRF-Chem simulates CO mixing ratios on terrain-following hydrostatic pressure levels while MOPITT retrievals are on pressure levels, so an interpolation of model levels to MOPITT ten-level pressure grid is mandatory. For locations where the surface pressure is less than 900 hPa, missing values are expected. For MOPITT Version 5 (V5) products, each retrieval level simply corresponds to a uniformly-weighted

layer immediately above that level. For example, the V5 surface-level retrieval product corresponds to the mean volume mixing ratio over the layer between the surface and 900 hPa. Thus, in the second step, a simple unweighted averaging to the model results in the layers above each retrieval level is applied. For the topmost MOPITT retrieval level at 100 hPa, the uniform-VMR layer extends from 100 hPa to 50 hPa. Assumed VMR values in the layer from 50 hPa to the top of atmosphere (TOA) are based on MOZART-4 model climatology and are fixed. Next, a common logarithm is taken of the averaged VMR so that it is consistent with the data in the observation sequence file. Now the model CO profile $\mathbf{Z_m}$ is available for use. Together with the weighted a priori and averaging kernel from the observation sequence file, the retrieval algorithm in Equation 4.1 is applied to get model equivalent CO retrievals ($\mathbf{Z'_r}$) by replacing the real observations $\mathbf{Z_o}$ with $\mathbf{Z_m}$:

$$\mathbf{Z'_r} = \mathbf{A}\mathbf{Z_m} + (\mathbf{I} - \mathbf{A})\mathbf{Z_a}. \tag{4.3}$$

To calculate the CO total column mixing ratio, the atmosphere is divided into ten partial columns according to the MOPITT ten-level grid. For the i^{th} partial column, the CO mixing ratio is:

$$C(CO) = Z'_r(i) \cdot \frac{\triangle P \cdot N_A}{m \cdot g}, \tag{4.4}$$

where $N_A = 6.022 \times 10^{23}\ mol^{-1}$ is Avogadro's number, A $= 1 \times 10^{-4}\ m^2$ is the unit area, m $= 28.966 \times 10^{-3}$ kg is the molecular weight of air and g $= 9.8\ m \cdot s^{-2}$ is the gravitational acceleration. $\triangle$P is the pressure differences between each retrieval level and the level immediately above, except for the topmost level at 100 hPa. As mentioned

before the partial column for the layer between 100 hPa and 0 hPa is actually composed of a fixed value based on MOZART-4 climatological VMR values from TOA to 50 hPa and a uniform-VMR layer from 50 hPa to 100 hPa. For total column comparisons, the layer from 100 hPa to 0 hPa can be represented as an "equivalent layer" with pressure width 74 hPa and the retrieved VMR at 100 hPa; this approximation yields total column values which are within a few percent of the actual retrieved total column values [Deeter, 2011]. Thus, $\triangle$P for vertical level i can be written as:

$$\begin{aligned} \triangle P_i &= (P_{sfc} - P_{isfc+1}) \quad && (i = isfc) \\ &= 100hPa && (isfc + 1 < i < 10) \\ &= 74hPa && (i = 10) \end{aligned} \tag{4.5}$$

Finally, a summation of the mixing ratio in all the partial columns gives the model equivalent CO total column, and will be compared directly with the MOPITT observations.

4.3.4 Localization and inflation

4.3.4.1 Localization

In fact, every EnKF with a finite number of members suffers from a sampling error. This issue becomes more severe when the correlation between an observation and a state variable is weak. One way to estimate those weak correlations precisely is to increase the ensemble size. Thousands of ensemble members may be required in order to remedy this problem, which is not feasible given the current computational power. Thus for the

limited size of ensemble in practical use we have to deal with the sampling error associated with weak correlations.

It is generally believed that the correlation weakens with physical distances between an observation and a state variable. An observation is strongly correlated to the state vector at the observed location and as the distance increases, this correlation decreases. At some point, when the state is far enough from the observation, their potential correlation can be expected to be insignificant. So we consider only the observations from a local domain surrounding the location of the analysis [Ott et al., 2004]. In DART, a localization function is used to weight the regression. The Gaspari-Cohn [Gaspari and Cohn, 1999], a 5^{th} order compactly supported polynomial function which decays from one to zero as the distance of the observations from the analysis grid point increases is used in this study. It corresponds to gradually increasing the uncertainty assigned to the observations until beyond a certain distance they have infinite uncertainty and thus no influence on the analysis. One parameter that describes the Gaspari-Cohn function is the half-width, the horizontal distance between the maximum and half maximum. When the distance between the model state variable and observation is greater than two times of the half-width, the observation has no impact on the state variable; while for that less than two times of the half-width, it behaves like a Gaussian function. The optimal value of half-width depends on the size of the ensemble. With larger ensemble size, correlations at larger distances can be closely estimated. Therefore, an appropriate half-width can be specified for a certain ensemble size.

Sensitivity tests were carried out to determine the optimal ensemble size and localization cutoff values. Experiments with ensemble size of 20, 40, and 60 were conducted. As the ensemble size increased, the results did not improve significantly, however, the computing time doubled or tripled. As recognized in Oczkowski et al. [2005] and Patil et al. [2001], the smaller the local domain in a model, the smaller the ensemble size that is necessary to properly represent the model dynamics in the local domain. In this thesis, a limited area domain is used and an ensemble size of 20 is enough to capture the uncertainties in the simulation. Given the ensemble size, a second sensitivity test to determine the localization cutoff value was performed. Four simulations using the same configuration only with different localization cutoff values: 0.1, 0.05, 0.025 and 0.0125 radian were conducted. The results showed that a cutoff value of 0.0125 radian, approximately 80 km, gave the best analysis results. This actual result met the initial expectation since a very small ensemble size was used in this study.

For the following case study, an ensemble size of 20 and a cutoff value of 0.0125 radian will be used.

4.3.4.2 Covariance inflation

Besides sampling errors, ensemble Kalman filters are also subject to other sources of errors, such as model errors and interpolation as well as representativeness errors. If these errors are not considered, the EnKF can potentially underestimate the forecast covariance and it becomes overconfident in the model forecast state [Li et al., 2009]. This problem

amplifies as the assimilation cycle goes on. Moreover, when the observations are dense, the covariance will be reduced massively. This would cause the filter to depend more on the model forecast and the analysis loses track of the truth. In order to compensate for this tendency, one approach is to artificially inflate the model forecast (analysis) error covariance matrix before (after) each analysis. This covariance inflation can also be thought as localizing the analysis in time [Wursch, 2013]. When an artificial inflation is applied to the model state covariance, it damps the influence of previous observations on the current analysis. As the analysis cycle goes on, the cumulative effect of inflation at each cycle is to diminish the influence of an observation on future analyses exponentially with time.

The standard covariance inflation method is to multiply the model ensemble covariance by a constant factor. But a single value of inflation is not appropriate for all state variables since the ensemble spread is very sensitive to the observation density. When the observations are dense, the ensemble spread is cut down excessively, and vice versa. So an adaptive inflation [Anderson, 2009], which estimates the appropriate inflation factor at each grid point on the fly would be better for this kind of problem. We also did a controlled experiment with a constant inflation factor and adaptive inflation factors, the results showed that the latter performed better than the former in terms of sustaining an approporate model covariance. Moreover, whether the inflation factor is applied before or after each analysis cycle does not make much difference and we will apply adaptive inflation before each assimilation in our case study.

4.3.5 Overarching driver

The above includes all the major components of the analysis and forecast system. Finally, an overarching program is designed to drive the ensemble forecast system to prepare the model and observation data, advance the model, transform data between WRF-Chem and DART and perform data assimilation. In the two experiments that were conducted in this study, this driver is somehow different and thus will be discussed in each experiment separately.

Chapter 5

Experiment I: Data assimilation

5.1 Introduction

The EnKF/WRF-Chem analysis and forecast system was tested on observations of a forest fire event in British Columbia (BC), Canada in 2010.

2010 was one of the most severe fire seasons for BC. There were over 100 notable fires and approximately 330,000 hectares (ha) was burned costing a financial loss of $220 million. Started in early July 2010, the hot sunshine dried out the forest quickly, but minimal lightning activity kept fire starts down. Nonetheless, on July 28, lightning storms hit the central interior and, in four days, the number of fires province-wide was nearly doubled. Conditions started to calm as mid-August approached, but it was just a brief respite. On August 18, a strong wind event passed through the central interior, causing significant and unprecedented growth on some forest fires. Nearly 100,000 hectares, approximately

one-third of the entire seasons total, were burned in only 24 hours. But as quickly as it started, the fire season petered out. Nevertheless, the long-lived chemical constitutions emitted from the fire event were still in the atmosphere and were transported further inland by the synoptic scale westerly winds. By the end of August, cooler temperatures and precipitation reduced fire activity and decreased the pollutant concentration through wet deposition.

Figure 5.1 [National weather service weather prediction center, a] is the surface analysis chart at 11 A.M. UTC on August 17, 2010. There was a strong high pressure system developing in Western Canada. The associated cold front across British Columbia resulted in a high wind condition, which forced the emissions from the forest fire to be transported in a southwest direction. As the high pressure system matured and moved further south on August 18 (Figure 5.2, National weather service weather prediction center [b]), the warm front on the west side of the high center moved across British Columbia and transported the emissions eastward to Alberta.

In this study, a model simulation of this event was conducted. As stated in Chapter 2, given its mid-range lifetime in the atmosphere, CO was a good tracer for such a forest fire disaster. Hence it was used as an indicator of the forest fire and its impacts on the downwind regions.

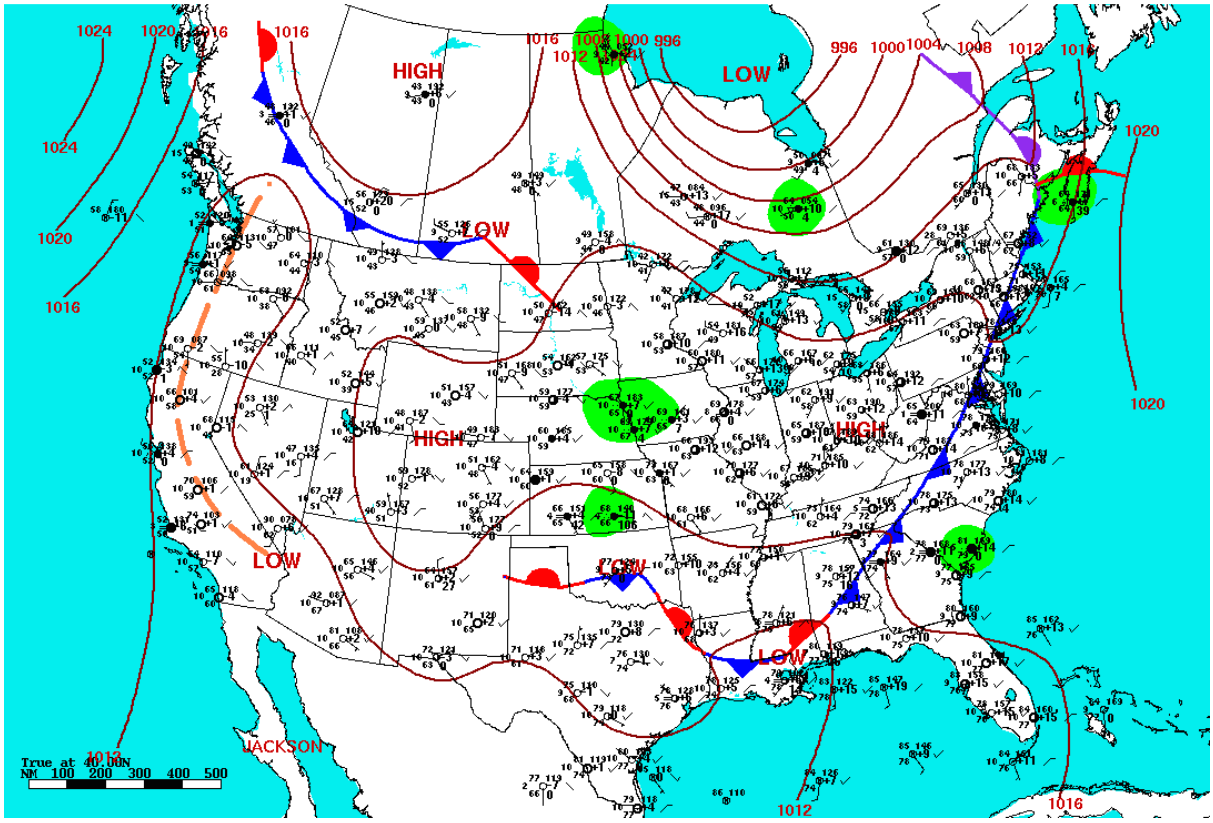

FIGURE 5.1: Surface analysis weather map and station weather plot for 11 A.M. UTC on August 17, 2010. [National weather service weather prediction center, a]

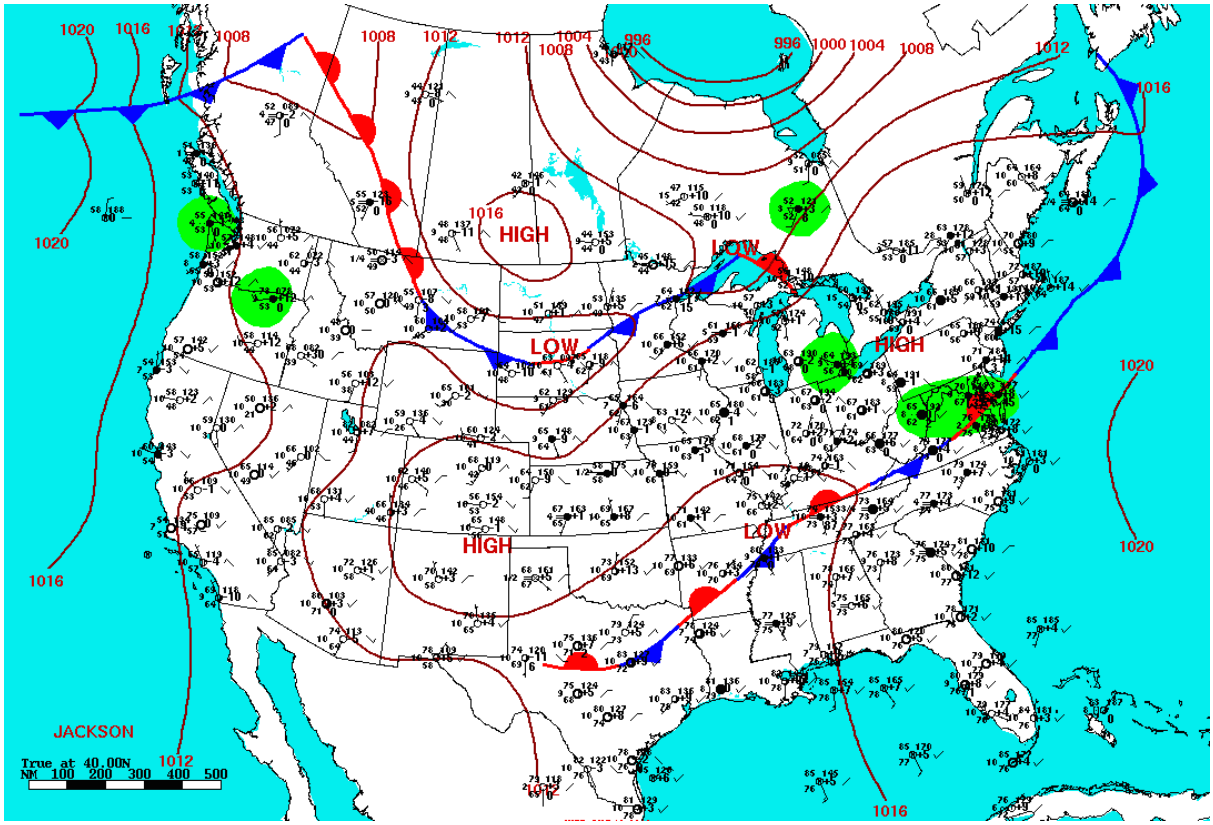

FIGURE 5.2: Surface analysis weather map and station weather plot for 11 A.M. UTC on August 18, 2010. [National weather service weather prediction center, b]

5.2 Simulation configuration

5.2.1 WRF-Chem configuration

5.2.1.1 Domain setup

A single study domain centered over British Columbia was used in this simulation. As showed in Figure 5.3, the domain extended from 45°N to 57°N in latitude and from 140°W to 95°W in longitude. It had 120 grid points in the east-west and 80 in the north-south direction with a uniform horizontal resolution of 30 km. The vertical co-ordinate used in the ARW solver is the terrain-following hydrostatic-pressure levels vary from a value of 1 at the surface to 0 at the upper boundary of the domain. 28 vertical levels, which extend to $\sim$ 50 hPa, were used is this simulation. A simulation with higher resolution of 5 km had also been tested. Results showed that these two experiments had similar broad CO distributions. In this study, the coarse resolution was implemented considering computational efficiency. The colour contour represented the United States Geological Survey (USGS) 24-category land use categories (deatiled index information is given in Appendix A). The fire icons marked the locations of the fire hot spots.

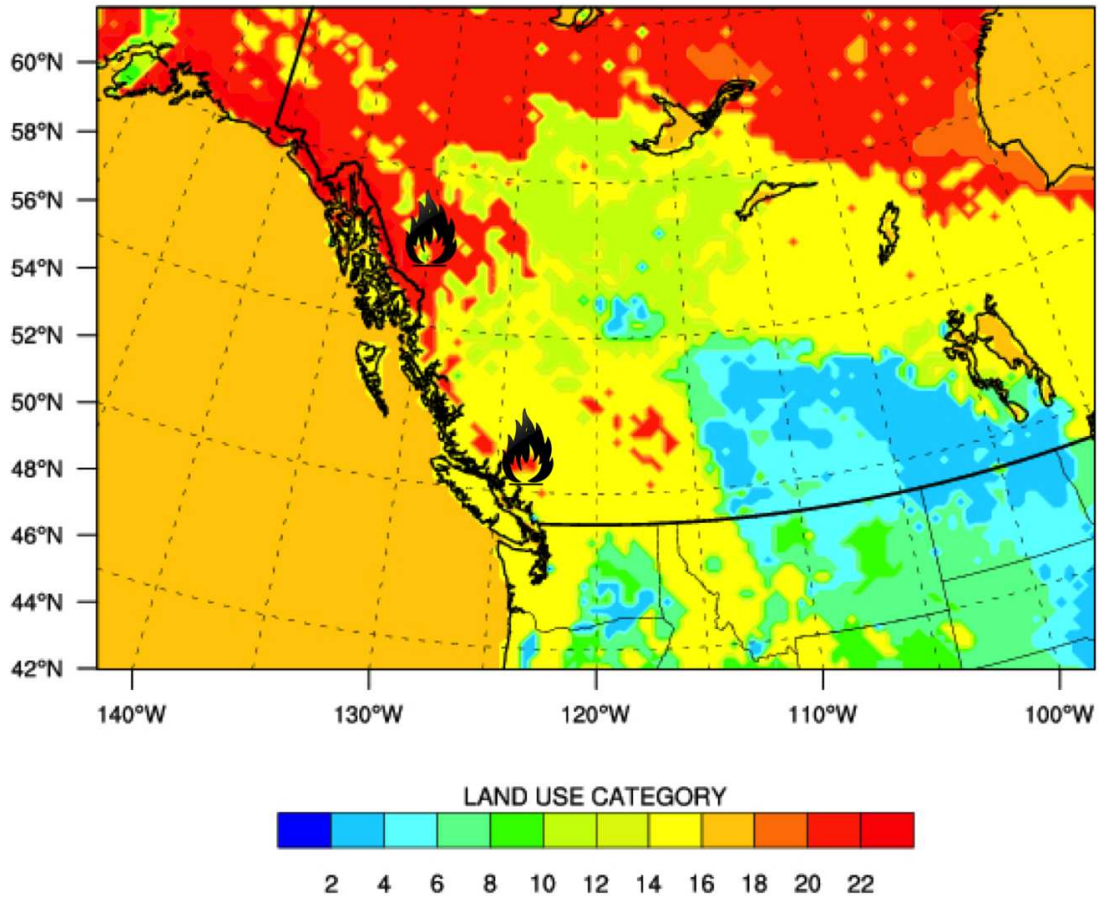

FIGURE 5.3: WRF-Chem simulation domain centered over British Columbia with a horizontal resolution of 30 km. The color contour represented the United States Geological Survey (USGS) 24-category land use categories (deatiled index information is given in Appendix A). The fire icons marked the location of the fire hot spots.

5.2.1.2 Physics and chemistry schemes

The physics and chemistry schemes selected for the WRF-Chem model simulation are given in Table 5.1.

TABLE 5.1: Physics and chemistry schemes selected for the WRF-Chem model simulation

Physics	Scheme
Microphysics	WRF Single-Moment 5-class scheme
Cumulus Parameterization	Grell-3D scheme
Planetary Boundary Layer Physics	Yonsei University scheme
Longwave Radiation	Rapid Radiative Transfer Model scheme
Shortwave Radiation	Goddard shortwave scheme
Chemistry	**Scheme**
Aerosol and Gas Phase Scheme	Regional Acid Deposition Model, 2^{nd} Generation
Photolysis	Madronich Photolysis scheme
Dust Emissions	MOSAIC and MADE/SORGAM Dust Emissions

Considerations for selecting particular schemes depended upon the requirements for this specific study and also their compatibility with other physics and chemistry schemes. The following two paragraphs give a simple overview of all the chosen schemes.

Physics schemes: Microphysical processes were treated in this study by the WRF Single-Moment 5-class (WSM5) scheme [Hong et al., 2004]. This is a mixed phase scheme with five categories of hydrometers: vapour, rain, snow, cloud ice and cloud water. Supercooled water can exist in this scheme. It predicted water vapour and condensate in cloud water, cloud ice, rain and precipitation ice. The sub-grid-scale effects of convective and shallow clouds were handled by the Grell 3D scheme [Grell and Devenyi, 2002]. A unique characteristic of the Grell 3D scheme is that it uses an ensemble of convective

schemes modulating convective triggering threshold parameters, updraft and downdraft entrainment and detrainment parameters, and precipitation efficiency parameters. After calculations for each convective scheme ensemble member, the parameterization passes to the dynamics of the ensemble mean time tendency for temperature, moisture, and cloud and precipitation hydrometeors. Vertical sub-grid-scale fluxes caused by eddy transport were treated through the Yonsei University (YSU) planetary boundary layer scheme [Hong and Dudhia, 2003]. Atmospheric Radiation was simulated by the Rapid Radiative Transfer Model (RRTM) for longwave [Mlawer et al., 1997] and the Goddard shortwave scheme [Chou et al., 1998] for shortwave radiation, respectively.

Chemistry schemes: There are two choices for gas-phase chemical reaction calculations in WRF-Chem: Carbon-Bond Mechanism version Z (CBM-Z) and Regional Acid Deposition Model, 2^{nd} generation (RADM2) [Stockwell et al., 1990]. The RADM2 was chosen in this study since it has the advantage of a good balance between chemical detail, accurate chemical predictions, and available computer resources. Among the inorganic species, there are 14 stable species, 4 reactive intermediates, and 3 abundant stable species (oxygen, nitrogen and water) included in this mechanism. Organic chemistry is represented by 26 stable species and 16 proxy radicals. It is widely used in atmospheric models to predict concentrations of oxidants and other air pollutants. Photolysis frequencies for the 21 photochemical reactions of the gas-phase chemistry model were calculated at each grid point according to Madronich photolysis scheme [Madronich, 1987]. Dry deposition in WRF-Chem is calculated by multiplying the concentrations in the lowest model layer by the spatially and temporally varying deposition velocity, v_d, to give the flux of trace

gases and aerosols to the surface. v_d is proportional to the sum of the aerodynamic, sub-layer, and surface resistance. The parameterization of the surface resistance is developed by Wesley [1989], and depends on the resistances of soil and plant surfaces, the diffusion coefficient, the reactivity, and water solubility of the reactive trace gas. Dust emissions were calculated online using the MOSAIC and MADE/SORGAM dust emissions scheme.

5.2.1.3 Initial and boundary conditions

For the meteorology, results from the NCEP final (FNL) Operational Global Analysis data served both as the initial and lateral boundary conditions. The FNL data had a grid resolution of $1^\circ \times 1^\circ$ in the horizontal and 27 vertical levels extended to 10 hPa. This data was interpolated onto the WRF-Chem grid using the WRF Preprocessing System (WPS). These input data contain the meteorological analyzed fields every six hours.

When it came to chemistry, the chemical initial and boundary conditions (ICBCs) were provided by the MOZART-4 model [Emmons et al., 2010] with $1.9^\circ \times 2.5^\circ$ grid resolution. As was done with the meteorological fields, the MOZART-4 data was first interpolated to the WRF-Chem grid, and the lateral boundaries was also updated every six hours.

5.2.1.4 Emissions

The standard WRF-Chem package, prep_chem_sources, was used to prepare anthropogenic emissions source files. It uses various databases as input and output chemical

emissions at a constant hourly emission rate. This emission was applied in the WRF-Chem model at the beginning of each hour.

The wildfire emissions were estimated following the Fire INventory from NCAR (FINN) [Wiedinmyer et al., 2011], which provides daily, 1-Km resolution, global estimates of the trace gas and particle emissions from wildfires and agricultural fires. Uncertainties in the emissions estimates arise from several steps of the method. The use of fire hot spots, assumed area burned, land cover maps, biomass consumption estimates, and emission factors can all introduce errors into the model. However, the global estimates agree reasonably well with other global inventories of biomass burning emissions for CO, CO_2. Because of its high temporal and spatial resolution and its global coverage, FINN emission estimates had been widely used for modeling atmospheric chemistry and air quality at scales from local to global. Hourly emission rates for various species were specified in a four dimensional array that was input to the WRF-Chem model at the beginning of each hour.

5.2.2 Generation of ensemble members

The initial and boundary conditions generated using the NCEP FNL analysis and MOZART-4 data were the best available estimation of the atmospheric state, and they would be treated as the ensemble mean ICBCs onto which the random perturbations were added to generate the ensemble members.

To perturb the meteorological initial conditions, the WRF model data assimilation system (WRFDA) was used to draw random perturbations from the WRFDA-based background error covariances [Torn et al., 2006]. It would add spatially correlated random perturbations to several specified model state variables including temperature, wind, pressure and water vapour mixing ratio. The boundary conditions were perturbed in a similar approach. The chemical ICBCs and emissions were also perturbed, but in a more straightforward manner. While mapping the MOZART-4 output to WRF-Chem input, instead of directly using the MOZART-4 data, they were first scaled using random numbers from a Gaussian distribution with mean of 1 and spread of 0.3. The updated MOZART-4 data were then combined with the perturbed meteorological ICBCs to provide the ensemble ICBCs for the experiments. For the emission source, a multiplicative parameter called the emission scale factor was defined. This emission scale factor should be one if the standard FINN emission were used. However, in order to generate some perturbations in the emission, the scale factors were also randomly drawn from a Gaussian distribution with mean of 1 and spread of 0.3. The ensemble members generated following these methods were not optimal due to the lack of spatial variation in chemistry. However, it was a minor problem and did not affect the results significantly.

5.2.3 Observations

5.2.3.1 Conventional meteorological observations

Assimilation of meteorological data plays an important role in improving a chemical weather forecast since the meteorological field controls the advection and diffusion processes that determine the distribution of the chemical species.

Binary Universal Form for the Representation (BUFR) of meteorological data is a widely used form for the representation, exchange and archiving of observational data that is approved by the World Meteorological Organization (WMO). NCEP "prepared" BUFR (PrepBUFR), BUFR data after quality control, was assimilated to update the model state vectors in this research. Observations that were used include the temperature and wind fields from radiosonde, aircraft, surface, and satellite observation platforms. The standard surface observations are taken at 00, 06, 12 and 18 UTC daily and radiosonde observations are available at 00 and 12 UTC daily; while aircraft and satellite observations do not have a fixed observation time. Under this circumstance, all observations were divided into four time periods centering on 00, 06, 12 and 18 UTC each with a time window of six hours, i.e. ± 3 hours on each side. All observations within that time window were assimilated sequentially at the same time in this experiment. Figure **??** shows the locations of conventional observations within the six-hour assimilation window centering at 18 UTC on August 16, 2010.

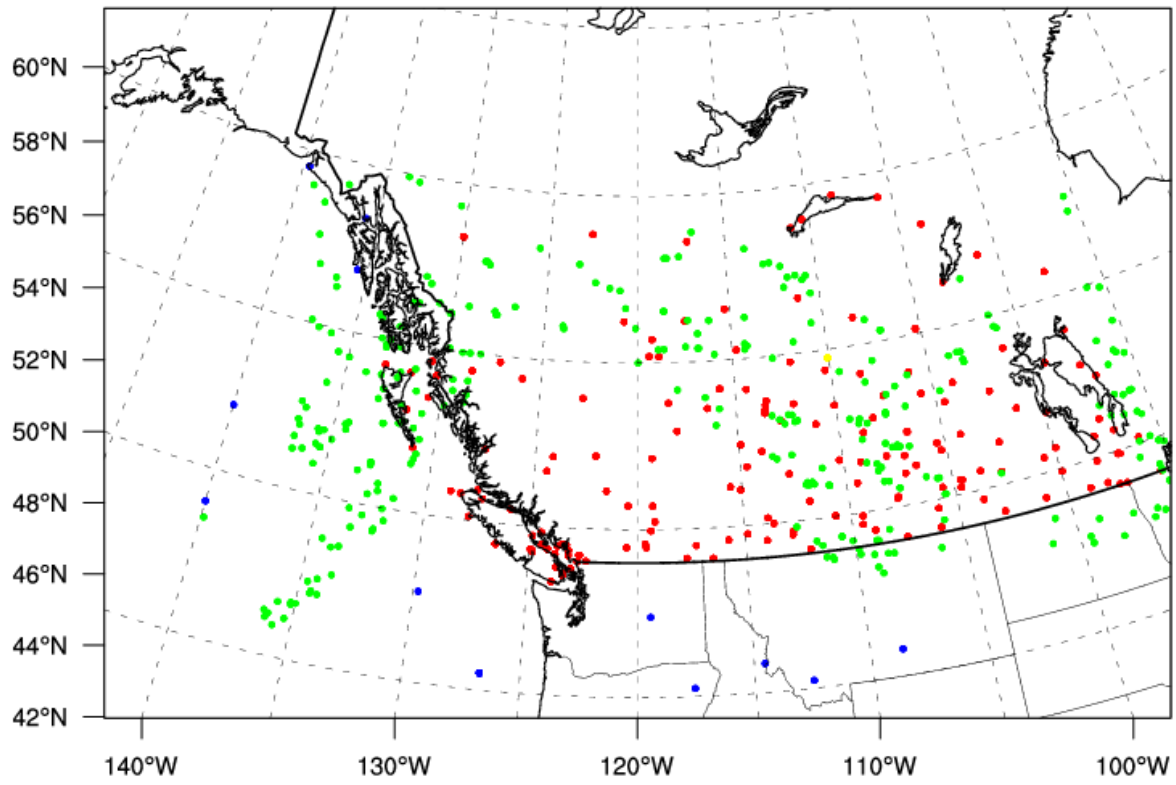

FIGURE 5.4: Observation locations within the six-hour assimilation window centering at 18 UTC on August 16, 2010.

5.2.3.2 MOPITT CO retrievals

The retrievals of CO by the MOPITT instrument onboard the Terra satellite provide an opportunity for the quantitative study of the transport and sources of CO in the atmosphere. A brief introduction of MOPITT was given in Chapter 4.

The Terra satellite is a polar-orbiting spacecraft. It passes a fixed location on Earth twice a day: one is in its descending (daytime) mode and the other is in its ascending (nighttime) mode. At nighttime, in the absence of sunlight, only the thermal infrared emission of CO at the wavelength of 4.7 μm could be measured by MOPITT. This results in larger errors in the retrieval. For this reason, only daytime retrievals were assimilated in this study. MOPITT passes British Columbia around 19 UTC daily. Given the same assimilation period as that in meteorological assimilation, MOPITT observations will be included at 18 UTC daily. Figure 5.5 showed the locations of MOPITT observations within the six-hour assimilation window centering at 18 UTC on August 16, 2010.

The associated MOPITT CO observation errors are estimated and distributed along with the retrievals. These values represent the cumulative error from smoothing error, model parameter error, forward model error, geophysical noise and instrument noise. The CO total column errors are generally one order of magnitude smaller than the retrieved CO total column. The averaged CO total column error from August 13 to August 17, 2010 is 2.6814×10^{17} molec $\cdot\, cm^{-2}$.

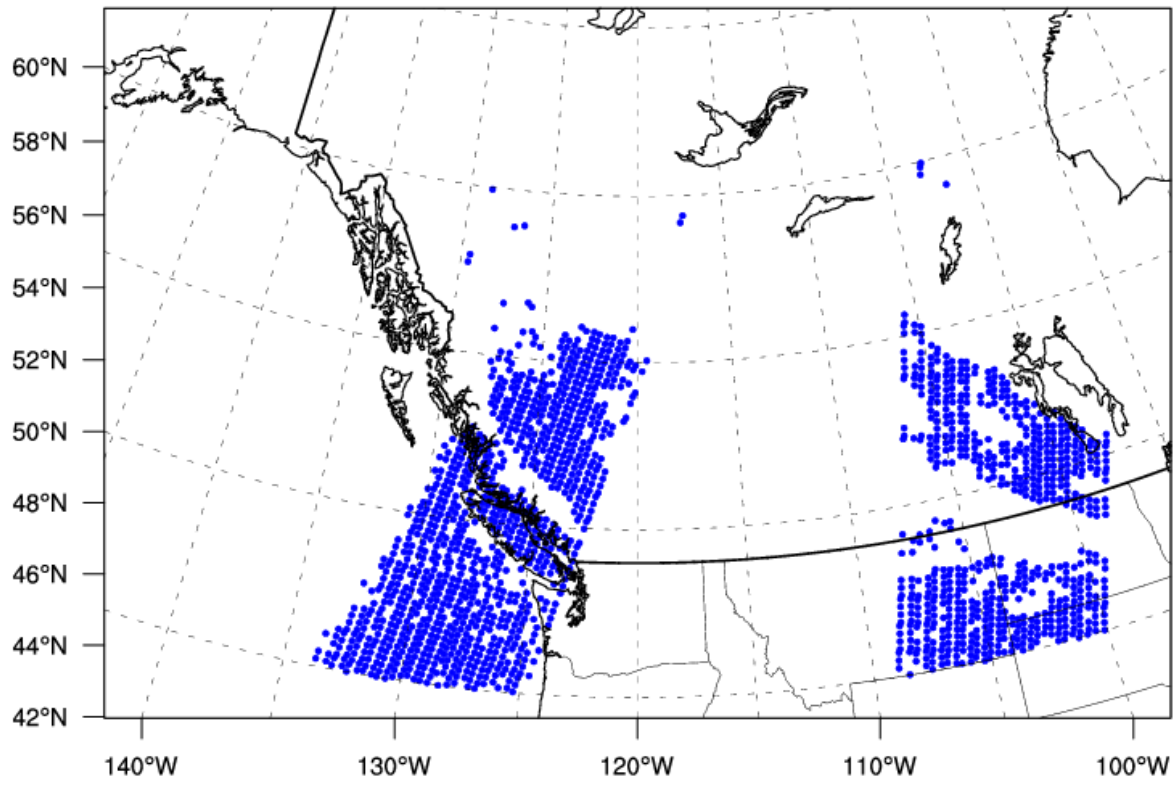

FIGURE 5.5: MOPITT observation locations within the six-hour assimilation window centred at 18 UTC on August 16, 2010.

5.3 Experiment design

In this experiment, two independent simulations were conducted: the control run and the MOPITT run. Both simulations went through three stages: the spin up stage, the assimilation stage and the forecast stage. The spin up stage began at 00 UTC on August 11, 2010. For convenience, this time is written as 2010/08/11/00 (yyyy/mm/dd/hh) UTC and the same simplified time form will be applied here after. The model first ran for two days, until 2010/08/13/00 UTC, to pick up the constituent concentration from the fire event. The two experiments only differed at the assimilation stage. The assimilation stage started right at 2010/08/13/00 UTC and lasted for five days until 2010/08/18/00 UTC. For the control run, only meteorological observations were assimilated at 00, 06, 12, 18 UTC daily. For the MOPITT analysis run, besides the meteorological observations, MOPITT CO total column retrievals were also assimilated but only once at 18 UTC daily. However, since both the observed and calculated CO total column were in molec $\cdot$ cm^{-2}, and the values were normally in the order of 10^{18}; while other state variables were in the order of 10^{0} - 10^{1}. In order to make the data assimilation system more effective, both the observed and calculated CO total column were scaled by 10^{-17} so that they were in the same order with other model state variables. At the end of the assimilation cycle, a six-day forecast was made from the best available initial conditions. The design of this experiment is summarized in Table 5.2.

TABLE 5.2: Design of experiment I: Data assimilation

Experiment	Spin up 2010/08/11/00 - 2010/08/13/00		Assimilation 2010/08/13/00 - 2010/08/18/00		Forecast 2010/08/18/00 - 2010/08/24/00	
	Emission	Assimilation	Emission	Assimilation	Emission	Assimilation
Control	Random perturbation	N.A.	Random perturbation	Meteorological obsrvations	Random perturbation	N.A.
MOPITT	Random perturbation	N.A.	Random perturbation	Meteorological and MOPITT observations	Random perturbation	N.A.

An overarching driver was written to drive the simulations. The driver script included preprocesses, analysis and forecast, and diagnostic phases. At the preprocess phase, the observation sequence files were generated in the DART data format. The ICBCs and emissions were prepared and 20 ensemble members were generated by adding perturbations to the ICBCs and emission scale factors. Settings got complicated at the analysis and forecast phase. For the spin up stage, since no data assimilation was performed, the driver directly called the WRF-Chem model to advance the 20 ensemble members 48 hours in time. At the analysis stage, the driver firstly called the WRF-Chem model to forward the 20 ensemble members to produce a six-hour forecast in parallel. When all of them were finished, it converted the ensemble state vectors to DART format to be read by DART filter. DART assimilated the meteorological observations in the control run and both the meteorological and MOPITT observations in the MOPITT run. Besides assimilation, DART also has an option to evaluate the observations, which will not be used to update the state vectors but will be only used to produce the statistical information about the differences between model simulated and real observations. For comparison purposes, DART will evaluate MOPITT observations in the control run at this stage. After a successful assimilation, the driver converted the updated state vectors back to WRF-Chem format and continued next assimilation cycle until the end of the assimilation stage. For the forecast stage, again, six hour forecast followed by evaluating both the meteorological and MOPITT observations were conducted. These settings are summarized in Table 5.3.

Finally, visualized diagnostics to access the impact of data assimilation were plotted.

TABLE 5.3: Forecast and analysis settings in the overarching driver script for Experiment I: Data Assimilation

		Spin up	Assimilation	Forecast
Control	Forecast period	48h forecast	6h forecast	6h forecast
	Assimilate	N. A.	Met	N. A.
	Evaluate	N. A.	MOPITT	Met and MOPITT
MOPITT	Forecast period	48h forecast	6h forecast	6h forecast
	Assimilate	N. A.	Met and MOPITT	N. A.
	Evaluate	N. A.	N. A.	Met and MOPITT

5.4 Results

5.4.1 Synoptic weather simulation

First of all, the WRF-Chem model performance in terms of its capability to reproduce synoptic scale weather patterns is examined. Figure 5.6 and Figure 5.7 are model-simulated Sea Level Pressure (SLP) valid at 12 UTC August 17 and August 18 respectively. At 12 UTC August 17, there was a strong high pressure system extending from the north to near the Canada-United States border. This high pressure system covered a large area throughout the domain. There was a small low pressure centre to the southeast side of the high system. This high and low system moved eastward as shown in the SLP coutour at 12 UTC August 18. There is a good similarity between model simulations and the surface analysis (Figure 5.1 and Figure 5.2).

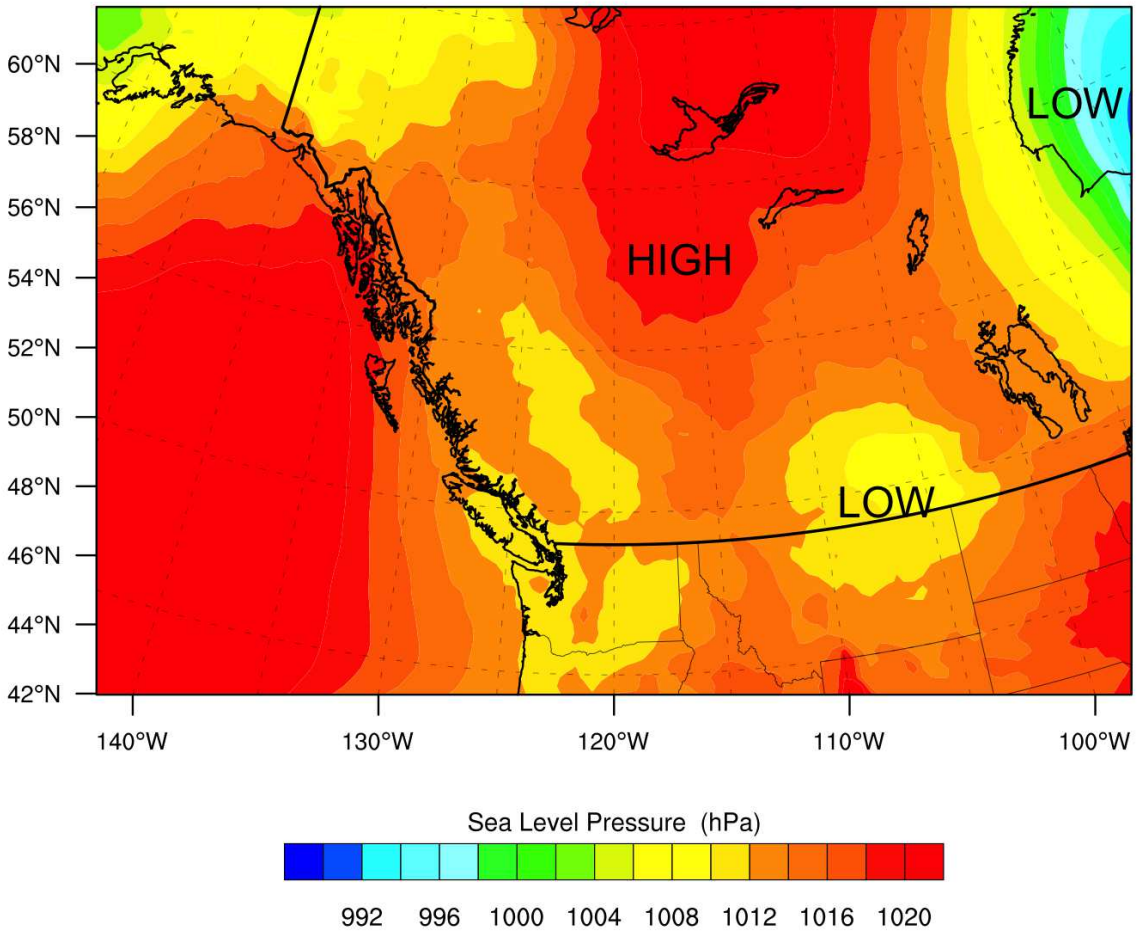

FIGURE 5.6: Contour of simulated Sea Level Pressure (SLP) at 12 UTC on August 17, 2010

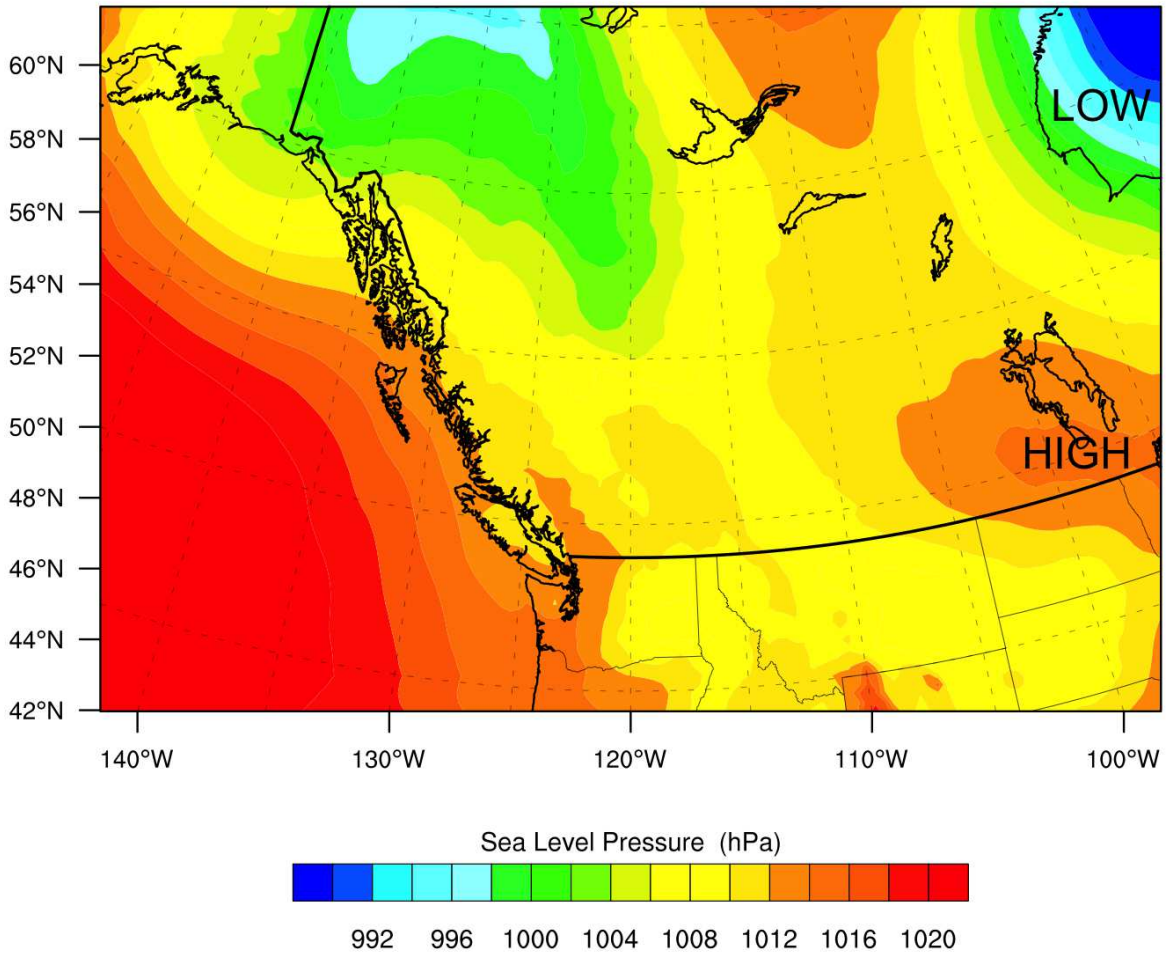

FIGURE 5.7: Contour of simulated Sea Level Pressure (SLP) at 12 UTC on August 18, 2010

5.4.2 Assimilation stage

To evaluate the assimilation performance of the MOPITT run, the domain averaged root-mean-square error (RMSE) of both the ensemble mean forecast (prior) and the analysis (posterior) from the MOPITT observations are computed. Figure 5.8 is a time evolution of the RMSE diagnostic for the assimilation stage. The black and red lines are the RMSEs of the MOPITT run forecast and analysis ensemble mean from the observations respectively. The averaged forecast RMSE is 6.0581 molec $\cdot$ cm^{-2}. Recall that a scale factor of 10^{-17} was applied before the assimilation, the real RMSE should be 6.0581×10^{17} molec $\cdot$ cm^{-2}. After assimilation, this RMSE is reduced to 3.4916×10^{17} molec $\cdot$ cm^{-2}. It could be seen that the analysis reduces the forecast RMSE by approximately 30% at most assimilation cycles, and even more than 65% at some assimilation cycles. In general, data assimilation performs better when there are large model errors. A comparison of the RMSEs between the forecast and analysis suggests that the data assimilation effectively reduces the CO forecast uncertainties and maintains the errors at a steady level. The blue line is the RMSE of the control run analysis ensemble mean from the observations. In the control run, only meteorological observations are assimilated, and this could help to improve the CO simulation in two ways. Firstly, CO mixing ratio can be adjusted through its correlations to the meteorological variables. Secondly, the physical processes, such as advection, are better represented using the updated meteorological variables. The averaged analysis RMSE from the control run is 5.4291×10^{17} molec $\cdot$ cm^{-2}. Although the impacts of assimilating only meteorological observations are not as significant as

assimilating both meteorological and MOPITT observations, it does improve the forecast results in terms of reducing model errors from observations.

The total number of available MOPITT observations at each assimilation time is marked by the blue open circles in Figure 5.8, and the number of MOPITT observations used is marked by the blue crosses. On average, there are 1500 MOPITT observations available at each analysis time, and 50% of them are assimialted while the rest are discarded either by the input quality control or outlier detection processes.

Figure 5.9 demonstrates the CO ensemble spread evolution. The black and red lines are the ensemble spreads of forecast and analysis respectively. The ensemble spread is reduced at each analysis cycle. Partly due to the adaptive inflation, the MOPITT CO spread does not collapse. It even increases as time goes on. This stem primarily from the fact that the forest fire was strongest at August 18, 2010, and the strong windstorm at the same time also increased the model uncertainties. Figure 5.10 is the difference in ensemble spread in U wind component from August 18 compared with August 13 at around 900 hPa level. The domain-averaged ensemble spread in U wind component at August 13 at the same level is 1.4833 m s^{-1}. This spread increses almost over the entire domain. A significant area has increments around 75% and even 200% at some locations.

According to Houtekamer and Mitchell [1998], the total spread (total spread = $\sqrt{(ensemble\ \ spread)^2 + (observation\ \ error)^2}$) shoud be equivalent to the RMSE of the ensemble mean to ensure that the real state of the atmosphere is encompassed by the ensemble. In this experiment, posterior ensemble spread has an average of 1.3692×10^{17}

molec $\cdot$ cm^{-2} and observation error average is 2.6814 $\times$ 10^{17} molec $\cdot$ cm^{-2}, which gives an averaged total spread of 3.0107 $\times$ 10^{17} molec $\cdot$ cm^{-2}. This approximately equals to the mean of the RMSE of the ensemble mean which is 3.4916 $\times$ 10^{17} molec $\cdot$ cm^{-2}. In terms of a short-range air quality forecast, this analysis and forecast system is capable of sustaining enough model spread for an ensemble forecasting.

Examine the analysis at 2010/08/15/18 UTC more closely. The posterior CO total column field is plotted in Figure 5.11. The relative increment in CO total column, i.e. changes in posterior from prior over the prior, after assimilation is plotted in Figure 5.12. Since a small localization radius ($\sim$ 80 km) is specified in the data assimilation process, the increments are only significant within a certain distance from the observations. The biggest increment occurred near the fire hot spot, where there are largest uncertainties. And from the dipole pattern of the increment, one could infer that the model has a lag in the westward transportation of CO and observations contributes to correct this lag.

A vertical profile of the increment near the hot spot is plotted in Figure 5.13. The largest increment is located in the low and mid-troposphere. Given the averaging kernel of MOPITT retrievals shown in Figure 5.14, it is easy to see that this vertical increment pattern is highly related to the shape of the averaging kernel. Whereever the averaging kernel is large, i.e. MOPITT is very sensitive, and the increment is correspondingly large.

Since the CO distribution is strongly affected by the wind field, the analysis of the horizontal wind field is also studied and presented. Figure 5.15 displays the vertical profile

of the five-day averaged RMSE of the horizontal wind from the radiosonde observations. Unlike CO assimilation, which is only sensitive at mid-troposphere, wind assimilation is effective throughout the entire atmosphere. This is because there is no weighting function associated with the horizontal wind field and each level has the same weighting while performing data assimilation. The errors increase as the altitude increases, and reach their maximum at the jet stream level at around 250 hPa. This is primarily the result of the increase of the horizontal wind speed with altitude and its local maximum at the jet stream level. Data assimilation reduces the model RMSE approximately 25% at each level. The time evolution of the model horizontal wind RMSE from radiosonde measurements at two levels: 850 hPa and 400 hPa are plotted in Figure 5.16 and Figure 5.17 respectively. Generally, the wind at 400 hPa is larger than that at 850 hPa, thus a larger RMSE is found at the 400 hPa level. However the performance of the data assimilation is almost the same at the two levels, with slightly better results for situations where the forecast RMSEs are large.

5.4.3 Forecast stage

The comparison of the CO total column RMSE of the six-day forecast of MOPITT assimilation and control run from the observations is plotted in Figure 5.18. The time series of the RMSE shows that the forecast error is reduced after assimilating MOPITT observations, and the benefits of data assimilation are more apparent when the control run has larger RMSE as in the first two days. It is also shown that the data assimilation has reduced effects as time goes on. This is because, as pointed out in the previous chapter,

the CTM depends more on the emission than the initial conditions. In this experiment, only initial conditions were optimized and the emissions scale factors were kept random and not adjusted by the observations. For this reason, a better model forecast result may reqiure to optimze not only the initial conditions but also the emission data.

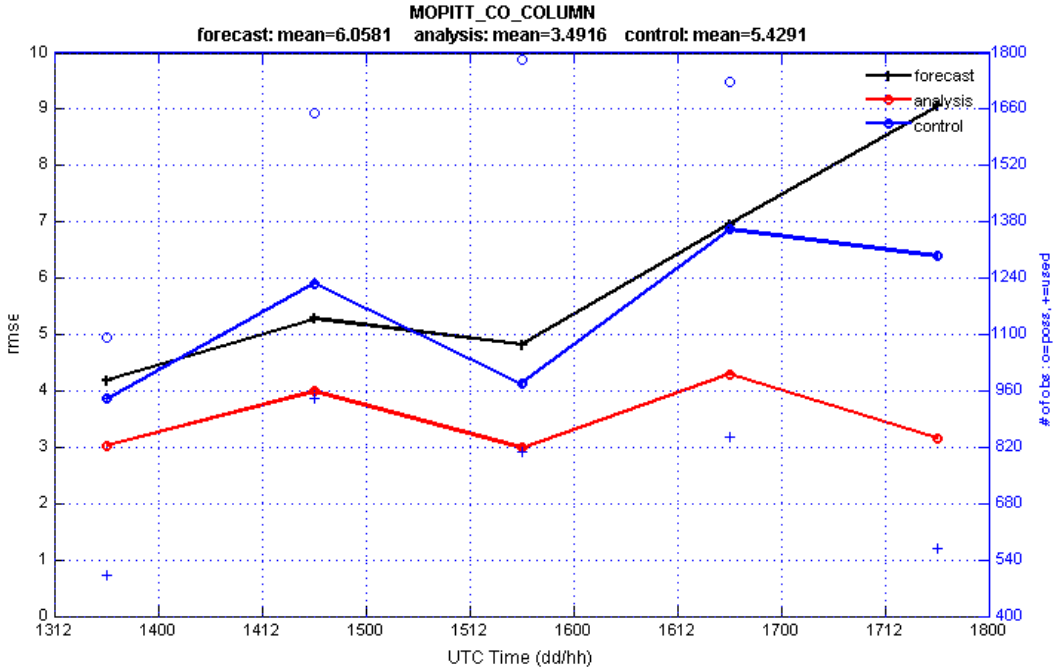

FIGURE 5.8: Domain-averaged MOPITT CO total column root-mean-square error (RMSE) evolution time series. The black and red lines are RMSEs of the forecast and analysis from MOPITT observations respectively. The blue line is the RMSE of the control run analysis ensemble mean from the observation. The open blue circles are the number of available observations and the blue crosses are the number of assimilated observations.

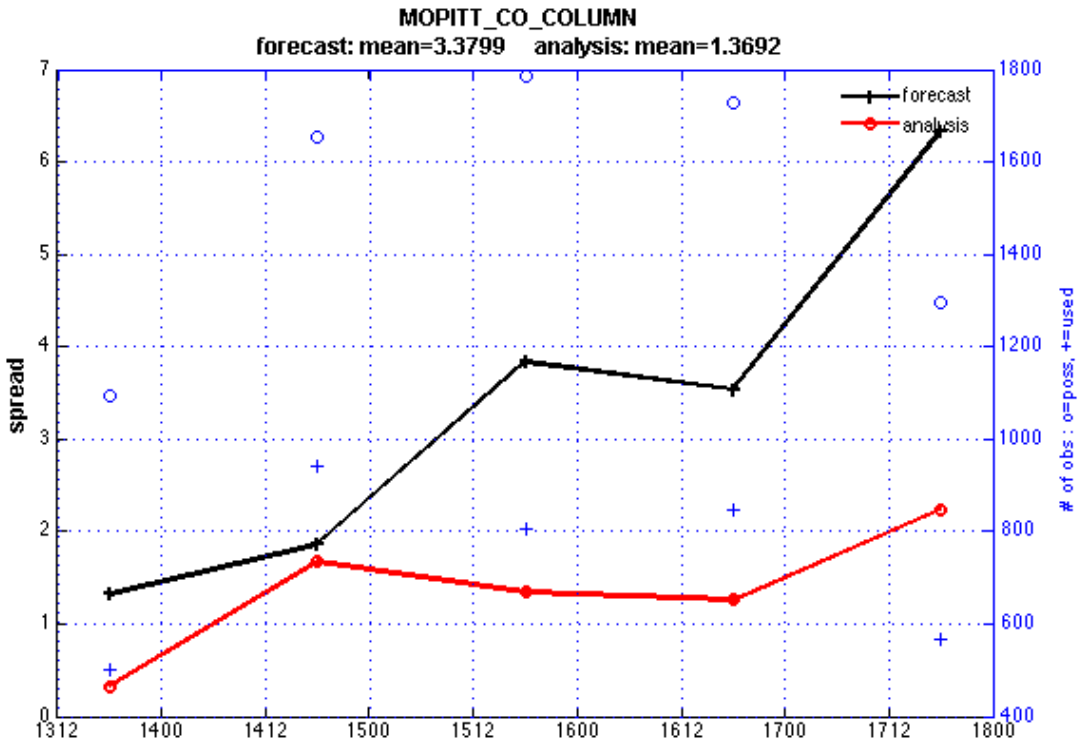

FIGURE 5.9: Domain-averaged MOPITT CO total column spread evolution time series. The black and red lines are the spreads of the forecast and analysis respectively. The open blue circles are the number of available observations and the blue crosses are the number of assimilated observations.

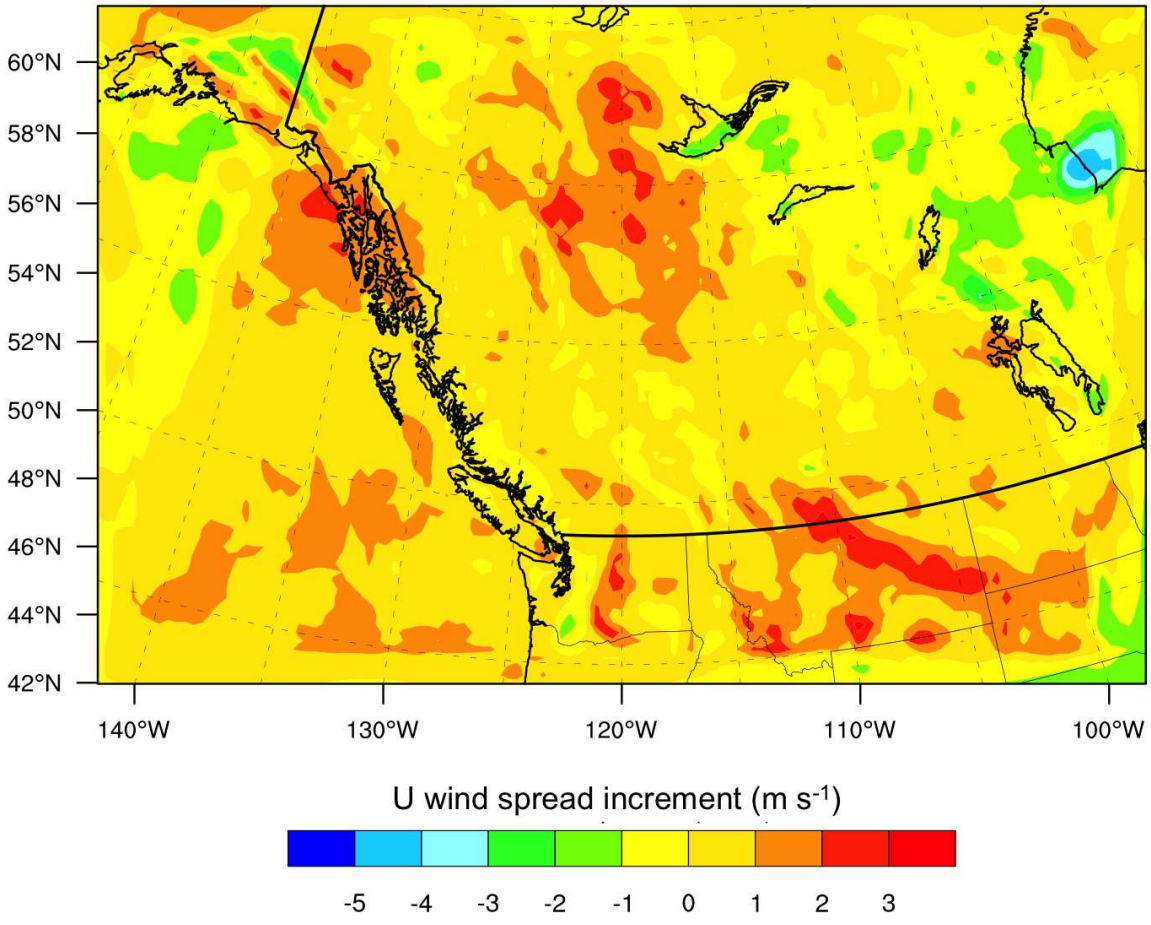

FIGURE 5.10: Difference in ensemble spread in the U wind component from August 18 to August 13 at around the 900 hPa level

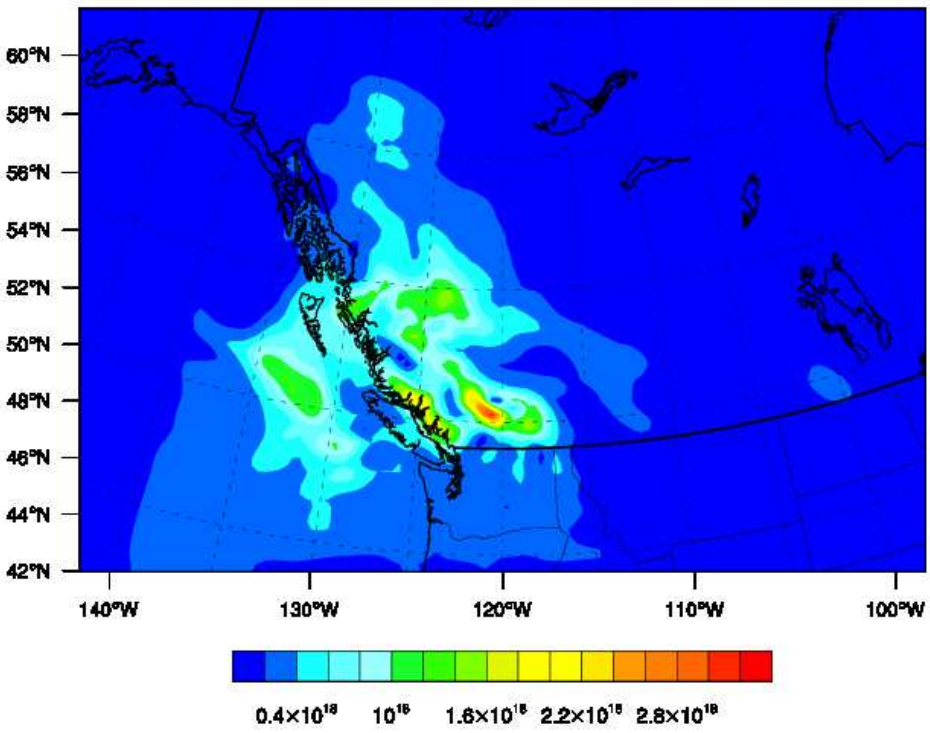

FIGURE 5.11: Posterior MOPITT CO total column at 1800 UTC on August 15, 2010

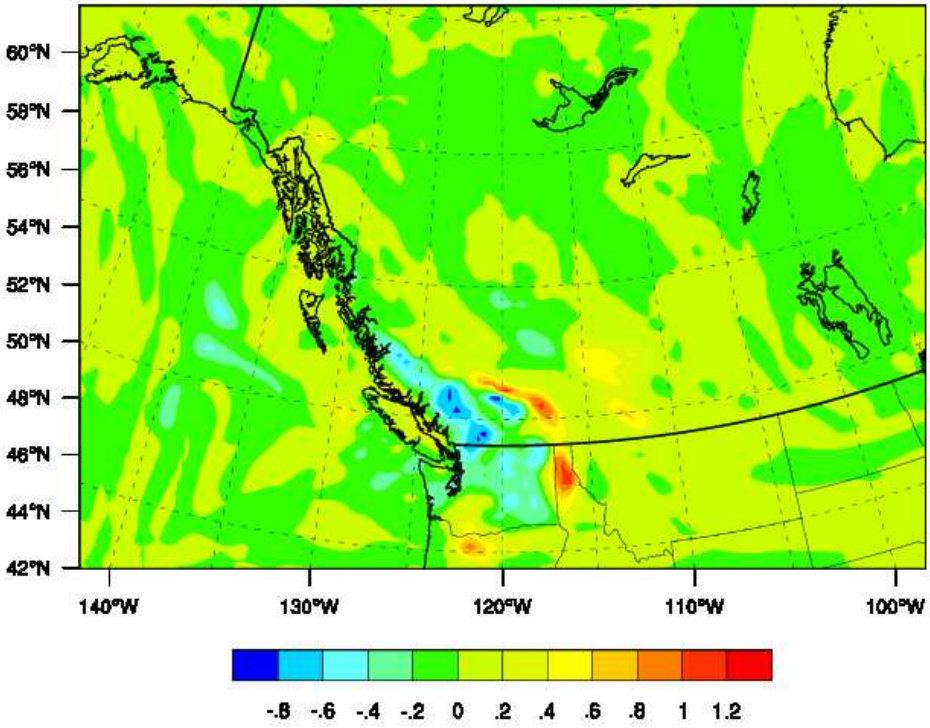

FIGURE 5.12: MOPITT CO total column relative increment, i.e. changes in posterior from prior over prior, at 1800 UTC on August 15, 2010

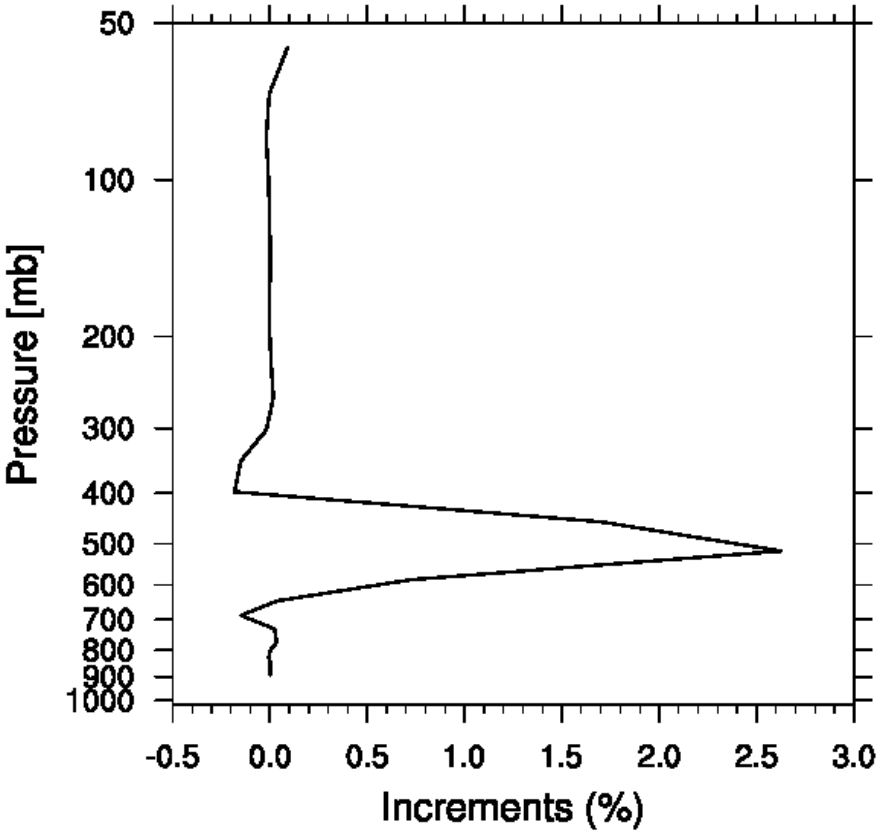

FIGURE 5.13: A vertical profile of CO relative increment near the fire hot spot at 1800 UTC on August 15, 2010

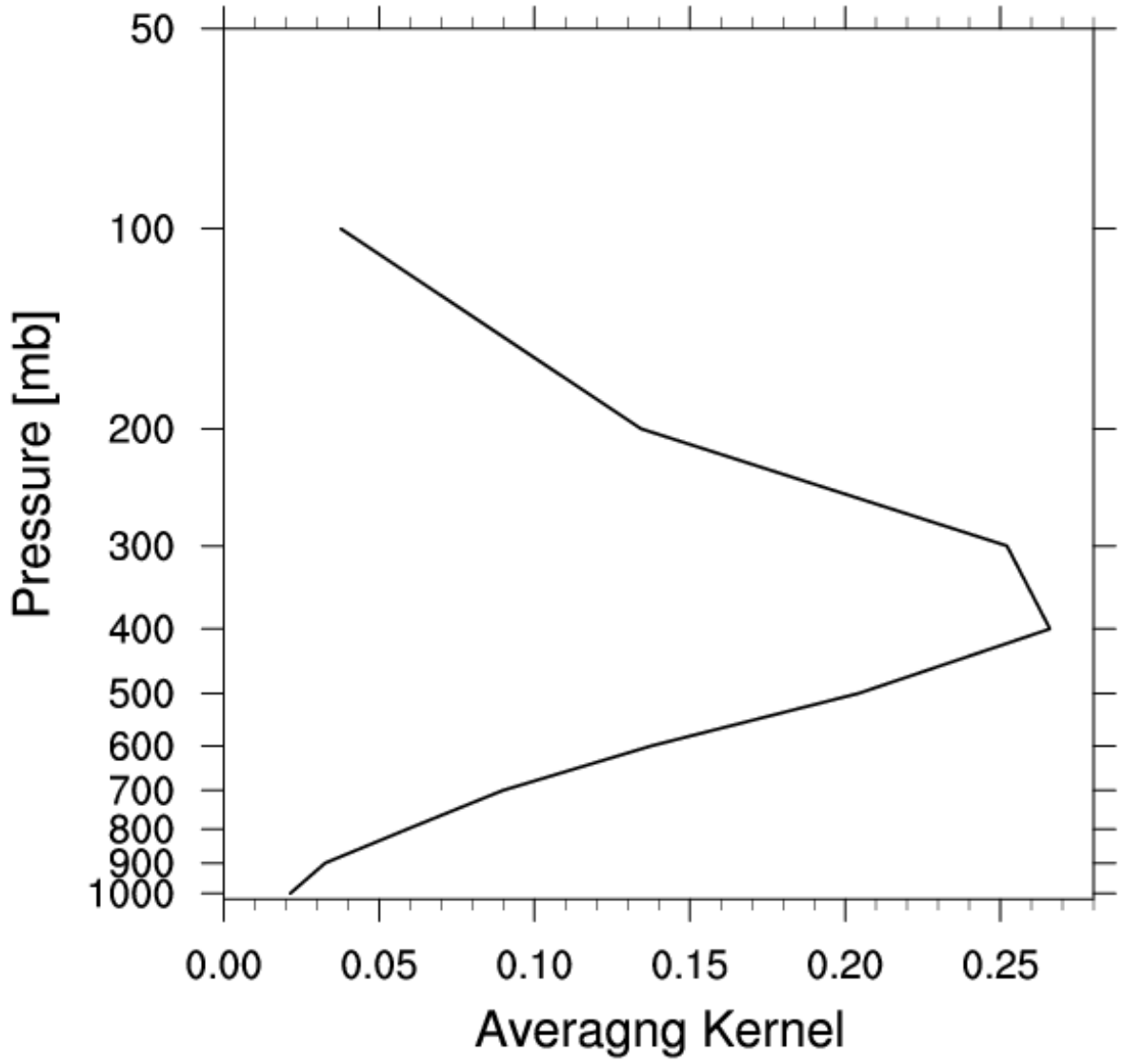

FIGURE 5.14: An averaging kernel profile for MOPITT CO retrievals near the fire hot spot at 1800 UTC at August 15, 2010

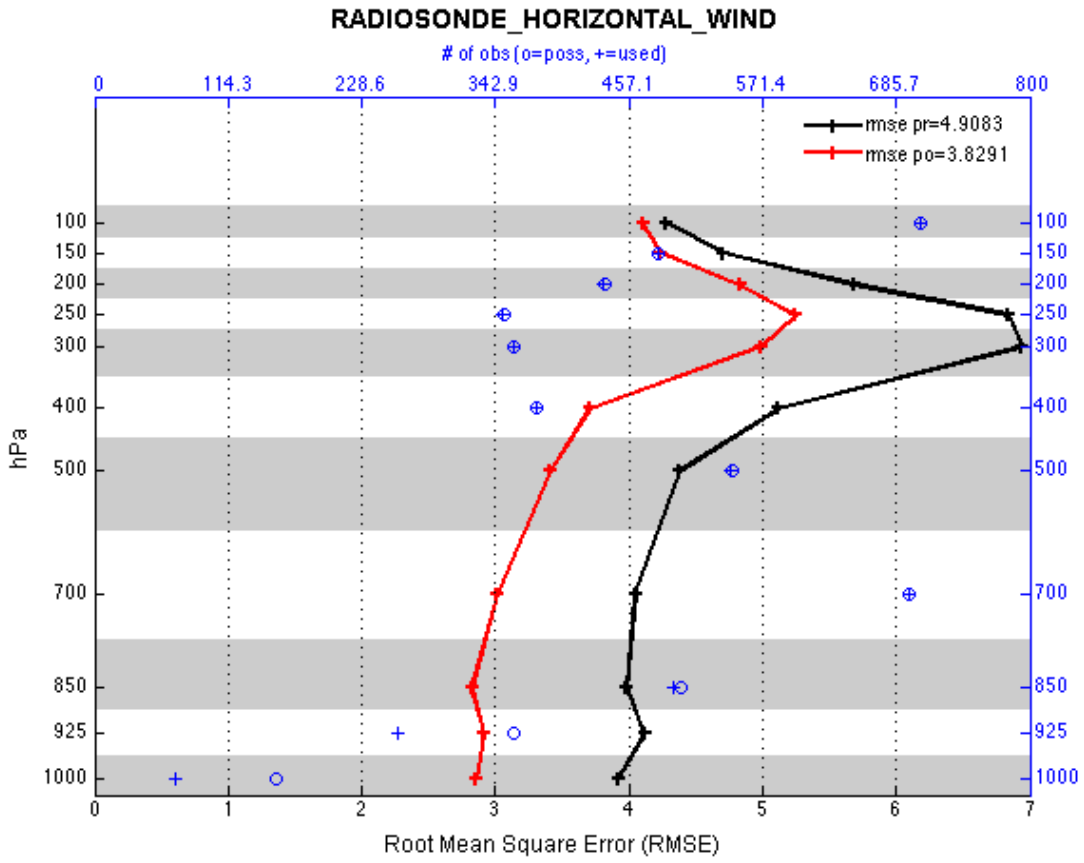

FIGURE 5.15: Five-day averaged vertical profile of the radiosonde horizontal wind root-mean-square error (RMSE). The black and red lines are RMSEs of the forecast and analysis from MOPITT observations respectively. The open blue circles are the number of available observations and the blue crosses are the number of assimilated observations.

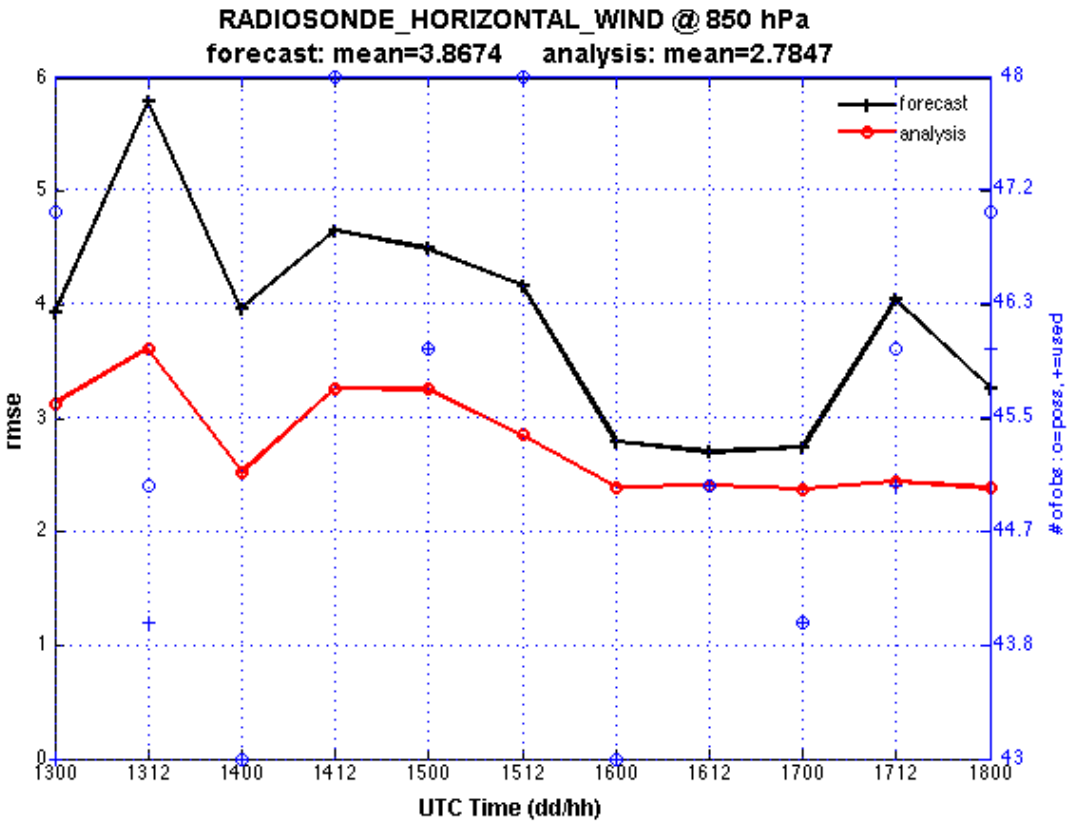

FIGURE 5.16: Radiosonde horizontal wind root-mean-square error (RMSE) evolution time series at 850 hPa. The black and red lines are the RMSEs of the forecast and analysis from MOPITT observations respectively. The open blue circles are the number of available observations and the blue crosses are the number of assimilated observations.

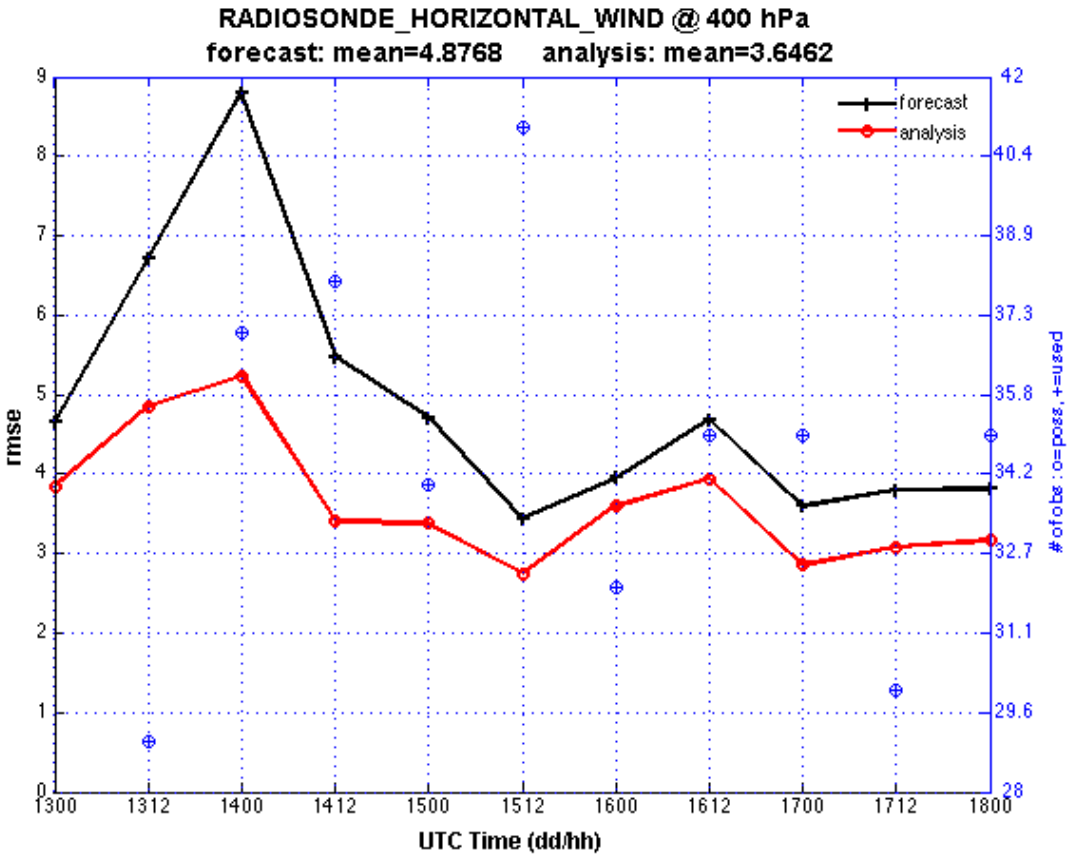

FIGURE 5.17: Radiosonde horizontal wind root-mean-square error (RMSE) evolution time series at 400 hPa. The black and red lines are the RMSE of the forecast and analysis from MOPITT observations respectively. The open blue circles are the number of available observations and the blue crosses are the number of assimilated observations.

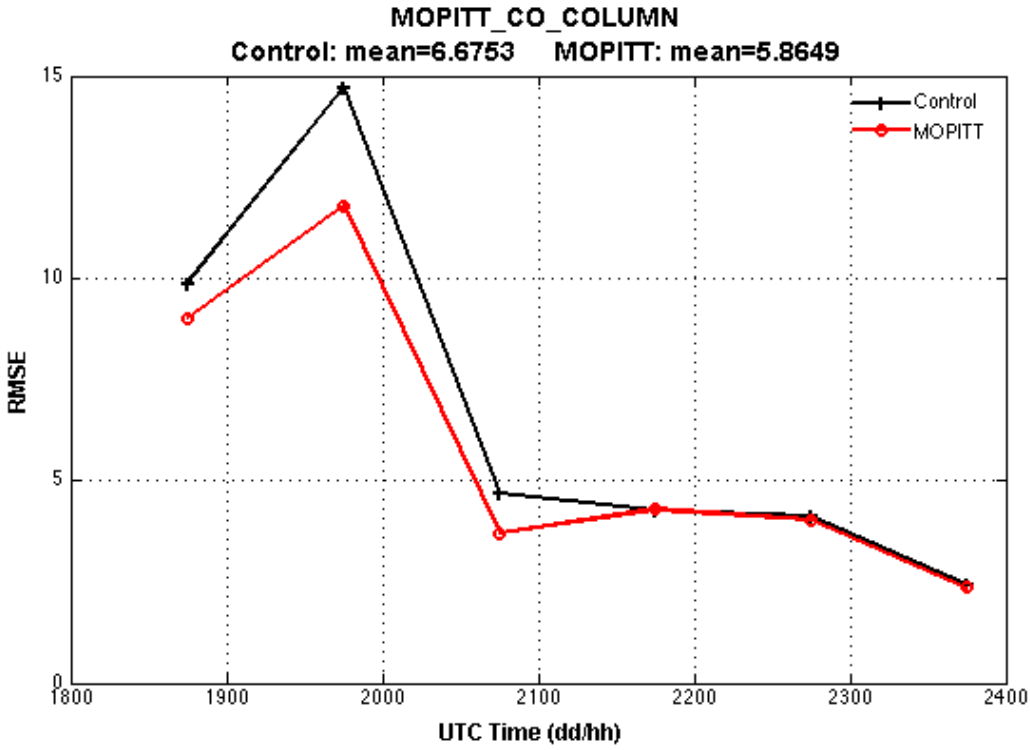

FIGURE 5.18: MOPITT CO total column root-mean-square error (RMSE) evolution time series. The black and red lines are the RMSEs of the control and MOPITT runs respectively.

Chapter 6

Experiment II: Inverse modeling

6.1 Introduction

In order to simulate the fire event, we used the Fire INventory from NCAR (FINN) as the input emissions scenario. This inventory is based on satellite observations of active fires and land cover, together with empirical emission factors and estimated fuel loading. Hence there are large uncertainties in the estimated emissions. Errors arise from the use of fire hot spots, assumed area burned, land cover maps, biomass consumption estimates, and emission factors. As a typical up-scaling approach for estimating emission sources, it has its own limitations.

Inverse modeling through data assimilation offered a completely different way to estimate emission sources. In data assimilation, measurements of atmospheric chemical species mixing ratios are used to determine source distributions that lead to optimal

agreements between model simulations and these observations. In this thesis, an emission scale factor, a multiplicative scaling function applied to the FINN data before it is used in the WRF-Chem model, is defined. If the FINN data are to be used in the simulation then this scale factor would be exactly one. Now this emission scale factor could also be considered as part of the model state alongside the conventional variables, and then the covariance sampled by the ensemble members can be used directly to update this parameter in exactly the same manner as for the state variables. This method is known as parameter estimation and the new state vector is known as the extended state vector. The FINN data multiplied by this optimal emission scale factor is used as the input emissions for next forecast cycle.

The capability of optimizing the CO emissions of the WRF-Chem/DART system was tested in the same BC forest fire case.

6.2 Experiment design

The model configuration used in this experiment was the same as that used in Experiment I.

An additional simulation, MOPITT analysis with optimal emission, was added on top of Experiment I. This study started directly from the analysis stage by using the available output from the spin up stage from the previous experiment. The assimilation stage started right at 2010/08/13/00 (yyyy/mm/dd/hh) UTC and lasted five days until

2010/08/18/00 UTC. At the assimilation stage, the emission scale factor for CO was treated as a model parameter and was appended at the end of the model state vector. The emission scale factor also got updated when assimilating the observations. The updated emission scale factor was used while advancing the model to the next time step. The observations assimilated during this stage were meteorological observations at 00, 06, 12, 18 UTC daily and MOPITT CO total column retrievals at 18 UTC daily. The scale factor 10^{-17}, which was used to scale the CO total column to the same order as other state variables in Experiment I was also applied in this simulation. Comparisons between the MOPITT run and the MOPITT optimal run are summarized in Table 6.1.

TABLE 6.1: Design of experiment II: Inverse modeling

Experiment	Spin up 2010/08/11/00 - 2010/08/13/00		Assimilation 2010/08/13/00 - 2010/08/18/00		Forecast 2010/08/18/00 - 2010/08/24/00	
	Emission	Assimilation	Emission	Assimilation	Emission	Assimilation
MOPITT	Random perturbation	N.A.	Random perturbation	Meteorological and MOPITT observations	Random perturbation	N.A.
MOPITT Optimal	Random perturbation	N.A.	Optimal esitimation	Meteorological and MOPITT observations	N.A.	N.A.

The overarching driver for this experiment was only a minor revise based on that given in the previous experiment. The observations and ICBCs were already available for use and the preprocessing phase was skipped in this experiment. At the analysis and forecast phase, the driver first called the WRF-Chem model to forward the 20 ensemble members to produce a six-hour forecast in parallel. When all of them were finished, it converted the ensemble state vectors as well as the emission factor to form an extended state vector that can be read by the DART filter. DART used both the meteorological and MOPITT observations to update the extended state vector. After a successful assimilation, the driver separated the extended state vector into two parts: the conventional state vector and the optimized emission scale factor. The conventional state vector was converted back to WRF-Chem format and the optimized emission scale factor was multiplied to FINN data to provide optimal emissions for the next forecast cycle. This processes continued until the end of the assimilation stage.

6.3 Results

As in the previous chapter, the domain-averaged ensemble mean root-mean-square errors of both the forecast and the analysis from the MOPITT observations are computed. The results, together with those from the MOPITT run, are plotted in Figure 6.1. The dashed black and red lines are the forecast and analysis RMSEs from the MOPITT run and the solid black and red lines are the forecast and analysis RMSEs from the MOPITT optimal run. In terms of the model forecast errors, the optimal run indeed

improved the model forecasting ability. The original average of the forecast RMSE from the MOPITT run is 6.0581 $\times$ 10^{17} molec $\cdot$ cm^{-2}, and is reduced to an average of 4.5820 $\times$ 10^{17} molec $\cdot$ cm^{-2} after using optimal emission scale factors. Especially when using random perturbations in emission scale factors gives a very large model error at the end of the simulation; applying optimal emission factors reduces this error significantly. This might also due to the accumulative effects of using optimal emissions. However, these two simulations give almost the same analysis RMSE throughout the entire study period. This is reasonable since the same observation data and the same configuration are used for the data assimilation processes. The simulation with optimal emission improved the consistency of the ensemble system.

A relative change in the emission scale factor, i.e. the change in emission scale factor over the prior estimate, is plotted in Figure 6.2 and Figure 6.3. At 2010/08/13/06 UTC (Figure 6.2), the largest change in the emission factor is located near the fire hot spot. The positive increment of the emission factor indicates that the original fire emission resultes in a CO distribution field that is smaller compared with observations. However, the negative increment of the emission factor at 2010/08/16/18 UTC (Figure 6.3) along the west coast indicates that the model tends to give a higher estimation of the CO mixing ratio. These results could be verified by plotting the forecast bias evolution time series of the MOPITT run (Figure 6.4). From the bias plot, it could be clearly seen that the model underestimates the CO mixing ratio at the beginning of the simulation (2010/08/13/00 - 2010/08/16/00) and overpredicts it towards the end of the simulation period (2010/08/16/00 - 2010/08/18/00). The optimal emission tries to correct the model

bias by adding a positive (negative) tendency to the emission whenever model has a low (high) bias. This explaines how the optimal emission contributes to improve the model forecast ability. This also indicates that, updating the emission factor helps to reduce the errors during each assimilation cycles. And these positive effects keep accumulating and result in the best performance at the end of the simulation cycle. Since assimilating the meteorological fields also has an effect on the emission factors, there are still some notable changes further east, however, they would not influence the simulation results since the fire emission hot spots are only along the west coast line.

In conclusion, using the optimal emission from the parameter estimation method has a positive impact on the model forecast precision. Although, this method is not perfect, and there are still model errors and bias, parameter estimation does provide a way to do inverse modeling to get a better estimation of the chemical emission sources using data assimilation techniques.

6.4 Discussion

The results presented above clearly shows that using optimal emission sources increases the model forecast ability. The forecast root-mean-square errors are reduced by 25% on average and more than one third at the end of the analysis cycle. Because of the accumulative effects of using optimal emissions, the model performance keeps improving. We could expect better model forecast results if the experiment were to run for a longer time period.

However, one could argue that this improvement could be the result of artificially forcing the model to the observations. The differences between the model forecast results from MOPITT observations might be coming from other sources, such as: imperfect physical and/or chemical schemes, representive errors, model errors and insufficient model resolution. Adapting emissions may mask the true error sources. Hence further study on this topic is required to quantitatively define this problem. Verification using other independent observations of CO profiles or CO surface flux measurements would be appropriate to address this problem. This would be one direction for future work.

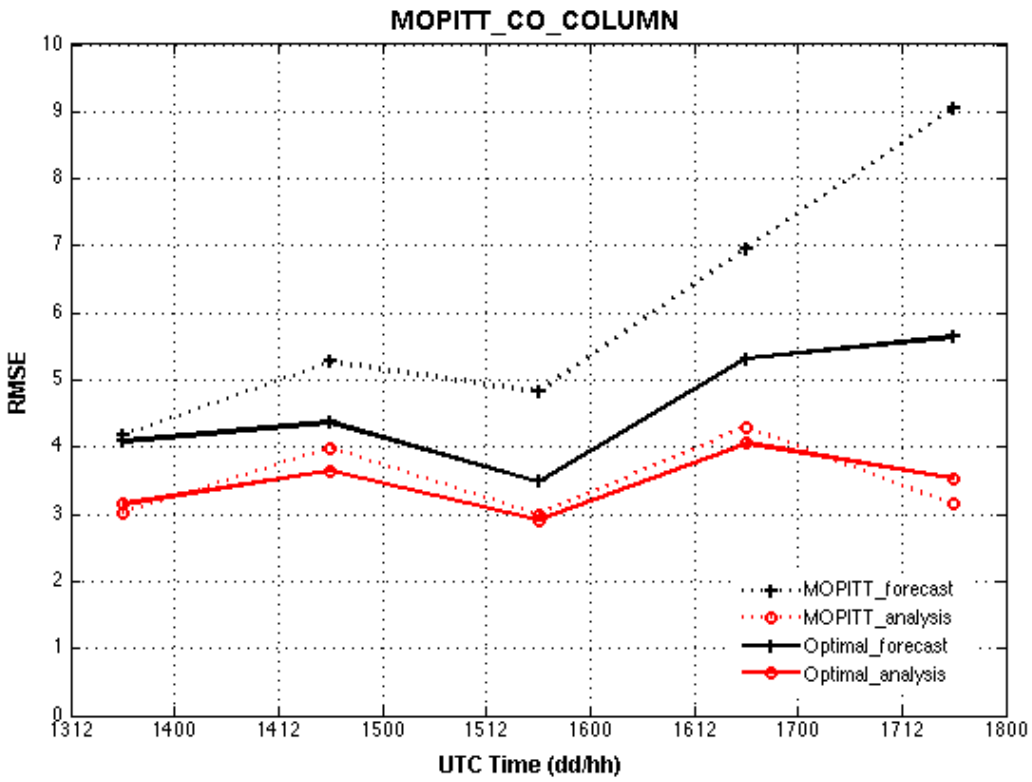

FIGURE 6.1: MOPITT CO total column root-mean-square error (RMSE) evolution time series. The dashed black and red lines are the forecast and analysis RMSEs from the MOPITT run and the solid black and red lines are the forecast and analysis RMSEs from the optimal run.

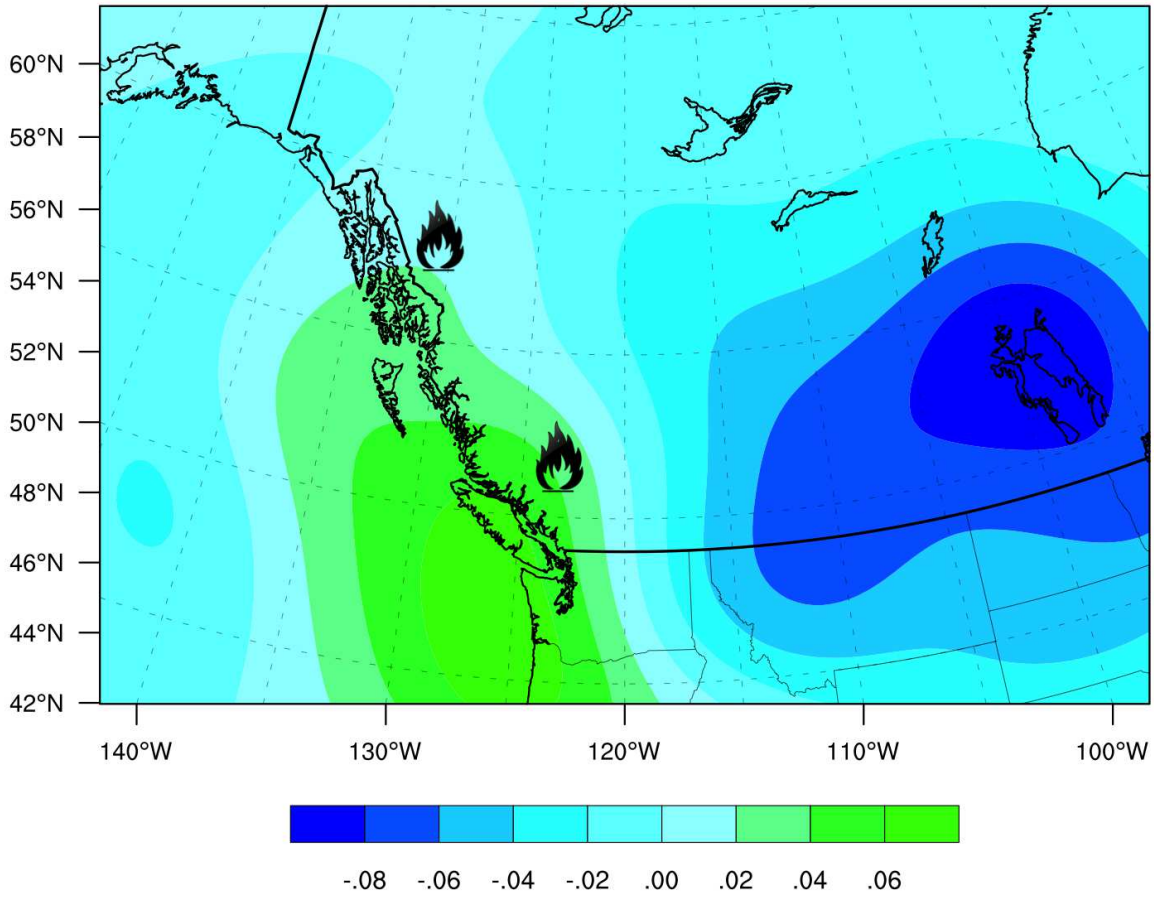

FIGURE 6.2: The relative increment of the emission scale factor, i.e. the change in the emission scale factor over the prior at 0600 UTC on August 13, 2010. The fire icons mark the locations of the fire hot spots.

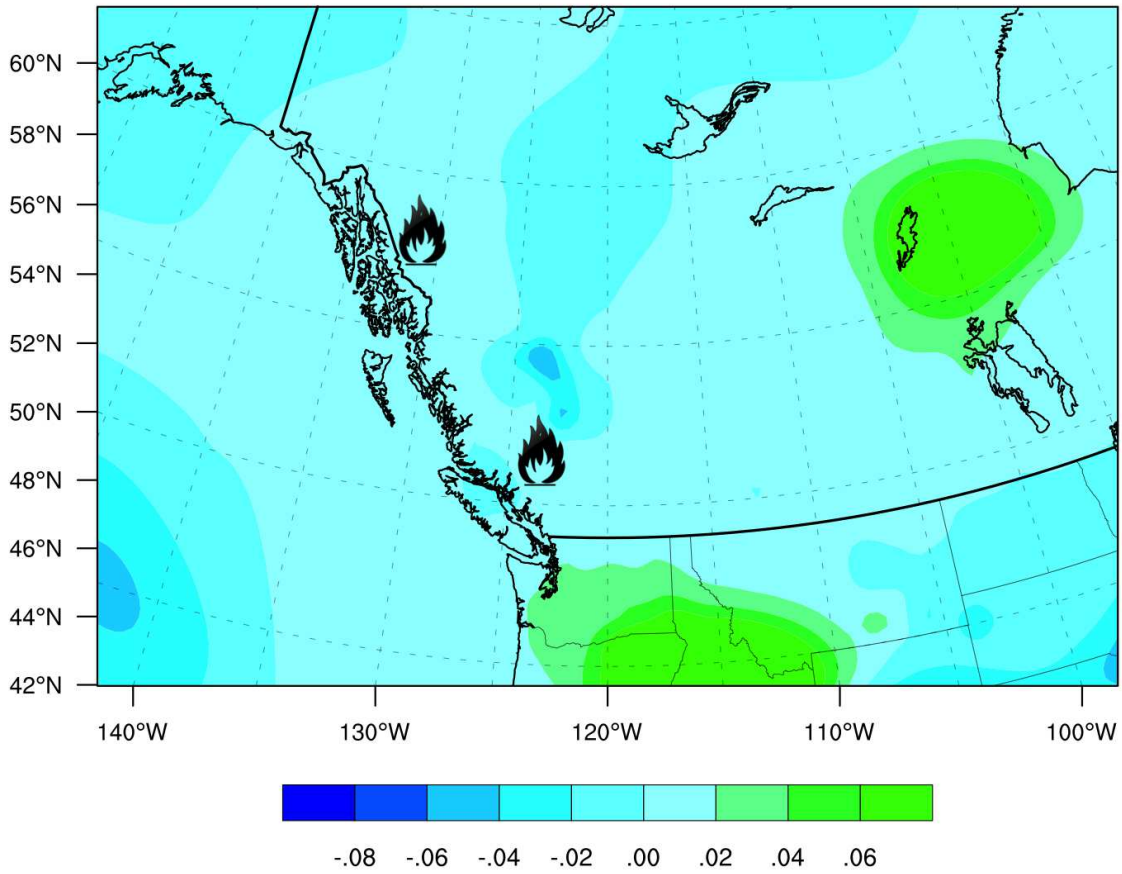

FIGURE 6.3: The relative increment of emission scale factor, i.e. the change in the emission scale factor over the prior at 1800 UTC on August 16, 2010. The fire icons mark the locations of the fire hot spots.

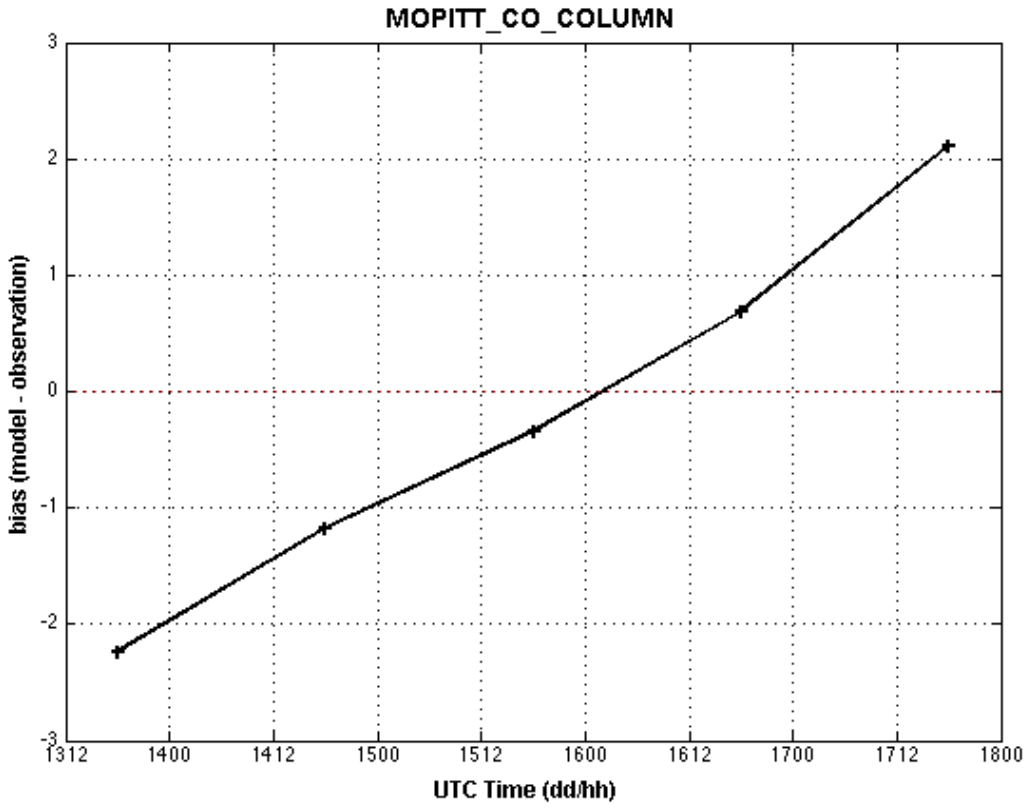

FIGURE 6.4: The time series of the model forecast CO total column bias from observations at the assimilation stage.

Chapter 7

Conclusion and future work

7.1 Thesis conclusions

Compared with numerical weather prediction, air chemistry forecasting still has a large gap to close. In a chemical weather forecast, one tries to maximize the forecast performance under the restriction of current available computational power.

A general introduction of the potential ability of EnKF in advancing air chemistry forecasting was first presented in Chapter 1. With the increasing use of space-based observations, more and more data with a large temporal and spatial coverage are becoming available to continually constrain the numerical models. In recent years, the EnKF has established itself as one of the most successful techniques. The EnKF represents the probability of the model parameters through an ensemble of model realizations and reduces

the dimensionality of the inverse problem from the number of unknown parameters to the number of realizations.

In this study, we focused on the short-range forecasting of the CO field in the atmosphere. Chapter 2 gave a brief introduction of CO's role in the atmosphere. It also introduced how CTMs were developed to describe atmospheric chemical conditions. Chapter 3 reviewed the formulation of the Kalman filter and some common issues and corresponding solutions associated with the ensemble filter.

An EnKF-based analysis and forecast system for atmospheric carbon monoxide prediction was developed for this thesis. Chapter 4 gave a thorough description of handling MOPITT observations in the system, coupling the assimilation program with the chemical models, and applying forward operators.

The analysis and forecast system developed in Chapter 4 were tested on a forest fire case in Chapter 5 and its performance was satisfying. Through localization and inflation, this analysis and forecast system was capable of sustaining enough model spread for an ensemble forecast and it reduced the CO total column forecast RMSE by 42% on average. At some assimilation cycles, this number was even larger than 65%. If we look closer into the vertical levels, the largest increment is located in the low and mid-troposphere, which was primarily due to the vertical profile of the averaging kernel. Assimilation results for the horizontal wind fields were also reviewed given the strong effects the wind field has on the distribution of CO. The horizontal wind field RMSE was reduced by approximately 25% at each level. In terms of the forecast, its RMSE was reduced after assimilating

MOPITT observations, and data assimilation was more effective when the control run had a larger RMSE. It was also shown that the data assimilation had a reduced impact on forecast as time went on. This was because that longer-term CTM depends more on the emission than the initial conditions.

A method to estimate surface CO emission rate via data assimilation was also introduced in Chapter 6. The optimal emission rate estimated from inverse modeling advances the model performance further. The forecast root-mean-square errors were reduced by 25% on average and more than one third at the end of the analysis cycle. Assimilation performed better at the end of the simulation when randomly perturbed emissions gave a very large model error. This might also be the result of the accumulative effects of using optimal emission estimation. The optimal emission rates tended to reduce model bias in each assimilation cycle, and these positive effects kept accumulating and resulted in the best performance at the end of the simulation cycle.

7.2 Future work

The results of this study, although not perfect, are promising. The model forecast ability improved significantly through data assimilation. Furthermore, the methodology of constraining the unobserved surface CO emission by assimilating atmospheric CO observations simultaneously with atmospheric observations provided an alternative approach to estimate CO fluxes. However, one could argue that this improvement could be the

result of artificially forcing the model to the observations. Hence one direction of further work would be to verify the results presented in this study using other independent observations of CO profiles or CO surface flux measurements. If the forecast results with optimal emissions also reduce the RMSE from the independent observations, we can safely come to the conclusion that this analysis and forecast system is efficient in improving short range CO forecast performance and is reliable for emission estimation.

Additionally, the CO observations used in this thesis should be not restricted to MOPITT observations only. Verified observations from the surface observation network, aircraft campaigns, satellite missions and et cetera could also be used to guide numerical model simulations. Further more, using the correlations between CO and other chemical species like CH_4, O_3 and NO_x, a better simulation result of CO field can help to estimate the distribution of other species.

Appendix A

Land use categories

Index	Land Use Description	Index	Land Use Description
1	Urban and Built-up Land	13	Evergreen Broadleaf
2	Dryland Cropland and Pasture	14	Evergreen Needleleaf
3	Irrigated Cropland and Pasture	15	Mixed Forest
4	Mixed Dryland/Irrigated Cropland and Pasture	16	Water Bodies
5	Cropland/Grassland Mosaic	17	Herbaceous Wetland
6	Cropland/Woodland Mosaic	18	Wooden Wetland
7	Grassland	19	Barren or Sparsely Vegetated
8	Shrubland	20	Herbaceous Tundra
9	Mixed Shrubland/Grassland	21	Wooded Tundra
10	Savanna	22	Mixed Tundra
11	Deciduous Broadleaf Forest	23	Bare Ground Tundra
12	Deciduous Needleleaf Forest	24	Snow or Ice

Appendix B

Acronyms

3D-VAR	**T**hree **D**imensional **V**ariational
4D-VAR	**F**our **D**imensional **V**ariational
AFWA	**A**ir **F**orce **W**eather **A**gency
ARW	**A**dvanced **R**esearch **W**RF
BC	**B**ritish **C**olumbia
BUFR	**B**inary **U**niversal **F**orm for the **R**epresentation of meteorological data
CBM-Z	**C**arbon **B**ond **M**echanism version **Z**
CTM	**C**hemical **T**ransport **M**odel
DA	**D**ata **A**ssimilation
DAReS	**D**ata **A**ssimilation **Re**search **S**ection
DART	**D**ata **A**ssimilation **R**esearch **T**estbed
EAKF	**E**nsemble **A**djustment **K**alman **F**ilter
EKF	**E**xtended **K**alman **F**ilter
EnKF	**En**semble **K**alman **F**ilter
EnSRF	**En**semble **S**quare **R**oot **F**ilter
ETKF	**E**nsemble **T**ransform **K**alman **F**ilter
FAA	**F**ederal **A**viation **A**dministration
FINN	**F**ire **IN**ventory from **N**CAR
FSL	**F**orecast **S**ystems **L**aboratory
IACPES	**I**ntegrating **A**tmospheric **C**hemistry and **P**hysics from **E**arth to **S**pace
ICBC	**I**nitial **C**ondition and **B**oundary **C**ondition
KF	**K**alman **F**ilter
KPP	**K**inetic **Pre**processor
LEKF	**L**ocal **E**nsemble **K**alman **F**ilter
MAP	**M**aximum **A** **P**osteriori

MOZART-4	**M**odel for **OZ**one and **R**elated **C**hemical **T**racers version 4
MOPITT	**M**easurements **O**f **P**ollution **I**n **T**he **T**roposphere
NASA	**N**ational **A**eronautics and **S**pace **A**dministration
NCAR	**N**ational **C**enter for **A**tmospheric **R**esearch
NCEP	**N**ational **C**enter for **E**nvironmental **P**rediction
NOAA	**N**ational **O**ceanic and **A**tmospheric **A**dministration
NRL	**N**aval **R**esearch **L**aboratory
NWP	**N**umerical **W**eather **P**rediction
PPMV	**P**arts **P**er **M**illion **V**olume
PrepBUFR	**P**repared **BUFR**
RADM2	**R**egional **A**cid **D**eposition **M**odel 2^{nd} edition
RMSE	**R**oot **M**ean **S**quare **E**rror
RRTM	**R**apid **R**adiative **T**ransfer **M**odel
SLP	**S**ea **L**evel **P**ressure
TOA	**T**op **O**f **A**tmosphere
USGS	**U**nited **S**tates **G**eological **S**urvey
UTC	**C**oordinated **U**niversal **T**ime
VMR	**V**olume **M**ixing **R**atio
WPS	**W**RF **P**reprocessing **S**ystem
WRF	**W**eather **R**esearch and **F**orecasting
WRF-Chem	**W**eather **R**esearch and **F**orecasting model coupled with **C**hemistry
WRFDA	**W**eather **R**esearch and **F**orecasting **D**ata **A**ssimilation
WSM5	**W**RF **S**ingle **M**oment 5-class
YSU	**Y**onsei **U**niversity

Chemical formula

CO	Carbon monoxide
CO_2	Carbon dioxide
CH_4	Methane
H_2S	Hydrogen Sulfide
HCFC	Hydrochlorofluocarbon
NMHC	Non Methane Hydrocarbon
O_3	Ozone
OH	Hydroxyl radical
SO_2	Sulfur dioxide

Appendix C

List of symbols

$\mathbf{A}_d$	Linear operator in EAKF
$\mathbf{A}$	Averaging kernal
$\mathbf{H}$	Linear observation operator
$\mathcal{H}$	Non-linear observation operator
$\mathbf{I}$	Identic matrix
$\mathbf{J}(\mathbf{x})$	Cost function
$\mathbf{K}$	Kalman gain
$\mathbf{M}$	Linear model dynamics
$\mathcal{M}$	Non-linear model dynamics
$\mathbf{N}$	Ensemble size
$\mathbf{P}^f$	Forecast error covariance
$\mathbf{P}^a$	Analysis error covariance
$\mathbf{Q}$	Random model error
$\mathbf{R}$	Observation error covariance
$\mathbf{v}_d$	Deposition velocity
$\bar{\mathbf{x}}$	Ensemble mean state vector
$\mathbf{X}'$	Ensemble state vector perturbation
$\mathbf{x}^a$	Analysis state vector
$\mathbf{x}^f$	Forecast state vector
$\mathbf{x}^{true}$	True state vector
$\mathbf{y}$	Observation
$\mathbf{Z}_a$	Priori profile
$\mathbf{Z}_m$	Model simulated profile
$\mathbf{Z}_o$	Observed profile
$\mathbf{Z}_r$	Retrived profile
$\mathbf{Z}'_r$	Model equivalent retrived profile

References

Aksoy, A., Zhang, F., and Nielsen-Gammon, J. W. (2006). Ensemble-based simultaneous state and parameter estimation with MM5. *Geophysical Research Letters*, 33(12).

Alexe, M. and Sandu, A. (2011). Adaptive solution of time-dependent inverse problems with the discrete adjoint method. In *International Conference on Computational Science 2011*, Bali, Indonesia.

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

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

Anderson, J. L., Hoar, T., Reader, K., Liu, H., Collins, N., Torn, R., and Avellano, A. (2009). The data assimilation research testbed (DART).

Barlow, R. J. (1989). *Statistics: A guide to the use of statistical methods in the physical sciences*. John Wiley and Sons.

Bei, N., de Foy, B., Lei, W., Zavala, M., and Molina, L. T. (2008). Using 3dvar data assimilation system to improve ozone simulations in the Mexico City basin. *Atmospheric Chemistry and Physics*, 8(24):7353–7366.

Bishop, C. H., Etherton, B. J., and Majumdar, S. J. (2001). Adaptive sampling with ensemble transform kalman filter. Part I: Theoretical aspects. *Monthly Weather Review*, 129:420–436.

Brasseur, G. P., Orlando, J. J., and Tyndall, G. S. (1999). *Atmospheric chemistry and global change*. Oxford University Press, New York.

Burgers, G., Leeuwen, P. J., and Evensen, G. (1998). Analysis scheme in the ensemble kalman filter. *Monthly Weather Review*, 126(6):1719–1924.

Chai, T., Carmichael, G. R., Tang, Y., Sandu, A., Heckel, A., Richter, A., and Burrows, J. P. (2009). Regional NOx emission inversion through a four-dimensional variational approach using SCIAMACHY tropospheric NO2 column observations. *Atmospheric Environment*, 43(32):5046–5055.

Chou, M. D., Suarez, M. J., Ho, C. H., Yan, M. M. H., and Lee, K. T. (1998). Parameterizations for cloud overlapping and shortwave single-scattering properties for use in general circulation and could ensemble models. *Journal of Climate*, 11:202–214.

Clark, H. L., Cathala, M. L., Teyssedre, H., Cammas, J. P., and Peuch, V. H. (2007). Cross-tropopause fluxes of ozone using assimilation of MOZAIC observations in a global CTM. *Tellus B*, 59(1):39–49.

Compo, G. P., Whitaker, J. S., and Sardeshmukh, P. D. (2006). Feasibility of a 100-year reanalysis using only surface pressure data. *Bulletin of the American Meteorological Society*, 87:175–190.

Constantinescu, E. M., Sandu, A., Chai, T., and Carmichael, G. R. (2007a). Assessment of ensemble-based chemical data assimilation in an idealized setting. *Atmospheric Environment*, 41(1):18–36.

Constantinescu, E. M., Sandu, A., Chai, T., and Carmichael, G. R. (2007b). Ensemble-based chemical data assimilation I: General approach. *Quarterly Journal of the Royal Meteorological Society*, 133:1229–1243.

Deeter, M. N. (2011). MOPITT (measurements of pollution in the troposphere) version 5 product user's guide. User's guide, National Center for Atmospheric Research.

Deeter, M. N., Worden, H. M., Edwards, D. P., Gille, J. C., Mao, D., and Drummond, J. R. (2011). MOPITT multispectral CO retrievals: Origins and effects of geophysical radiance errors. *Journal of Geophysical Research*, 116:D15303–D15303.

Derber, J. (1989). A variational continuous assimilation scheme. *Monthly Weather Review*, 117:2437–2446.

Dethof, A. and Holm, E. V. (2004). Ozone assimilation in the ERA-40 reanalysis project. *Quarterly Journal of the Royal Meteorological Society*, 130:2851–2872.

Drummond, J. R. (1996). MOPITT mission description document. Internal report, University of Toronto, Toronto, Ontario, Canada.

Emerson, S. R. and Hedges, J. I. (2008). *Chemical Oceanography and the Marine Carbon Cycle*. Cambridge University Press, Cambridge, UK.

Emmons, L. K., Walters, S., Hess, P. G., Lamarque, J. F., Pfister, G. G., Fillmore, D., Granier, C., Guenther, A., Kinnison, D., Laepple, T., Orlando, J., Tie, X., Tyndall, G., Wiedinmyer, C., Baughcum, S. L., and Kloster, S. (2010). Description and evaluation of the model for ozone and related chemical tracers, version 4 (MOZART-4). *Geoscientific Model Development*, 3:43–67.

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

Evensen, G. (2003). The ensemble kalman filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, 53:343–367.

Fast, J. D., Gustafson, W. I., Easter, R. C., Zaveri, R. A., Barnard, J. C., Chapman, E. G., Grell, G. A., and Peckham, S. E. (2006). Evolution of ozone, particulates, and aerosol direct radiative forcing in the vicinity of houston using a fully coupled meteorology-chemistry-aerosol model. *Journal of Geophysical Research*, 111:D21305–D21305.

Flemming, J., Inness, A., Jones, L., Eskes, H. J., Huijnen, V., Schultz, M. G., Stein, O., Cariolle, D., Kinnison, D., and Brasseur, G. (2011). Forecasts and assimilation experiments of the antarctic ozone hole 2008. *Atmospheric Chemistry and Physics*, 11(5):1961–1977.

Gaspari, G. and Cohn, S. E. (1999). Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, 125:723–757.

Grell, G. A. and Devenyi, D. (2002). A generalized approach to parameterizing convection combining ensemble and data assimilation techniques. *Geophysical Research Letters*, 29:38–1–38–4.

Grell, G. A., Fast, J. D., Gustafson, W. L., Peckham, S. E., McKeen, S. A., Salzmann, M., and Freitas, S. (2011). On-line chemistry within WRF: Description and evaluation of a state-of-the-art multiscale air quality and weather prediction model. In Baklanov, A., Mahura, A., and Sokhi, R., editors, *Integrated Systems of Meso-Meteorological and Chemical Transport Models*. Springer.

Grell, G. A., Peckham, S. E., Schmitz, R., McKeen, S. A., Frost, G., Skamarock, W. C., and Eder, B. (2005). Fully coupled online chemistry within the WRF model. *Atmospheric Environment*, 39:6957–6975.

Gruber, N., Friedlingstein, P., Field, C. B., Valentini, R., Heimann, M., Richey, J. E., Lankao, P. R., Schulze, D., and Chen, C. A. (2004). The vulnerability of the carbon cycle in the 21st century: An assessment of carbon-climate-human interactions. In Field, C. B. and Raupach, M. R., editors, *The Global Carbon Cycle: Integrating Humans, Climate and the Natural World*. Island Press, Washington D. C.

Gupta, M, L., Cicerone, R. J., and Elliott, S. (1998). Perturbation to global tropospheric oxidizing capacity due to latitudinal redistribution of surface sources of NOx, CH4 and CO. *Geophysical Research Letters*, 25(21):3931–3943.

Hamill, T. M. (2004). Ensemble-based atmospheric data assimilation. Technical report, University of Colorado and NOAA-CIRES Climate Diagnostics Center, Boulder, Colorado, USA.

Hong, S. Y. and Dudhia, J. (2003). Testing of a new non-local boundary layer vertical diffusion scheme in numerical weather prediction applications. In *20th Conference on*

Weather Analysis and Forecasting/16th Conference on Numerical Weather Prediction, Seattle, WA, US.

Hong, S. Y., Dudhia, J., and Chen, S. H. (2004). A revised approach to ice microphysical processes for the bulk parameterization of clouds and precipitation. *Monthly Weather Review*, 132:103–120.

Houtekamer, P. L., Lefaivre, L., Derome, J., Ritchie, H., and Mitchell, H. L. (1996). A system simulation approach to ensemble prediction. *Monthly Weather Review*, 124:1225–1242.

Houtekamer, P. L. and Mitchell, H. L. (1998). Data assimilation using an ensemble kalman filter technique. *Monthly Weather Review*, 123(3):796–811.

Hunt, B. R., Kostelich, E. J., and Szunyogh, I. (2007). Efficient data assimilation for spatiotemporal chaos: A local ensemble transform kalman filter. *Physical D: Nonlinear Phenomena*, 230:112–126.

Jackson, D. R. (2007). Assimilation of EOS MLS ozone observations in the Met Office data-assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 133(628):1771–1788.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Transactions of the ASME - Journal of Basic Engineering*, 82(D):95–108.

Kalman, R. E. and Bucy, K. (1961). New results in linear prediction filtering theory. *Transactions of the ASME - Journal of Basic Engineering*, 83(D):95–108.

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

Khare, S. P. and Anderson, J. L. (2006). An examination of ensemble filter based adaptive observation methodologies. *Tellus A*, 58(2):179–195.

Khattatov, B. V., Lamarque, J. F., Lyjak, L. V., Menard, R., Levelt, P., Tie, X., Brasseur, G. P., and Gille, J. C. (2000). Assimilation of satellite observations of long-lived chemical species in global chemistry transport models. *Journal of Geophysical Research: Atmospheres*, 105(D23):29135–29144.

Lamarque, J. F., Khattatov, B. V., and Gille, J. C. (2002). Constraining tropospheric ozone column through data assimilation. *Journal of Geophysical Research: Atmospheres*, 107(D22):ACH9–1–ACH9–11.

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

Logan, J. A., Prather, M. J., Wofsy, S. C., and McElroy, M. B. (1981). Tropospheric chemistry: a global perspective. *Journal of Geophysical Research*, 86(C8):7210–7254.

Lynch, P. (2006). *The emergence of numerical weather prediction: Richardson's dream*. Cambridge University Press.

Madronich, S. (1987). Photo dissociation in the atmosphere: 1. Actinic flux and the effects of ground reflections and clouds. *Journal of Geophysical Research*, 92:9740–9752.

Maki, T., Tanaka, T. Y., Sekiyama, T. T., and Mikami, M. (2011). The impact of ground-based observations on the inverse technique of aeolian dust aerosol. *Scientific Online Letters on the Atmosphere*, 7A(Special Edition):21–24.

Menut, L. (2003). Adjoint modeling for atmospheric pollution process sensitivity at regional scale. *Journal of Geophysical Research*, 108:ESQ5–1–ESQ5–17.

Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., and Clough, S. A. (1997). Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *Journal of Geophysical Research*, 102:16663–16682.

National weather service weather prediction center. Daily weather maps (august 17, 2010).

National weather service weather prediction center. Daily weather maps (august 18, 2010).

Oczkowski, M., Szunyogh, I., and Patil, D. (2005). Mechanisms for the development of locally low-dimensional atmospheric dynamics. *Journal of Atmospheric Science*, 62:1135–1156.

Ott, E., Hunt, B. R., Szunyogh, I., Zimin, A. V., Kostelich, E. J., Corazza, M., Kalnay, E., Patil, D. J., and Yorke, J. A. (2004). A local ensemble kalman filter for atmospheric data assimilation. *Tellus A*, 56:415–428.

Pan, L., Gille, J. C., Edwards, D. P., Bailey, P., and Rodgers, C. D. (1998). Retrieval of tropospheric carbon monoxide for the mopitt experiment. *Journal of Geophysical Research*, 103:32277–32290.

Patil, D., Hunt, B. R., Kalnay, E., Yorke, J. A., and Ott, E. (2001). Local low dimensionality of atmospheric dynamics. *Physical Review Letters*, 86:5878–588q.

Pierce, R. B., Schaack, T., Al-Saadi, J. A., Fairlie, T. D., Kittaka, C., Lingenfelser, G., Natarajan, M., Olson, J., Soja, A., Zapotocny, T., Lenzen, A., Stobie, J., Johnson, D., Avery, M. A., Sachse, G. W., Thompson, A., Cohen, R., Dibb, J. E., Crawford, J., Rault, D., Martin, R., Szykman, J., and Fishman, J. (2007). Chemical data assimilation estimates of continental U.S. ozone and nitrogen budgets during the intercontinental chemical transport experiment - North America. *Journal of Geophysical Research: Atmospheres*, 112(D12):D12S21–D12S21.

Rodgers, C. (2000). Inverse methods for atmospheric sounding: Theory and practice. *World Scientific*.

Salzmann, M. (2008). WRF-Chem/KPP Coupler (WKC) for WRF V3, Users' and developers' guide v2.0. Technical report, Princeton University, Princeton, NJ, USA.

Simmons, A. J. and Hollingsworth, A. (2002). Some aspects of the improvement in skill of numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, 128(580):647–677.

Singh, K. and Sandu, A. (2009). Improving GEOS-Chem model forecasts through profile retrievals from tropospheric emission spectrometer model forecasts through profile retrievals from tropospheric emission spectrometer. *Paper presented an International Conference on Computational Science 2009*.

Singh, K., Sandu, A., Bowmanand, K., Parrington, M., Jones, D., and Lee, M. (2011). Ozone data assimilation with GEOS-Chem: A comparison between 3D-Var, 4D-Var, and suboptimal Kalman filter approaches. *Atmospheric Chemistry and Physics*, submitted.

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Duda, M. G., Huang, X. Y., Wang, W., and Powers, J. G. (2008). A description of the advanced research WRF version 3. Technical report, National Center for Atmospheric Research.

Stewart, R. W. (1993). Multiple steady states in atmospheric chemistry. *Journal of Geophysical Research*, 98(D11):20601–20611.

Stockwell, W. R., Middleton, P., Chang, J. S., and Tang, X. (1990). The second generation regional acid deposition model chemical mechanism for regional air quality modeling. *Journal of Geophysical Research*, 95:16343–16376.

Tang, Y., Carmichael, G. R., Horowitz, L. W., Uno, I., Woo, J. H., Streets, D. G., Dabdub, D., Kurata, G., Sandu, A., Allan, J., Atlas, E., Flocke, F., Huey, L. G., Jakoubek, R. O., Millet, D. B., Quinn, P. K., Roberts, J. M., Worsnop, D. R., Goldstein, A., Donnelly, S., Schauffler, S., Stroud, V., Johnson, K., Avery, M. A., Singh, H. B., and Apel, E. C. (2004). Multiscale simulations of tropospheric chemistry in the eastern pacific and on the U.S. west coast during spring 2002. *Journal of Geophysical Research*, 109(D23):D23S11.

Thompson, A. M. (1992). The oxidizing capacity of the earth's atmosphere: Probable past and future changes. *Science*, 256:1157–1165.

Torn, R., Hakim, G. J., and Snyder, C. (2006). Boundary conditions for limited-area ensemble Kalman filters. *Monthly Weather Review*, 133:2490–2502.

Uppala, S. M., KÅllberg, P. W., Simmons, A. J., Andrae, U., Bechtold, V. D. C., Fiorino, M., Gibson, J. K., Haseler, J., Hernandez, A., Kelly, G. A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R. P., Andersson, E., Arpe, K., Balmaseda, M. A., Beljaars, A. C. M., Berg, L. V. D., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B. J.,

Isaksen, L., Janssen, P. A. E. M., Jenne, R., Mcnally, A. P., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N. A., Saunders, R. W., Simon, P., Sterl, A., Trenberth, K. E., Untch, A., Vasiljevic, D., Viterbo, P., and Woollen, J. (2005). The ERA-40 re-analysis. *Quarterly Journal of the Royal Meteorological Society*, 131(612):2961–3012.

Wesley, M. L. (1989). Parameterizations of surface resistances to gaseous dry deposition in regional-scale numerical models. *Atmospheric Environment*, 23:1293–1304.

Whitaker, J. S. and Hamill, T. M. (2002). Ensemble data assimilation without perturbed observations. *Monthly Weather Review*, 130:1913–1924.

Wiedinmyer, C., Akagi, S. K., Yokelson, R. J., Emmons, L. K., Al-Saadi, J. A., Orlando, J., and Soja, A. J. (2011). The fire inventory from NCAR (FINN): a high resolution global model to estimate the emissions from open burning. *Geoscientific Model Development*, 4:625–641.

Wu, L., Mallet, V., Bocquet, M., and Sportisse, B. (2008). A comparison study of data assimilation algorithms for ozone forecasts. *Journal of Geophysical Research: Atmospheres*, 113(D20).

Wursch, M. (2013). *Testing data assimilation methods in idealized models of moist atmospheric convection*. PhD thesis, Meteorologisches Institut of the University of Munich, Munich, Germany.

Yumimoto, K. and Uno, I. (2006). Adjoint inverse modeling of CO emissions over eastern asia using four-dimensional variational data assimilation. *Atmospheric Environment*, 40(35):6836–6845.

Zaveri, R. and Peters, L. K. (1999). A new lumped structure photochemical mechanism for largescale applications. *Journal of Geophysical Research*, 104:30387–30415.

Zhang, F., Snyder, C., and Sun, J. (2004). Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble kalman filter. *Monthly Weather Review*, 132:1238–1253.

Zhang, L. and Sandu, A. (2007). *Data assimilation in multiscale chemical transport models*, volume 4487 of *Lecture Notes in Computer Science*. Springer.