Efficient Computational Tools for Variational Data Assimilation and Information Content Estimation

Kumaresh Singh

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Computer Science

Adrian Sandu, Chair
Wu-chun Feng
Calvin J. Ribbens
Yang Cao
Traian Iliescu
Kevin W. Bowman

August 10, 2010
Blacksburg, Virginia

Keywords: Data Assimilation, Error Covariance Matrices, Information Theory, Chemical Transport Models, Global Ozone Measurements, Model Adjoint Construction, Adjoint Sensitivity Analysis.

Efficient Computational Tools for Variational Data Assimilation and Information Content Estimation

Kumaresh Singh

(ABSTRACT)

The overall goals of this dissertation are to advance the field of chemical data assimilation, and to develop efficient computational tools that allow the atmospheric science community benefit from state of the art assimilation methodologies. Data assimilation is the procedure to combine data from observations with model predictions to obtain a more accurate representation of the state of the atmosphere.

As models become more complex, determining the relationships between pollutants and their sources and sinks becomes computationally more challenging. The construction of an adjoint model (capable of efficiently computing sensitivities of a few model outputs with respect to many input parameters) is a difficult, labor intensive, and error prone task. This work develops adjoint systems for two of the most widely used chemical transport models: Harvard's GEOS-Chem global model and for Environmental Protection Agency's regional CMAQ regional air quality model. Both GEOS-Chem and CMAQ adjoint models are now used by the atmospheric science community to perform real sensitivity analyses and data assimilation studies.

Despite the continuous increase in capabilities, models remain imperfect and models alone cannot provide accurate long term forecasts. Observations of the atmospheric composition are now routinely taken from sondes, ground stations, aircraft, and satellites, etc. This work develops three and four dimensional variational data assimilation capabilities for GEOS-Chem and CMAQ which allow to estimate chemical states that best fit the observed reality.

Most data assimilation systems to date use diagonal approximations of the background covariance matrix which ignore error correlations and may lead to inaccurate analyses. This dissertation develops computationally efficient representations of covariance matrices that allow to capture spatial error correlations in data assimilation.

Not all observations used in data assimilation are of equal importance. This work proposes techniques to estimate the information content of observations used in assimilation; information-theoretic metrics are used.

The four dimensional variational approach to data assimilation provides accurate analyses but requires an adjoint construction, and uses considerable computational resources. This work studies versions of the four dimensional variational methods that use approximate gradients and are less expensive to develop and run.

Variational and Kalman filter approaches are both used in data assimilation, but their relative merits and disadvantages in the context of chemical data assimilation have not

been assessed. This work provides a careful comparison on a chemical assimilation problem with real data sets. The assimilation experiments performed here demonstrate for the first time the benefit of using satellite data to improve estimates of tropospheric ozone.

Dedication

To my parents,

Jai Singh and Shanti Singh.

Acknowledgments

This work has been supported in parts by Houston Advanced Research Council, National Science Foundation and NASA (ROSES-2005 AIST).

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The overall goals of this dissertation are to advance the field of chemical data assimilation, and to develop efficient computational tools that allow the atmospheric science community benefit from state-of-the-art assimilation methodologies.

Data assimilation is the procedure to combine data from observations with model predictions to obtain a more accurate representation of the state of the atmosphere. Data assimilation is an essential ingredient for obtaining meaningful model forecasts. Since models cannot capture exactly the complex dynamics of the atmosphere, models alone provide solutions that become increasingly inaccurate with simulation time. Observations are sparse and observations alone cannot represent the three dimensional distribution of atmospheric constituents.

The work focuses on chemistry transport models, which determine the evolving chemical state of the atmosphere by solving the fundamental equations that govern physical and chemical transformations. Variation in the chemical composition of the atmosphere can affect planet's habitability through short term changes in air quality at ground level, and through long term changes in the climate.

This dissertation develops computationally efficient representations of covariance matrices that allow to capture spatial error correlations in data assimilation (Chapter 4); techniques to estimate the information content of observations used in assimilation (Chapter 5); and variational data assimilation algorithms based on inexact gradients (Chapter 6). Different data assimilation techniques are evaluated in a real data setting (Chapter 3).

The computational tools developed include adjoint systems of Harvard's GEOS-Chem global model and for Environmental Protection Agency's regional CMAQ regional air quality model (Chapter 2). The construction of an adjoint model is a difficult, labor intensive, and error prone task. Both GEOS-Chem and CMAQ adjoint models are now used by the atmospheric science community to perform real sensitivity analysis and data assimilation studies.

1.1 Motivation and Objectives

The computational complexity of a chemistry transport model increases significantly with increase in spatial resolution, number of chemical species, and physical and chemical science processes incorporated into the system; these are the primary factors that decide the accuracy of the model. As models become more complex, determining the relationships between pollutants and their sources and sinks become computationally more challenging. Sensitivity analysis calculates the variations in model outputs due to variations in model input parameters and thus is a key tool in understanding complex interactions among chemical species. *This work develops adjoint sensitivity analysis capabilities for two of the most widely used chemical transport models: CMAQ and GEOS-Chem.*

Despite the continuous increase in capabilities, models remain imperfect and models alone cannot provide accurate long term forecasts. Owing to the increasing interest in climate change and weather forecasting, observations of the atmospheric composition are now routinely taken from a number of instruments including sondes, ground stations, aircraft, and satellites. *This work develops computational tools for data assimilation, which allow to bring together complex models and complex data sets, and to estimate chemical states that best fit the observed reality.*

An important element for data assimilation is the a priori (or background) information. The background error covariance matrix accounts for the background errors magnitude and correlations. The background covariance matrix impacts directly the assimilation results; for example, it determines how the information from a local observation is spread spatially (to neighboring grids) and among chemical species. Most data assimilation systems to date use diagonal approximations of the background covariance matrix for two reasons. First, a full rank background covariance matrix cannot be computed or stored due to the large dimension of the state space; next, the error correlation information (giving the non-diagonal terms) is many times unavailable. Diagonal approximations ignore any correlation and may lead to inaccurate analyses. *This work develops a methodology that allows for an economic representation of non-diagonal background covariance matrices, and for computationally efficient linear algebra operations involving them*.

Not all observations used in data assimilation are of equal importance. Erroneous and redundant observations not only affect the quality of analysis but also add unnecessary computational expense to the assimilation system. It is therefore important to quantify the amount of information brought in by each observation into the assimilation system. This can help prune the observation data set, asses the effectiveness of the assimilation system, develop optimal observation networks, and define strategies for targeting observations. Uncertainty reduction in analysis provides a way of computing the information content [Wahba et al., 1995; Rodgers, 1996; Rabier et al., 2002; Fisher, 2003; Abramov, 2004; Majda, 2006], but the approach is not available for variational data assimilation. *This work develops computationally feasible techniques for estimating the information content of*

observations in the context of four dimensional variational data assimilation.

Data assimilation methods currently in use are based on two approaches: variational and Kalman filter. Variational techniques are rooted in optimal control theory [LeDimet and Talagrand, 1986; Evensen, 1994; Khattatov et al., 2000; Elbern and Schmidt , 2001]. Kalman filter techniques are rooted in statistical estimation theory [Kalman, 1960]; different approximations are required in order to obtain computationally feasible assimilation algorithms for large problems [Anderson and Moore, 1979; Miller et al., 1994; Evensen, 1992b; Constantinescu et al., 2007b,c,d; Menard et al., 2000; Lamarque et al., 2002; Segers et al., 2005; Pierce et al., 2007; Parrington et al., 2009]. The relative merits and disadvantages of the two families in the context of chemical data assimilation have not been extensively explored. *This work provides a careful comparison of several methods applied to a chemical assimilation problem with real data sets.*

Among the variational techniques, the three dimensional version (3D-Var) is easier to implement as it does not involve constructing model adjoints, and is computationally inexpensive to run. The four dimensional variational method (4D-Var) is more accurate but requires an adjoint construction, and more computational time and memory to run. Deriving assimilation techniques that are more accurate than 3D-Var but less expensive than 4D-Var has the potential to provide important practical benefits. *This work studies versions of 4D-Var that use approximate gradients and are less expensive to develop and run.*

1.2 Research Accomplishments

In this work we construct adjoints of two large scale chemical transport models: the global GEOS-Chem model (`http://acmg.seas.harvard.edu/geos`) [Bey et al., 2001] which covers the entire atmosphere, and the regional Community Multiscale Air Quality (CMAQ) model (`http://www.epa.gov/AMD/CMAQ`) [Byun and Ching , 1999; Byun and Schere , 2006]. Adjoints of individual science processes are constructed and validated extensively against their finite difference approximations. The developed adjoint model is hybrid in that a combination of discrete, continuous, and hand coded adjoints are implemented. The performance of the adjoint model is improved through parallelization, and through a balanced combination of checkpoints and recalculations of dependent variables.

Adjoint based sensitivity analysis is an efficient method of computing the changes in few model outputs with respect to changes in large number of model input parameters, making it the most suitable choice to help understand the relationship between chemical species and their sources and sinks for large scale models. We demonstrate the sensitivity of tropospheric ozone with respect to its trace gas precursors, emissions, and dry deposition for a single GEOS-Chem model adjoint run covering three simulation days. A similar sensitivity study of tropospheric ozone with respect to initial nitrogen dioxide

concentrations on a regional scale is conducted with CMAQ.

A data assimilation framework including 3D-Var and 4D-var algorithms is implemented for Geos-4 v7 of GEOS-Chem. The framework is capable of assimilating vertical profile retrievals from the Tropospheric Emission Spectrometer (TES) [Beer et al., 2001], the first dedicated infrared instrument which provides information about the global and vertical distribution of tropospheric ozone. We provide the first results of improved global tropospheric ozone estimates by assimilating ozone profile retrievals from TES into the GEOS-Chem through variational approaches. We also provide the first side by side comparison of 3D-Var, 4D-Var and suboptimal Kalman filter (KF) [Parrington et al., 2009] data assimilation systems in terms of computational cost and the quality of assimilations. The accuracy of the estimated ozone concentration is verified against independent data from ozonesonde measurements. All the three assimilation systems lead to improved ozone distributions: the relative difference between mean ozone estimates and ozonesondes is reduced by up to 75% for sequential methods (3D-Var, suboptimal KF) and by up to 90% for 4D-Var over a two week assimilation. This work enables GEOS-Chem users worldwide to choose an assimilation system based on their requirements of accuracy and availability of computational resources.

A realistic estimation of the background error distribution is of considerable importance in data assimilation. We propose a computationally efficient approach of constructing full rank background error covariance matrices that account for correlations in both horizontal and vertical directions. This multi-dimensional correlation is achieved by exploiting the tensor products of one-dimensional correlation matrices. Thus, the explicit construction and storage of full covariance matrices is avoided. Both the Kalman filter algorithms (which involve matrix-vector multiplications) and the variational data assimilation algorithms (which require the solution of liner systems and the square root of the covariance matrix) benefit directly from the derived efficient linear algebra operations. We illustrate the impact of non-diagonal background error covariance matrices on data assimilation systems by assimilating ozone profile retrievals from TES into GEOS-Chem using the 3D-Var and 4D-Var algorithms. Both assimilation systems lead to significant improvements of the global ozone estimates. We propose a method of estimating the local spatial correlation length though computational experimentation.

In order to prune erroneous and redundant observations from the available data set, one needs to quantify the information brought in by each observation point into the data assimilation system. We propose a cost efficient ensemble-based approach to estimate information theoretic metrics in the context of 4D-Var. We consider the signal information to measure the adjustment of the mean of the distribution, and the Fisher information matrix, the Shannon information, and the degrees of freedom for signal to measure the decrease in the variance of the error. We first demonstrate a (vertical) level-wise information gain through the assimilation of TES ozone profile retrievals during the 00:00 GMT August 1, 2006 - 00:00 GMT August 6, 2006 period, into GEOS-Chem. We next present the amount of information brought in by each observation window of

4 hours, using signal and degrees of freedom for signal information. This cumulative information is then apportioned to individual data points. We perform 4D-Var data assimilation using data subsets selected based on the signal information content. Results show that the quality of analysis generated using the top 2.3% observation points is similar to that using the bottom 27%, while the quality of analysis generated using top 27% is comparable to that of using all the observations. We discuss the contributions of individual data points using degrees of freedom for signal. The results indicate a difference in the quality of the analyses generated using subsets of observations, however, the most effective data points seem to be the ones with lower degrees of freedom for signal. Further research work is necessary here. We provide a methodology to asses the potential impact of virtual observations. This is useful for planning new field campaigns, and for guiding the design of optimal observing networks.

Data assimilation systems based on 3D-Var and 4D-Var approaches have been implemented for numerous models in the fields of atmospheric sciences, meteorology and oceanography. The 3D-Var is easier to implement and is computationally less expensive, while 4D-Var usually provides better estimates but requires higher development and run costs. No effort have been made to date in developing an assimilation system that is cost efficient, yet does not compromise much on the quality of analysis. We propose assimilation techniques that use inexact gradient information and call them Quasi 4D-Var (Q4D-Var). One approach involves accumulating the adjoint forcing calculations of a 3D-Var system for each observation window and carry out optimization at the start of the assimilation window. This is conceptually equivalent to conducting a 4D-Var assimilation with model adjoint treated as an identity operator. Another approach performs a 4D-Var assimilation using only the continuous adjoint of advection process. As described in Chapter 2, the continuous adjoint of advection is easily implemented by reversing the direction of the wind fields and does not require checkpointing of forward variables. The Q4D-Var approaches are slightly more expensive than 3D-Var, but provide more accurate and smoother analyses. There are definitely limitations to our approach as the proposed methodologies would work as long as the difference between the model adjoint and the inexact gradients is not large. Hence, the quality of analysis is expected to deteriorate with increase of the assimilation window length. For relatively short assimilation windows the results indicate that the Q4D-Var analyses are comparable with those obtained by strongly-constrained 4D-Var.

1.3 Dissertation Layout

This dissertation is organized as follows. Chapter 2 presents a mathematical overview of adjoint modeling for chemistry transport models. Details on the construction and validation of adjoints of GEOS-Chem and CMAQ models are provided. Adjoint sensitivity analysis studies using both models are presented. Chapter 3 provides a detailed

discussion on how observations of reality are integrated with the model using various data assimilation systems. A comparison of these approaches on a tropospheric ozone estimation problem is presented. The methodology for constructing multidimensional covariance matrices and for performing linear algebra operations is discussed in Chapter 4. The non-diagonal covariance matrices on the quality of analysis is assessed. Chapter 5 develops computationally feasible estimation techniques to quantify the information content of observations in the context of 4D-Var. The effectiveness of the proposed approach is demonstrated on a simulation with real data. New data assimilation algorithms based on inexact gradients are proposed in Chapter 6. Their cost effectiveness and accuracy is illustrated. Chapter 7 summarizes the work of this dissertation and provides future research directions.

Chapter 2

Construction of Adjoints of Global (GEOS-Chem) and Regional (CMAQ) Air Quality Models

Abstract

Adjoint models are powerful tools for estimating the sensitivity of (a function of) a model output with respect to a large number of input parameters. Adjoints are extensively used in receptor-oriented sensitivity analysis studies and for providing gradients in the solution of inverse problems. The construction of an adjoint model is a difficult, labor intensive, and error prone task. This chapter discusses the construction of adjoints for Harvard's GEOS-Chem global model and for Environmental Protection Agency's regional CMAQ regional air quality model. Such models describe the evolution of the chemical composition of the atmosphere due to natural and anthropogenic factors.

Chemical transport models account for a variety of interacting physical processes like emissions, deposition, transport, gas and liquid phase chemistry, particulate processes, radiation effects, etc. Adjoints for individual science processes are developed using a combination of multiple approaches: symbolic preprocessing, analytic derivation of adjoint equations, hand coding, and automatic differentiation. The developed adjoints are validated extensively against finite-differences on an individual process basis as well as for the full model. The adjoint model performance is improved through parallelization, and through a balanced combination of checkpoints and recalculations.

The adjoint framework developed here is already in use by the GEOS-Chem and the CMAQ communities to perform sensitivity analysis and data assimilation for improved initial tracer concentrations, boundary conditions, and emission sources.

2.1 Introduction

Chemical constituents of the atmosphere play an important role in determining its temperature, radiation and dynamics, consequently affecting the habitability of the planet through changes in climate and air quality at the ground. Several trace gases are listed as pollutants; some because of their heat entrapment capabilities, highly oxidative properties, and for escalating breathing and sight problems, while others because they act as precursors or as catalyst in the formation of hazardous trace gases through chemical reactions. There are several regional and global atmospheric and weather forecasting agencies that provide information on the climate and air quality through chemical transport and global circulation models, while regulatory agencies create policies and strategic plans to control human activities that cause an increase in the pollutant levels. In order to compose such restraining policies, regulatory agencies require precise quantification of each source and sink of the pollutants in a cost effective manner.

In an attempt to replicate the complex physical and chemical processes of the atmosphere, the chemical transport models currently in use include several chemical species interacting with each other over millions of grid boxes. Sensitivity analysis computes the changes in the model outputs with respect to small changes in model parameters. It can help quantify the relationship between chemical tracers and their precursors. The gradient information provided by sensitivity calculations is useful in various applications including parameter estimation, design of optimal control strategies, and data assimilation.

Sensitivity calculations can be performed either in forward mode or in backward mode. The conventional direct sensitivity analysis propagates perturbations in model parameters forward in time along model trajectories from input sources to various model outputs (receptors). The Direct Decoupled Method (DDM) [Hakami et al. , 2003; Yang et al. , 1998, 2000] uses this technique to provide sensitivities of all the model outputs with respect to a few parameters. However, this technique is infeasible for a system with large number of input parameters. In the adjoint sensitivity analysis, the changes seen in the model outputs are propagated backwards in time along the model adjoint trajectory to the model parameters. Adjoint sensitivity analysis can efficiently calculate the derivatives of a cost functional with respect to a large number of input parameters.

Adjoint sensitivity analysis was first applied in the field of atmospheric sciences by [Marchuk , 1974; Lamb et al. , 1975], while the mathematical overview of sensitivity analysis for nonlinear models was discussed in [Cacuci , 1981a,b]. The method was later applied extensively in meteorology and oceanography for sensitivity analysis [Hall et al. , 1982; Hall and Cacuci , 1983; Errico and Vukicevic , 1992] and variational data assimilation [LeDimet and Talagrand, 1986; Talagrand and Courtier , 1987]. Due to the increase in computational resources and wide applicability, adjoint sensitivity analysis received huge attention in the past decade. Beginning with Lagrangian models [Fisher

and Lary , 1995; Elbern et al. , 1997], the method was later augmented to more detailed three dimensional chemistry transport models [Elbern et al. , 1997; Elbern and Schmidt , 1999; Vukicevic and Hess , 2000; Elbern et al. , 2000; Elbern and Schmidt , 2001]. More recently adjoint based applications have been developed for regional and global CTMs including CHIMERE [Vautard et al. , 2000; Menut et al. , 2000; Schmidt and Martin , 2003], Polair [Mallet and Sportisse , 2004, 2006], IMAGES [Muller and Stavrakou , 2005; Stavrakou and Muller , 2006], CIT urban-scale model [Martien et al. , 2006; Martien and Harley , 2006], TM4 [Meirink et al. , 2006], DRAIS [Nester and Panitz , 2006] and STEM [Sandu et al. , 2005a; Hakami et al. , 2005, 2006; Chai et al. , 2006].

In this chapter we focus on two chemical transport models: GEOS-Chem is a global CTM (`http://acmg.seas.harvard.edu/geos/`) [Bey et al., 2001] covering the entire atmosphere, while Community Multiscale Air Quality (CMAQ) (`http://www.epa.gov/AMD/CMAQ/`) [Byun and Ching , 1999; Byun and Schere , 2006] is a regional CTM.

Original work on the adjoint of GEOS-Chem began in 2003, focusing on the adjoint of offline aerosol simulations [Henze et al. , 2007]. By 2005, the adjoint was expanded to include a tagged CO simulation and a full chemistry simulation as well as observational operators for MOPITT (CO) and IMPROVE networks (aerosols) [Kopacz et al. , 2009]. [Henze et al. , 2007] discusses the construction of adjoint for GEOS-3 v6 of GEOS-Chem. GEOS-4 v7 of GEOS-Chem is significantly different from its previous version in that it includes hybrid sigma grid as compared to pure sigma grid for advection equation discretization in the vertical direction; pure sigma grids follow the surface terrain structure and lead to noisy winds in the upper atmosphere, while, hybrid grids use sigma grids near surface and fixed pressure grids for mid and upper atmosphere. The underlying advection mechanism in both the versions is multidimensional flux form semi-Lagrangian transport model [Lin and Rood , 1996]. For convection, Geos-4 has separate treatments of deep and shallow convections following the schemes developed by [Zhang and McFarlane , 1995] and [Hack , 1994] compared to the Relaxed Arakawa-Schubert convection scheme [Moorthi and Suarez , 1992] used in Geos-3. In chemistry, several more chemical species and corresponding reaction equations were introduced, and new emission inventories were added to v7. This makes GEOS-4 v7 a stand alone chemistry transport model, the adjoint model of which was required not only for sensitivity studies, but also for data assimilation to generate optimal initial conditions, parameter estimation and policy making on intercontinental pollution transport. Here we present a detailed discussion on construction and validation of an adjoint model for GEOS-4 v7 of GEOS-Chem providing a framework for adjoint sensitivity analysis. Current development also aims to standardize the implementation of the adjoint model and make it more user friendly, with the wider goal of making this adjoint publicly available as part of the standard GEOS-Chem code. This provides the community of GEOS-Chem users not only with the tools needed for performing sensitivity analysis and data assimilation, but also enable them to use more recent observation data.

Environmental Protection Agency's CMAQ model is the most widely used regional air

quality model worldwide. The need for adjoint sensitivity increases with resolution as the method could be used to locate precisely the sources and quantify their rates of emissions to eventually impose regulations. Also, the concentration of pollutants at regional scale determine heavily a community's air quality and weather conditions. It becomes important to study the physical and chemical interactions that lead to a particular state of a region. Discussed previously in [Hakami et al. , 2007], CMAQ v4.5 adjoint was built in collaboration between California Institute of Technology and Virginia Tech to provide users worldwide with tools to carry out sensitivity with respect to tracers and boundary conditions and 4D-Var data assimilation. Since most of the internal science processes and underlying governing equations in CMAQ are similar to GEOS-Chem, we will only present the validation of developed adjoint and sensitivity results here.

We employ a variety of strategies for the construction of adjoint models: symbolic preprocessing, derivation of adjoint differential equations and their numerical solution, and automatic differentiation. Kinetic PreProcessor (KPP) [Damian et al. , 2002] chemistry was first interfaced with GEOS-Chem and its adjoint in [Henze et al. , 2007], see Appendices therein. Here we improve upon this implementation in terms of automation, performance, benchmarking, and documentation. KPP has a provision to generate chemical solvers that are capable of performing discrete adjoint calculations [Daescu et al. , 2000, 2003; Sandu et al. , 2003]. These adjoint subroutines were plugged in automatically through the developed parsers in a collaborative effort [Eller et al. , 2009]. Convection and wet-deposition adjoints are discrete adjoints and have been constructed in a hybrid fashion using the automatic differentiation software TAMC [Giering and Kaminski , 1998] together with manual coding. Advection adjoint on the other hand is continuous and is obtained by calling the forward subroutine with reverse wind fields. We also provide a way of calculating the scaled emission and dry-deposition adjoints by modifying the KPP chemistry integrator. All these pieces are tested extensively and integrated together to build the full adjoint GEOS-Chem model.

Some discrete adjoint calculations require intermediate variable values from forward calculations. In such cases, either these variables are recomputed in the adjoint mode or are written to checkpoint files [Griewank and Walther , 2000] during the forward calculation and read in the adjoint mode. We recalculated variable values in the adjoint mode wherever this had a minimal performance penalty. The standard version of GEOS-Chem adjoint is completely parallel. The continuous advection adjoint inherits the parallelism of the forward subroutine. For rest of the science process adjoints, we have implemented OpenMP parallel versions.

This chapter is organized as follows. Section 2.2 presents the governing equations for chemistry transport models and introduces the concepts of tangent linear, continuous adjoint, and discrete adjoint models for sensitivity analysis. The general methodology applied for the construction and validation of model adjoints is introduced in Section 2.3. Details on the construction of GEOS-Chem science process adjoints are given in Section

2.4 (chemistry), Section 2.5 (advection), Section 2.6 (convection), Section 2.7 (turbulent mixing), Section 2.8 (emission and dry deposition), Section 2.9 (wet deposition), and Section 2.10 (stratosphere-troposphere ozone exchange). The validation of adjoints and performance metrics are presented in Section 2.11 for GEOS-Chem and in Section 2.12 for CMAQ. Examples of adjoint sensitivity analysis studies using both models and real data are given in Section 2.13. Section 2.14 draws conclusions and points to future work.

2.2 Mathematical Overview of Adjoint Modeling

In this section we present the underlying governing equations used in chemical transport models such as GEOS-Chem and CMAQ. The discretization in time via operator splitting is presented. The space discretization does not impact the main thread and its presentation is omitted. We discuss the tangent linear model of the chemical transport equations. Continuous and discrete adjoint models are derived in this context.

2.2.1 Chemical Transport Modeling

Starting with a certain state, chemical transport models generate a concentration field as a prediction of a future state of the atmosphere. This evolution of concentration in time can be described by the following set of equations defined over a domain Ω

$$\frac{\partial c_i}{\partial t} = -u \cdot \nabla c_i + \frac{1}{\rho}\nabla \cdot (\rho K \nabla c_i) + \frac{1}{\rho} f_i(\rho c) + E_i - D_i, \quad t^0 \leq t \leq T \tag{2.1a}$$

$$c_i(t^0, x) = c_i^0(x) \tag{2.1b}$$

In the above, c_i is the mole-fraction concentration of chemical species i ($1 \leq$ i $\leq$ s), u is the wind field vector, K the turbulent diffusivity tensor, ρ the air density in $moles/cm^3$, E_i and D_i are the rates of emissions and dry depositions for species i, and f_i is the rate of chemical transformations that depends on the absolute concentration values.

The system (2.1a)-(2.1b) is referred as forward model and c^0 is considered as the initial state of the model. It is important to note that we have not presented the boundary conditions here. This is applicable to the GEOS-Chem global simulations which has circular boundary, however, if we try to conduct a nested grid simulation or regional air quality simulations using CMAQ, the boundary conditions need to be accounted for. It is also worth mentioning that in GEOS-Chem, the emissions and dry depositions are included in the chemistry simulations through fake chemical reactions and therefore are represented as constant rate equations in (2.1a). Notice the positive sign for emission which indicates addition in the concentration of certain species while deposition leads to a decrease for certain species in the system.

If we think of the forward model as a numerical operator $\mathcal{M}$ that transforms a given chemical state as per the governing equation, then for a sequence of N timesteps of length Δt taken between t^0 and $t^N = T$

$$c^{n+1} = \mathcal{M}_{t^n \to t^{n+1}} \circ c^n, \quad c^N = \prod_{n=0}^{N-1} \mathcal{M}_{t^n \to t^{n+1}} \circ c^0 \tag{2.2}$$

where, at each time step $t^n = t^0 + n\Delta t$, one calculates the approximation $c^n(x) \approx c(t^n, x)$.

The numerical operator $\mathcal{M}$ can be further split into operators that imitate individual science processes that comprise of the governing equation. This technique is referred to as operator splitting and is widely used in air quality modeling [Lanser and Verwer , 1998]. Due to the fact that chemical reactions in nature have slower dynamics than transport which depend on the fast varying wind and temperature, the chemistry operator is applied only once during a single time step, while all other operators are applied multiple times. The step size decreases in proportion to the number of times the operator is applied. In GEOS-Chem, chemistry calculation is performed every hour while transport and convection are performed every 30 minutes at $4^\circ \times 5^\circ$ resolution and every 15 minutes at $2^\circ \times 2.5^\circ$ resolution. If we denote by O_{ad}, O_{cv}, and O_{ch} the numerical operators representing advection, convection and chemistry processes, then for a $4^\circ \times 5^\circ$ simulation, we have

$$\mathcal{M}_{t^n \to t^n + \Delta t} = O_{ch}^{\Delta t} \circ O_{cv}^{\Delta t/2} \circ O_{ad}^{\Delta t/2} \circ O_{cv}^{\Delta t/2} \circ O_{ad}^{\Delta t/2} \tag{2.3}$$

This technique not only provides solutions for intermediate steps enabling simulations to be carried out for various time interval but also extends the option of using selective processes as per the requirements of an experiment. The added benefit is that it assists in the development of tangent linear and adjoint models that we will be discussing shortly. On the downside, this method does introduce numerical errors into the system, however, these splitting errors remain within bounds for evolving time [Lanser and Verwer , 1998].

2.2.2 Tangent Linear Model and its Adjoint

The forward model produces a forecast field that is unique for a particular initial condition and other input parameters such as emission and dry deposition rates. A small variation in any of these parameters would cause the resultant to be different. Let us consider a response functional $\mathcal{J}$ that accounts for the changes in the model solution as per the variation in the input parameters

$$\mathcal{J}(c^0) = \int_{t^0}^{T} \int_{\Omega} g(c(t,x)) dx dt \tag{2.4}$$

An infinitesimal change δc^0 in the initial condition will bring a variation in the response functional approximated as

$$\delta \mathcal{J}(c^0) = \int_{t^0}^{T} \left(\int_{\Omega} \delta c(t,x) \cdot \frac{\partial g}{\partial c}(c(t,x)) dx \right) dt \tag{2.5}$$

and perturbations $\delta c(t,x)$ in each predicted concentration c_i through the following set of equations

$$\frac{\partial \delta c_i}{\partial t} = -u \cdot \nabla \delta c_i + \frac{1}{\rho} \nabla \cdot (\rho K \nabla \delta c_i) + F_i(\rho c) \delta c, \quad t^0 \leq t \leq T \tag{2.6a}$$

$$\delta c_i(t^0, x) = \delta c_i^0(x) \tag{2.6b}$$

where, F is the Jacobian of the function f. The system (2.6a)-(2.6b) is referred to as the tangent linear model. The direct sensitivity analysis approach discussed in Section 2.1 requires the model (2.1a)-(2.1b) together with tangent linear model to be solved forward in time. If we denote by $\mathcal{L}$ the tangent linear model operator, then equation (2.6a) reduces to

$$\frac{\partial \delta c}{\partial t} = \mathcal{L}(c(t)) \delta c, \quad t^0 \leq t \leq T \tag{2.7}$$

The above discussion leads to an interesting problem where through small perturbations in the initial condition, one could reach a state such that the amount of variation in the response functional is minimal. In more precise terms, the search for such a state transforms into the following optimization problem

$$\begin{aligned} &\textit{minimize} \quad \mathcal{J}(c^0) = \int_{t^0}^{T} \int_{\Omega} g(c(t,x)) dx dt \\ &\textit{subject to} \quad \frac{\partial c}{\partial t} = -u \cdot \nabla c + \frac{1}{\rho} \nabla \cdot (\rho K \nabla c) + \frac{1}{\rho} f(\rho c) + E - D, \quad t^0 \leq t \leq T \end{aligned} \tag{2.8}$$

Using a new variable λ, the Lagrange multiplier, let us define the Lagrange function as

$$\begin{aligned} \Lambda(c^0, \lambda) = \int_{t^0}^{T} dt \int_{\Omega} g(c(t,x)) - \Bigg(&\frac{\partial c}{\partial t} + u \cdot \nabla c \\ &- \frac{1}{\rho} \nabla \cdot (\rho K \nabla c) - \frac{1}{\rho} f(\rho c) - E + D \Bigg) \cdot \lambda dx \end{aligned} \tag{2.9}$$

If c^0 is an optimal point for the original constrained problem (2.8), then there exists a λ such that (c^0, λ) is a stationary point for the Lagrange function (2.9). At this stationary point,

$$\delta \Lambda(c^0, \lambda) = 0 \tag{2.10}$$

Using equations (2.1a),(2.5),(2.6a),(2.9), and (2.10), we obtain

$$\int_{t^0}^{T} dt \int_{\Omega} \left(\delta c \cdot \frac{\partial g}{\partial c}(c(t,x)) + \mathcal{L}(c)\delta c \cdot \lambda \right) dx = \int_{t^0}^{T} dt \int_{\Omega} \frac{\partial \delta c}{\partial t} \cdot \lambda \tag{2.11}$$

If we now define $\mathcal{L}^*$, an adjoint of the tangent linear model operator, then through the Lagrange identity, we have

$$\int_{t^0}^{T} dt \int_{\Omega} \mathcal{L}(c)\delta c \cdot \lambda dx = \int_{t^0}^{T} dt \int_{\Omega} \delta c \cdot \mathcal{L}^*(c)\lambda dx \tag{2.12}$$

After integrating by parts the right hand side and (2.12) it follows that

$$\int_{t^0}^{T} dt \int_{\Omega} \delta c \cdot \left(\frac{\partial g}{\partial c}(c(t,x)) + \mathcal{L}^*(c)\lambda \right) dx = \int_{\Omega} \delta c(T) \cdot \lambda(T) - \int_{\Omega} \delta c(t^0) \cdot \lambda(t^0) \tag{2.13}$$

If $\lambda(t,x)$ is defined as the solution of the adjoint system

$$\begin{aligned} \frac{\partial \lambda}{\partial t} &= -\mathcal{L}^*(c)\lambda - \frac{\partial g}{\partial c}(c(t,x)) \\ \lambda(T) &= 0 \end{aligned} \tag{2.14}$$

then

$$\delta \mathcal{J}(c^0) = \int_{\Omega} \delta c^0 \cdot \lambda(t^0) dx \tag{2.15}$$

where, $\lambda(t^0,x)$ represents the sensitivity to the initial conditions. The adjoint operator $\mathcal{L}^*(c)$ in equation (2.14) is given by

$$\mathcal{L}^*(c)\lambda = \nabla \cdot (u\lambda) + \nabla \cdot \left(\rho K \nabla \frac{\lambda}{\rho} \right) + F^{\mathrm{T}}(\rho c)\lambda \tag{2.16}$$

Continuous Adjoint Sensitivity

The continuous adjoint model is defined as the adjoint of the tangent linear model (associated with the original system of partial differential equations). Therefore, from the previous discussion, the continuous adjoint variables evolve in time as per the following set of equations

$$\frac{\partial \lambda_i}{\partial t} = -\nabla \cdot (u\lambda_i) - \nabla \cdot \left(\rho K \nabla \frac{\lambda_i}{\rho} \right) - \left(F^{\mathrm{T}}(\rho c)\lambda \right)_i - \varphi_i, \quad T \geq t \geq t^0 \tag{2.17a}$$

$$\lambda_i(T,x) = \lambda_i^N(x) \tag{2.17b}$$

$$\varphi_i(t,x) = \frac{\partial g(c_1, c_2, \ldots, c_n)}{\partial c_i}(t,x), \quad \lambda_i^N(x) = 0 \tag{2.17c}$$

Here each adjoint variable $\lambda_i(t,x)$ is associated with its corresponding concentration $c_i(t,x)$ for all the species $1 \leq i \leq s$. These adjoint variables also known as influence functions represent the sensitivities of the response functional with respect to the variations in the state variable

$$\lambda_i(t,x) = \frac{\partial \mathcal{J}}{\partial c_i(t,x)} \tag{2.18}$$

Similar to forward model, the adjoint model (2.17a)-(2.17c) is again a chemical transport equation and can be solved by appropriate numerical methods of choice. In adjoint sensitivity analysis, the adjoint variable is initialized at the final time T and is integrated backward in time down to t^0 through the adjoint model equations. Notice that the adjoint model equations also depend on the concentration fields from the forward run, and thus would require their values at each time step in the backward integration. Therefore, similar to direct sensitivity analysis, the adjoint sensitivity analysis also requires the forward model solutions. The only difference being the requirement that the forward model be solved first to save the concentration state vector $c(t,x)$ at each time step and then reuse these values in the backward integration. If we denote by A, the adjoint model operator, being applied at each time step in the backward direction, then we have

$$\lambda^n = \mathcal{A}_{t^{n+1}\rightarrow t^n} \circ \lambda^{n+1}, \quad \lambda^0 = \prod_{n=N-1}^{0} \mathcal{A}_{t^{n+1}\rightarrow t^n} \circ \lambda^N \tag{2.19}$$

The continuous adjoint sensitivity is also know as discretization of the adjoint (DA) approach. The response functionals are defined based on the problem at hand, and the forcing φ and adjoint variables at time t^N are chosen such that the adjoint variables are the sensitivities of the response functional defined.

Discrete Adjoint Sensitivity

Also known as adjoint of the discretization (AD), in the discrete adjoint sensitivity approach we start with the numerical discretization (2.2) of the forward model (2.1a)-(2.1b). Consider a computational grid that covers the domain Ω and on which the solution is discretized. Let $m = (j,k,l)$ be a grid point with coordinates j (longitude), k (latitude), and l (vertical level). A response function can be defined in terms of the discrete solution

$$\widehat{\mathcal{J}}(c^0) = \sum_{n=0}^{N} \sum_{m} g(c^n[m]) \tag{2.20}$$

Here $c^n[m]$ is the concentration vector of all species at time n and grid point m.

The aim is to derive the gradients of the discrete response functional (2.20) with respect to the discrete initial condition $c^0[m]$. A small change δc^0 in the initial condition evolves

in time governed by the following discrete tangent linear operation

$$\delta c^{n+1} = \mathcal{M}'_{t^n \to t^{n+1}} \circ \delta c^n, \quad \delta c^N = \prod_{n=0}^{N-1} \mathcal{M}'_{t^n \to t^{n+1}} \circ \delta c^0 \tag{2.21}$$

Here, $\mathcal{M}'$ is the tangent linear operator associated with the forward model numerical operator $\mathcal{M}$ (2.2). Splitting the operator $\mathcal{M}'$ into corresponding individual science process operators leads to

$$\mathcal{M}'_{t^n \to t^n + \Delta t} = O'^{\Delta t}_{ch} \circ O'^{\Delta t/2}_{cv} \circ O'^{\Delta t/2}_{ad} \circ O'^{\Delta t/2}_{cv} \circ O'^{\Delta t/2}_{ad} \tag{2.22}$$

As per the following duality principle, to each linear operator L there exists an adjoint operator L^*

$$\langle u, Lv \rangle_n = \langle L^* u, v \rangle_n \tag{2.23}$$

where, $\langle \cdot, \cdot \rangle_n$ denotes the inner product in $\mathbb{R}^n$. Applying equation (2.23) to the tangent linear operation (2.22) provides the discrete adjoint operator

$$\mathcal{M}'^*_{t^n + \Delta t \to t^n} = O'^{*\Delta t/2}_{ad} \circ O'^{*\Delta t/2}_{cv} \circ O'^{*\Delta t/2}_{ad} \circ O'^{*\Delta t/2}_{cv} \circ O'^{*\Delta t}_{ch} \tag{2.24}$$

such that the resulting discrete adjoint model is

$$\lambda^n = \mathcal{M}'^*_{t^{n+1} \to t^n} \circ \lambda^{n+1} + \varphi^n, \quad n = N-1, N-2, \ldots, 0\,. \tag{2.25}$$

The response functionals are defined based on the problem at hand, and the forcing φ and adjoint variables at time t^N are chosen such that the adjoint variables are the sensitivities of the response functional defined with respect to the state variables $c^n[m]$

$$\lambda_i^n[m] = \frac{\partial \widehat{\mathcal{J}}(c^0)}{\partial c_i^n[m]}\,. \tag{2.26}$$

From this point on we will drop the hat notation that distinguishes the discrete cost function from the continuous one.

2.3 Construction and Validation of Model Adjoints

This section describes the general procedure we used for the construction and validation of the adjoint of GEOS-4 v7 of GEOS-Chem. Since most of the underlying science processes in CMAQ v4.5 are present in the GEOS-Chem model as well, the adjoint construction follows similar steps. We do not provide the (redundant) details of adjoint construction for CMAQ. Validation results will be provided for individual processes as well as for the full model.

As seen in the previous section, operator splitting technique allows the development of tangent linear and adjoint models for an atmospheric chemical transport model with much relative ease. We will now utilize equations (2.19) and (2.24), to derive the adjoint of each science process individually. Discrete adjoints in principle have an advantage over continuous adjoints in the sense that the earlier provides an exact derivative of the discrete response functional being minimized. However, as discussed in [Sirkes and Tziperman , 1997], discrete adjoints show strong oscillatory numerical artifacts if the forward function under consideration has discontinuities or is non-smooth. A small example showcasing this artifact is presented for CMAQ advection adjoint derived through discrete approach using TAMC. CMAQ advection process is based on full monotonicity constrained Piecewise Parabolic Method (PPM). The smoother applied at the boundaries for monotonicity in the forward equation is not translated well in the discrete adjoint code leading to formation of wiggles at the boundaries which grow unboundedly over time. Hence, we utilize the hybrid adjoint modeling approach in the sense that discrete adjoints are implemented for all the science processes with the exception of advection, in which case the continuous adjoint is fairly easy to construct and performs better than its discrete equivalent.

The continuous adjoint is derived manually, while automatic differentiation tools such as Tangent and Adjoint Model Compiler (TAMC) are used to generate the discrete adjoint codes. Due to multiple dependencies between functions and variables involved in large model codes, there is a lot of effort involved in creating stand alone versions of the science processes that need to be submitted to the automatic differentiation tool. In addition, based on the number of forward variables on which the adjoint equation depends upon, the automatic differentiation tools produce large chunks of codes and additional loops to recalculate these forward values. A two-level checkpointing [Griewank and Walther , 2000; Sandu et al. , 2005a] is implemented for this purpose. With computational efficiency in mind, we recalculate the intermediate variable values wherever it is fairly straightforward, while for more involving codes, we write out checkpoint files with intermediate values in the forward mode and use those to avoid longer times spent on recalculations during the backward integration.

An important aspect of adjoint based sensitivity studies is the correctness of the adjoint values. A mismatch between the true gradient of the response functional and the calculated adjoint values might lead to a divergence in optimization during data assimilation or produce inexact analysis fields if longer assimilation windows are used; an elaborate introduction and discussion on data assimilation is provided in Chapter 3 while discussion on usage of inexact gradients is provided in Chapter 6. In order to validate the developed adjoint model, we create a finite difference test bed. Recall from calculus that

the derivative of a function $F(x)$ can be approximated through

$$\textit{forward difference:} \quad F^{'}(x) \approx \frac{F(x+h) - F(x)}{h} + O(h) \tag{2.27}$$

$$\textit{central difference:} \quad F^{'}(x) \approx \frac{F(x+h) - F(x-h)}{2h} + O(h^2) \tag{2.28}$$

These approximations are derived through the Taylor series expansion of F about the value of x, where the errors indicated by O notation increase with the amount of perturbation in x. Considering an infinitesimal perturbation in the initial state vector c^0, the tangent linear model operator $\mathcal{M}^{'}$ (2.21) at time step t^n could be approximated (forward) as

$$\mathcal{M}^{'}_{t^n \rightarrow t^{n+1}} \cdot \delta c^0 \approx \mathcal{M}_{t^n \rightarrow t^{n+1}} \circ c(n, c^0 + \delta c^0) - \mathcal{M}_{t^n \rightarrow t^{n+1}} \circ c(n, c^0) \; . \tag{2.29}$$

Using the above approximation, it is evident that the validation of direct sensitivity analysis is easy to achieve and can be performed at any time step since both tangent linear and forward model are integrated forward in time. Validation of adjoint sensitivities is more involved and for each pair of forward and adjoint model simulations, there is only one point-to-point comparison available. In practice, the adjoint sensitivities are compared with the finite difference approximations at the initial time t^0 that would reflect discrepancies if any during any time step in the integration window. Applying forward finite difference approximation to the equation (2.26) provides

$$\lambda_i^0[m] = \frac{\mathcal{J}(c^0 + \delta c_i^0 \, \mathbb{e}_i[m]) - \mathcal{J}(c^0)}{\delta c_i^0[m]}, \quad \forall \; i, m \; . \tag{2.30}$$

Here $\mathbb{e}_i[m]$ is a vector of zeros with a one for the species i and grid cell m (the perturbation is applied to a single species i in a single grid cell m). The model is run forward with the perturbed initial conditions to obtain the perturbed value of the cost function. This approach requires one forward model run per each entry in the adjoint vector.

For example, choosing $\mathcal{J}$ (2.20) to be a linear response functional as

$$\mathcal{J}(c^0) = \sum_m c_\ell^N[m] \tag{2.31}$$

and using equation (2.32) leads to

$$\lambda_i^0[m] = \frac{\prod_{n=0}^{N-1} \mathcal{M}_{t^n \rightarrow t^{n+1}} \circ (c^0 + \delta c_i^0 \, \mathbb{e}_i[m]) - \prod_{n=0}^{N-1} \mathcal{M}_{t^n \rightarrow t^{n+1}} \circ (c^0)}{\delta c^0[m]}, \; \forall i, m \; . \tag{2.32}$$

The central finite difference approximation can be deduced in a similar fashion. For chemistry only calculations the validation at each grid point m is independent of other grid points, and the validation of the three dimensional adjoint can be thought of as an ensemble of numerous box models providing a point-to-point comparison.

In order to validate the adjoint of an individual science process, all other processes are turned off. The forward and adjoint model equation then consist only of the science process being tested. The testing criteria for each process is different and is designed based on how the concentration state vector evolves in time governed by the forward model. For validation, we used central finite difference approximations for all the processes although, forward finite difference approximations would have been adequate for testing the linear functions. In a central finite difference approach, the forward model is run twice; first with initial concentration field being added with positive perturbations, and with negative perturbations in the second. A third run consists of a forward and an associated adjoint model run where the forward model integration is performed with unaltered initial conditions and adjoint variables (receptors) initialized at carefully chosen sensitive locations are integrated backwards to be tested against the finite difference approximations. Except for emissions where perturbations are brought in the emission rate coefficients, we use a 10% perturbation in the initial state vector for validation of all the science processes. For emissions, we use a scaling factor of 0.001 that changes the emission rates by 0.1% which is still a large quantity considering the magnitude of the emission rate coefficients.

Details on the construction and validation of individual processes and the combined performance are provided next.

2.4 Chemistry Adjoint Model

This section discusses the construction and implementation of adjoint gas-phase chemistry in GEOS-Chem via symbolic preprocessing.

The atmospheric chemical kinetics is highly non linear because some species are extremely reactive in nature while some are long lived. Also, the reaction rates for photochemical reactions are heavily affected based on the time of the day while it changes for other reactions due to change in temperature, air density or other natural phenomenon such as lightning. This makes atmospheric chemical kinetics a stiff ordinary differential equation system that entails special numerical integration methods which maintain stability, preserve mass balance, and are computationally efficient. As [Sandu et al. , 1997] elaborates, Rosenbrock methods are well suited for solving such problems.

The mathematical formulation of chemical kinetics as presented in the forward model equation (2.1a) is

$$\frac{dc}{dt} = f(t, c; p), \quad c(t^0, x) = c^0(x), \quad t^0 \leq t \leq T \tag{2.33}$$

where $c(t, x)$ represents the concentration vector evolving in time, p represents the reaction rate coefficients and other model parameters, $f = \dot{c}$ represents the non-linear

product-loss function that determines the rate of change in the concentration vector.

The discrete chemical model advances the forward solution at time t^n as per following s-stage Rosenbrock method [Hairer and Wanner , 1991]

$$\begin{aligned} C_i &= c^n + \sum_{j=1}^{i-1} a_{i,j} k_j, \quad T_i = t^n + \alpha_i h, \\ \left(\frac{1}{h\gamma} - J(t^n, c^n)\right) &= f(T_i, C_i) + \sum_{j=1}^{i-1} \frac{b_{i,j}}{h} k_j + h\gamma_i f_t(t^n, c^n), \\ c^{n+1} &= c^n + \sum_{i=1}^{s} q_i k_i \end{aligned} \tag{2.34}$$

where s is the number of stages. The formula coefficients $(q_j, a_{i,j}, b_{i,j}, \alpha_i, \gamma_i)$ provide the order of consistency and stability properties. $f_t(.,.)$ is the partial time derivative of the function $f_t(t,c) = \partial f(t,c)/\partial t$, $J(.,.)$ is the Jacobian $J(t,c) = \partial f(t,c)/\partial c$, and C_i, T_i, k_i are internal stage quantities defined by the method. At each stage of the method, a linear system of equations with unknowns k_i and matrix $\left(\frac{1}{h\gamma} - J\right)$ must be solved.

The adjoint of the above system (2.34) is obtained by differentiating it with respect to c^n

$$\begin{aligned} v_i &= J(T_i, C_i) \cdot u_i, \quad i = s, s-1, \ldots, 1, \\ \left(\frac{1}{h\gamma} - J(t^n, c^n)\right) \cdot u_i &= q_i \lambda_c^{n+1} + \sum_{j=i+1}^{s} \left(a_{j,i} v_j + \frac{b_{j,i}}{h} u_j\right), \\ \lambda_c^n &= \lambda_c^{n+1} + \sum_{i=1}^{s} (H(t^n, c^n) \times k_i)^{\mathrm{T}} \cdot u_i + h J_t(t^n, c^n) \cdot \sum_{i=1}^{s} \gamma_i u_i + \sum_{i=1}^{s} v_i \end{aligned} \tag{2.35}$$

where $J_t(.,.)$ is the partial time derivative of the Jacobian $J_t(t,c) = \partial J(t,c)/\partial t$, $H(.,.)$ is the Hessian $H(t,c) = \partial^2 f(t,c)/\partial c^2$, and u_i, v_i are the internal stage vectors defined by the method. Since the reaction rates are constant in GEOS-Chem, the adjoint system (2.35) reduces to

$$\lambda_c^n = \lambda_c^{n+1} + \sum_{i=1}^{s} (H(t^n, c^n) \times k_i)^{\mathrm{T}} \cdot u_i + \sum_{i=1}^{s} J(T_i, C_i) \cdot u_i \tag{2.36}$$

2.4.1 Implementation of forward GEOS-Chem chemistry using KPP

We have implemented chemistry simulations in GEOS-Chem using the Kinetic Pre-Processor (KPP) [Sandu et al. , 2003] retaining the native SMVGEARII solver [Jacobson and Turco , 1994; Jacobson , 1998] that comes with the package. KPP provides a library of chemical solvers together with their tangent linear and adjoint integrators. It generates

very effective sparse matrix computational kernels which lead to high computational efficiency. We can take advantage of the efficient stiff ODE solvers implemented including the mechanism-specific sparse linear algebra routines. We next discuss the interfacing of KPP generated gas-phase chemistry simulation code in to the GEOS-Chem.

Automatic translation of SMVGEARII inputs to KPP inputs

KPP requires a description of the chemical mechanism in terms of chemical species, chemical equations, and mechanism definition. Based on this input KPP generates all the Fortran90 files required to carry out the numerical integration of the mechanism with the numerical solver of choice. GEOS-Chem SMVGEARII uses the mechanism information specified in the "globchem.dat" file; this file contains a chemical species list and a chemical reactions list. The perl parser geos2kpp_parser.pl translates the chemical mechanism information from "globchem.dat" into the KPP syntax and outputs it in the files "globchem.def", "globchem.eqn", and "globchem.spc".

The perl parser geos2kpp_parser.pl makes the translation of the chemical mechanism information a completely automatic process. The user invokes the parser with the "globchem.dat" input file and obtains the KPP input files.

Creation of KPP model to interface with GEOS-Chem

Once the input files are ready, KPP is called to generate all Fortran90 subroutines needed to describe and numerically integrate the chemical mechanism. We have slightly modified the KPP code generation engine to assist the integration of the KPP subroutines within GEOS-Chem. A new KPP input command (#GEOSCHEM) instructs to produce code that can interface with GEOS-Chem. KPP is invoked with this input file and generates complete code to perform forward and adjoint chemistry calculations. The model files are named gckpp_* or gckpp_adj_*, respectively. These files are then copied into the GEOS-Chem code directory.

Automatic modification of GEOS-Chem source code to interface with KPP using gckpp_parser.pl

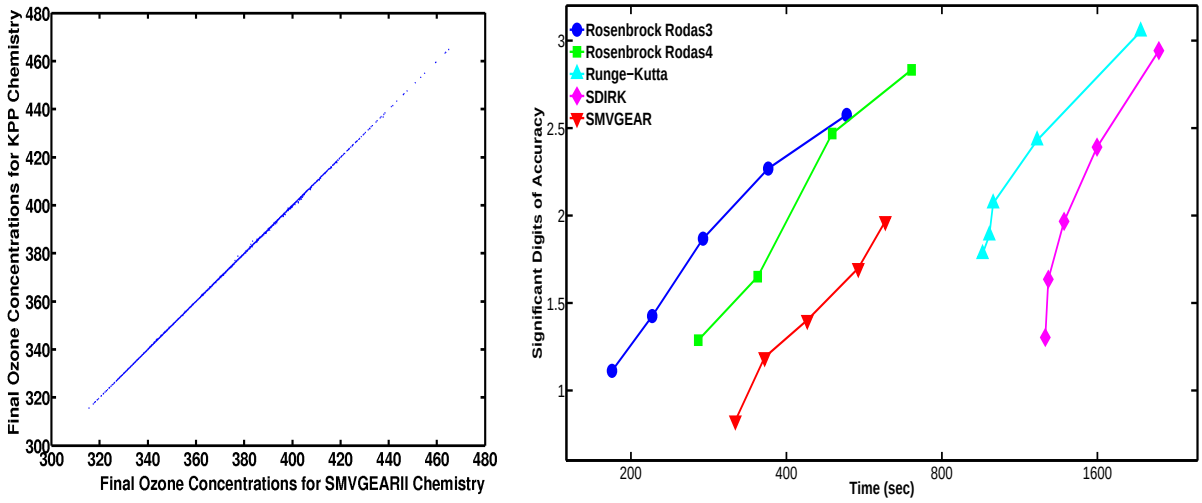
The last step involves running the gckpp_parser.pl to modify the GEOS-Chem source code to interface with KPP. This parser is run once in the GEOS-Chem code directory, modifies existing files to use KPP instead of SMVGEARII for the chemistry step, and adds new files with subroutines for adjoint calculations.

The GEOS-Chem code modifications are necessary for a correct data transfer to and from KPP. KPP performs a reordering of the chemical species to enhance sparsity gains, and provides subroutines to map GEOS-Chem data structures to KPP data structures

and vice versa. The GEOS-Chem calculated rate coefficients are mapped from the SMVGEARII reaction ordering to the KPP ordering via a reshuffling. The resulting GEOS-Chem code now uses KPP for gas-phase chemistry.

Forward Integration Results

The newly interfaced chemistry mechanism needs to be validated as an incorrect chemical kinetics would lead to wrong model predictions. Figure 2.1(a) provides the scatter plot of SMVGEARII vs KPP/Rodas3. The simulation interval is one week, between 00:00 GMT on 07/01/2001 and 00:00 GMT on 07/08/2001, using an absolute tolerance of ATOL=10^{-1} and a relative tolerance of RTOL=10^{-3}. Both simulations are started with the same initial conditions and they differ only in the chemistry module. Each point in the scatter plot corresponds to the final O_x concentration in a different grid cell. The O_x concentrations obtained with KPP and with SMVGEARII are very similar to each other.



(a) Scatterplot of O_x concentrations computed with SMVGEARII and KPP

(b) Work-precision diagram (Significant Digits of Accuracy versus run time)

Figure 2.1: Scatterplot of O_x concentrations (molecules/cm^3) computed with SMVGEARII and KPP for a one week simulation (Mean Error = 0.05%, Median Error = 0.03%). Work-precision diagram (Significant Digits of Accuracy for O_x versus run time) for a seven day chemistry-only simulation for RTOL=1.0E-1, 3.0E-2, 1.0E-2, 3.0E-3, 1.0E-3 (lower left to upper right). Rodas4 does not have a point for RTOL=1.0E-1 due to not producing meaningful results.

We also compare the efficiency of the forward SMVGEARII and KPP solvers over the simulation period. Specifically we compare the computed concentrations of each species in each grid cell against a reference solution for each solver. Calculations are performed

using relative tolerances (RTOL) in the range $10^{-1} \leq \text{RTOL} \leq 10^{-3}$ and absolute tolerances ATOL = $10^4 \times$ RTOL molecules cm^{-3}. Reference solutions are calculated using RTOL=10^{-8} and ATOL=10^{-3} molecules cm^{-3}.

Following Henze et al. (2007) and Sandu et al. (1997), we use significant digits of accuracy(SDA) to measure the numerical errors. We calculate SDA using

$$SDA = -\log_{10}\left(\max_{k}\left(\sum_{c_{k,j}\geq a} 1\right)^{-1} \cdot \sum_{c_{k,j}\geq a}\left|\frac{c^{\text{ref}}_{k,j} - \widehat{c}_{k,j}}{c^{\text{ref}}_{k,j}}\right|^2\right)$$

This uses a modified root mean square norm of the relative error of the solution ($\widehat{c}_{k,j}$) with respect to the reference solution ($c^{\text{ref}}_{k,j}$) for species k in grid cell j. A threshold value of a=10^6 molecules cm^{-3} avoids inclusion of errors from species concentrations with very small values.

Figure 2.1(b) presents a work-precision diagram where the number of accurate digits in the solution is plotted against the time needed by each solver to integrate the chemical mechanism. Chemistry-only simulation are carried out using the KPP Rosenbrock Rodas3, Rosenbrock Rodas4, Runge-Kutta, and Sdirk integrators and with the SMVGEARII integrator for each tolerance level. The results indicate that, for the same computational time, the KPP Rosenbrock integrators produce a more accurate solution than the SMVGEARII integrator. The Sdirk and Runge-Kutta integrators are slower at lower tolerances in comparison to the Rosenbrock and SMVGEARII integrators. The Runge-Kutta and Sdirk integrators normally take longer steps, but they take fewer steps at higher tolerances. Since we normally use tolerances of 10^{-3} or lower for GEOS-Chem, we do not get these advantages. These results are similar to the results produced by [Henze et al. , 2007] demonstrating that the KPP Rosenbrock solvers are about twice as efficient as SMVGEARII for a moderate level of accuracy. Note that none of the KPP solvers uses vectorization.

We have used $4^\circ \times 5^\circ$ resolution in all our experiments. There are 46 x 72 latitude-longitude grid boxes at this resolution, and 55 vertical levels, each integrating 106 chemical species (87 variables species and 19 fixed species) and 311 reactions. An approximate of 35MB of data to hold species concentrations and 104MB of data to hold the reaction rates is required. Experiments are performed on a Dell Precision T5400 workstation with 2 quadcore Intel(R) Xeon(R) E5410 processors, clock speed 2.33GHz, and a RAM of 16GB shared between the two processors.

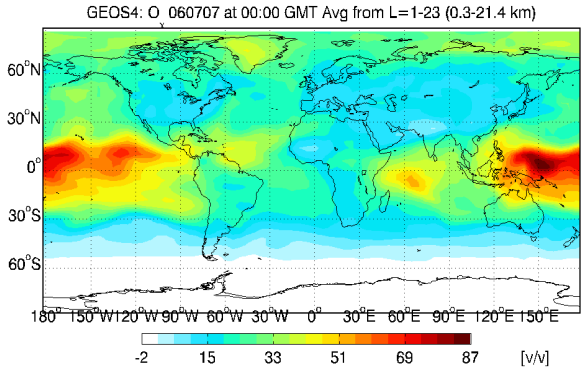
2.4.2 Implementation of adjoint GEOS-Chem chemistry using KPP

The Kinetic Pre-Processor generates chemistry adjoint solver routines in a similar fashion as it generates the forward chemistry code. The generated files are interfaced with the GEOS-Chem adjoint code, updating the KPP global variables, parameters and initialization files. There are separate checkpoint files for the chemical concentrations (CSPEC) and for reaction rate coefficients (RRATE). The CSPEC and RRATE arrays written each hour at the beginning of the forward chemical step are retrieved in the backward run before each chemistry forward-adjoint hourly calculations.

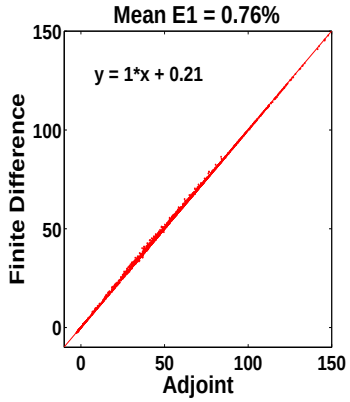
In order to test the accuracy of the adjoint of chemistry, we perform a box model test. Switching off all other transport processes restricts any exchange of mass between grid boxes and a change in the concentration field in any of the grid boxes evolved in time due to chemical reactions is contained with in that box. Thus a finite difference test could be set up and tested for each grid box simultaneously. Consider a chemical species (say S) with respect to which the derivative (dC_I/dC_S) of the concentration of another tracer species (say I) has to be calculated. This leads to the cost function (2.31) being $\mathcal{J}(c^0) = \sum_m c_I^N[m]$. For central finite difference calculations, the GEOS-Chem forward model is run twice; first with added perturbed concentration ($C^0_{S+} = C^0_S + \delta C^0_S$), second with negative perturbation ($C^0_{S-} = C^0_S - \delta C^0_S$). The third is the adjoint run in which a forward run is performed with original concentration field and in the backward mode, the adjoint tracer array for species I is initialized with unit concentration over the grid points under consideration at the final time (T) integrating backwards to the initial time. At $t = 0$, the adjoint tracer concentrations for species S (λ^0_S) are compared with their finite difference approximations in the following manner

$$\lambda^0_S \approx \frac{(C^f_{I+} - C^f_{I-})}{(2\ \delta C^0_S)} \tag{2.37}$$

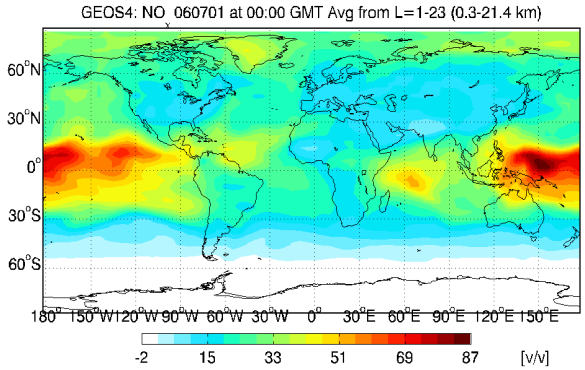
To validate the adjoint of chemistry, we perturb the initial concentration of NOx and measure the changes in Ox concentrations in each grid box over a period of 6 days from 00:00 GMT July 1, 2006 to 00:00 GMT July 7, 2006. Presented in Figure 2.2 are the validation results where subfigures 2.2(a) and 2.2(c) represent the finite difference approximation and adjoint values at each longitude-latitude point averaged over 23 GEOS-Chem levels as described in equation (2.37). A visual inspection of the plots show that the constructed adjoint of chemistry routine agrees well with its finite difference approximation. For a more rigorous check, we generate a scatter plot (subfigure 2.2(b)) and a relative error cumulative distribution function (subfigure 2.2(d)) of adjoint values versus their approximations. The plots reflect that the two values agree with a mean relative difference of less than 1% where 97% of the points are within 5% of this difference.



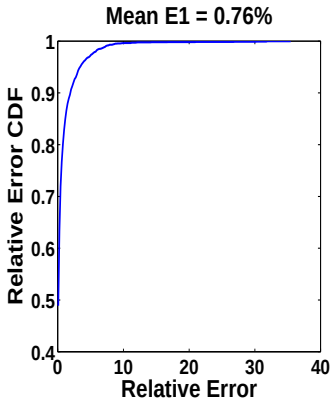
(a) Finite difference approximation dO_x/dNO_x at 00:00 GMT 07/07/2006.



(b) Scatter plot of finite difference approximation versus adjoints.



(c) Adjoint dO_x/dNO_x at 00:00 GMT 07/01/2006.



(d) Finite difference approximation versus adjoints relative error cumulative distribution function plot.

Figure 2.2: Comparison between adjoint and central finite difference approximation generated by running GEOS-Chem v7 adjoint, **chemistry** only simulation for 6 days from 00:00 GMT 07/01/2006 to 00:00 GMT 07/07/2006, for Ox with respect to NOx concentrations averaged over 23 layers.

2.5 Advection Adjoint Model

This section discusses the implementation of continuous adjoint of advection in GEOS-Chem.

The one dimensional (longitudinal) advection equation written in the flux form is

$$\frac{\partial \mu}{\partial t} = -\frac{\partial (u\mu)}{\partial x} \tag{2.38}$$

where u is the wind velocity and μ is the mass-based concentration vector ($\mu = \rho c$, ρ is the air density). If the total mass in the system is conserved, the flux form (2.38) could

be rewritten as

$$\frac{\partial c}{\partial t} = -u\frac{\partial c}{\partial x} \tag{2.39}$$

The above equation (2.39) follows directly from the forward model equation (2.1a). An infinitesimal perturbation in c^0 follows the tangent linear equation that is equivalent to (2.39) since u is scalar. Therefore, using the Lagrange identity (2.12), the one dimensional continuous adjoint of the mass conserved advection equation is

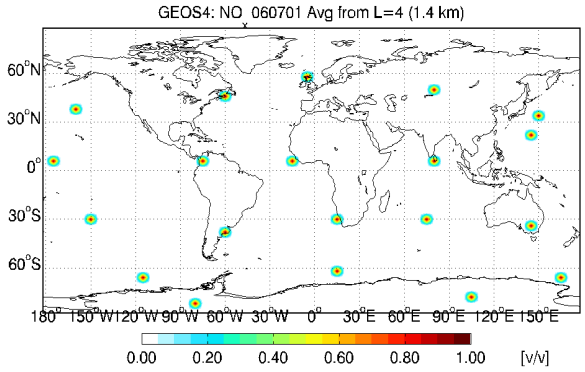
$$\frac{\partial \lambda}{\partial t} = -\frac{\partial (u\lambda)}{\partial x} \tag{2.40}$$

where the adjoint variable λ in the mass conserved form is related to λ_m in mixing ratio form by $\lambda = \lambda_m/\rho$. To achieve the mass conserved form of advection in the forward mode, functions "COUPLE" and "CONVERT_UNITS" are provided in CMAQ and GEOS-Chem respectively. In the backward integration mode, the adjoint variables are applied with the reverse of these operations.

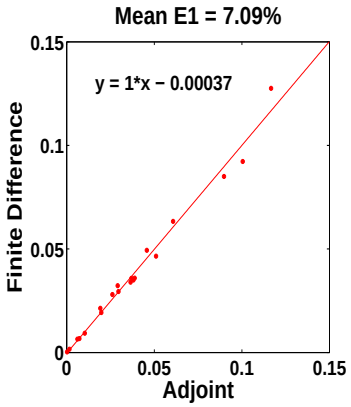
Validating the adjoint of advection routine is a rigorous task. The concentration field in one grid box is spread and transported to other grid boxes according to the wind velocity at that location. The task of locating the spread for each grid box over a certain time interval is highly intricate and it gets further complex if the spreads overlap. Therefore, in order to test the accuracy of advection adjoint, we first choose a tracer and then carefully locate a set of points and a simulation interval such that if we start with a concentration field for only that tracer from the chosen locations and evolved forward in time using advection process, then at the end of the simulation interval, the spread in the concentration fields for each starting location do not interfere. Using this spread information, we place receptors at locations with in the spread such that when adjoint variables initialized at these receptor points are integrated backwards through advection adjoint, they are spread back to the starting locations.

In order to verify the accuracy of advection adjoint, we choose tracer NOx and 20 starting locations over the globe at level 4 with 5 points evenly spaced in the longitudinal direction (5,15,35,50,65) and 4 points evenly spaced in the latitudinal direction. We then perform the runs required for central finite difference calculations for a period of 2 days starting from 00:00 GMT July 1, 2006 and ending at 00:00 GMT July 3, 2006. Presented in Figure 2.3 are the validation results and plots showing the selected locations. Subfigure 2.3(a) represents the receptor locations at which adjoint variables for NOx are initialized while 2.3(c) represents the spread in the NOx adjoint field. It is evident from the plots that at the end of backward integration, the adjoint variables retrieve back the starting location described above.

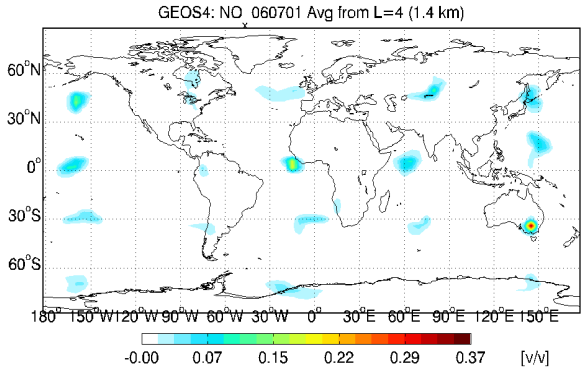
Subfigures 2.3(b) and 2.3(d) reflect that the adjoint values agree with their finite difference approximations with a mean relative difference of 7% with 95% points within 10%



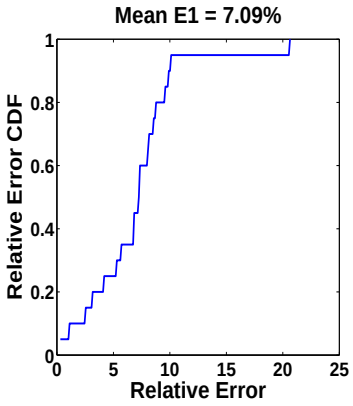
(a) Finite difference approximation $dNO_{x_{rl}}/dNO_{x_{sl}}$ at 00:00 GMT 07/03/2006.



(b) Scatter plot of finite difference approximation versus adjoints.



(c) Adjoint $dNO_{x_{rl}}/dNO_{x_{sl}}$ at 00:00 GMT 07/01/2006.



(d) Finite difference approximation versus adjoints relative error cumulative distribution function plot.

Figure 2.3: Comparison between adjoint and central finite difference approximation generated by running GEOS-Chem v7 adjoint, **advection** only simulation for 2 days from 00:00 GMT 07/01/2006 to 00:00 GMT 07/03/2006, for NOx concentrations for a set of carefully chosen receptors at level 4. *rl* and *sl* refer to the receptor and starting locations respectively.

of this difference. It is important however to note that the relative difference for advection adjoint is highest among all the inidividual science processes owing partly to the fact that continuous adjoints are not as accurate as their discrete counterparts, and partly to the fact that locating the receptors based on the spread information is an erroneous task.

2.6 Convection Adjoint Model

This section discusses the implementation of adjoint convection in GEOS-Chem.

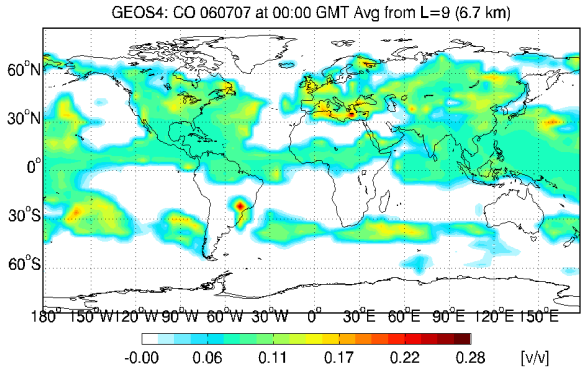
The GEOS-4 convection scheme in GEOS-Chem is a simplified cumulus parameterization scheme proposed by [Hack , 1994] and [Zhang and McFarlane , 1995]. Cumulus convection affects large-scale temperature and moisture fields through subgrid-scale evaporation and condensation described as

$$C_p\left(\frac{\partial T}{\partial t}\right) = -\frac{1}{\rho}\frac{\partial(M_u S_u + M_d S_d - M_c S)}{\partial z} + L(c-\epsilon) \tag{2.41}$$

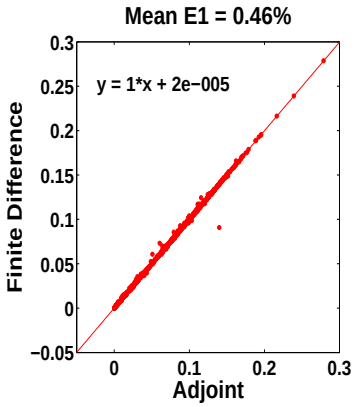
$$\left(\frac{\partial q}{\partial t}\right) = -\frac{1}{\rho}\frac{\partial(M_u q_u + M_d q_d - M_c q)}{\partial z} + (\epsilon - c) \tag{2.42}$$

where the net vertical mass flux within the convective region, M_c, is made up of upward (M_u) and downward (M_d) components. Here, c and e are respectively the large-scale mean rates of condensation and of evaporation, L is the latent heat and ρ is the air density; q, q_u and q_d are respectively the large-scale, the convective scale updraft and downdraft components of the specific humidity field; and S, S_u, S_d are respectively the corresponding values of dry static energy (defined in the usual way as $S_{u,d} = C_p T_{u,d} + gz$); $C_p T$ is the enthalpy of vaporization.

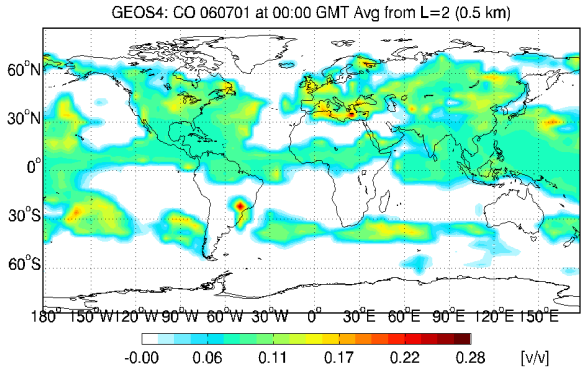
The adjoint of convection routine is discrete and has been constructed using the tangent linear and adjoint model complier (TAMC). To avoid multiple loops of recalculations during backward integration, dependent variables are checkpointed every dynamic time step in the forward model run and read in during the adjoint run. In order to test the accuracy of the developed convection adjoint, we perform a column testing. With all the other processes switched off, forward convection routine evolves the concentration field through vertical columns with the mass being conserved. For the validation test set up, a chemical species (say S) and a layer (say L) are selected with respect to which the derivative (dC_{S_H}/dC_{S_L}) of the concentration of the same species at a higher layer (say H) has to be calculated. We performed a column testing for carbon monoxide where perturbations were introduced in the concentration vector for CO at 00:00 GMT on July 1, 2006 for each GEOS-Chem longitude-latitude grid point at level 2 and the central finite difference approximations were calculated at level 9 at the end of 6 days, 00:00 GMT on July 7, 2006. Figures 2.4(a) and 2.4(c) present a visual conformation between the adjoint values and their finite difference approximations, while Figures 2.4(b) and 2.4(d) reflect that the two agree with a mean relative error of 0.49% where 98% of the points are within 2% of their relative errors.



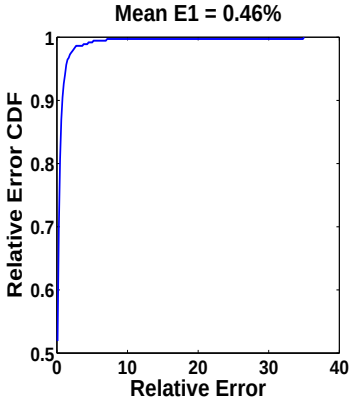
(a) Finite difference approximation dC_{CO_9}/dC_{CO_2} at level 9, 00:00 GMT 07/07/2006.



(b) Scatter plot of finite difference approximation versus adjoints.



(c) Adjoint dCO_9/dCO_2 at level 2, 00:00 GMT 07/01/2006.



(d) Finite difference approximation versus adjoints relative error cumulative distribution function plot.

Figure 2.4: Comparison between adjoint and central finite difference approximation generated by running GEOS-Chem v7 adjoint, **convection** only simulation for 6 days from 00:00 GMT 07/01/2006 to 00:00 GMT 07/07/2006, for CO concentrations in GEOS-Chem vertical columns.

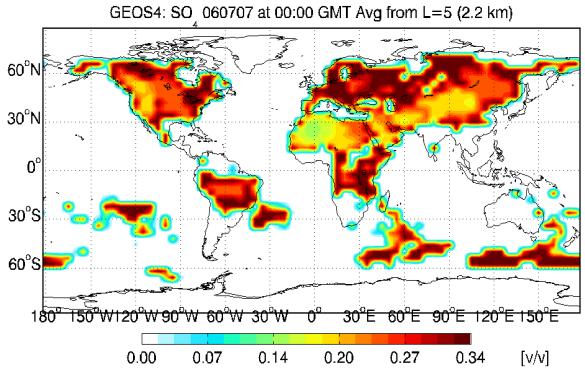
2.7 Turbulent Mixing Adjoint Model

This section discusses the implementation of adjoint turbulent mixing in GEOS-Chem.

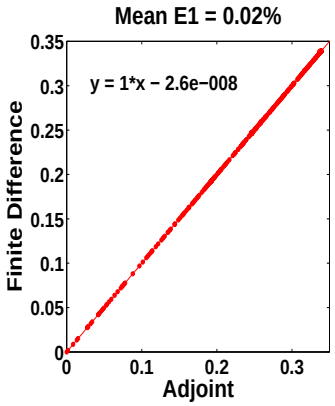
Turbulent mixing in GEOS-Chem computes the planetary boundary layer (PBL) height and mixes the tracers underneath the top boundary layer. The mass-weighted mixing algorithm provided by [Allen , 1996], is applied to the concentration vector in the forward model every internal time step as described in the operator splitting formulation (2.3)

$$v_{i,k}^{n+1} = \frac{\sum_{l=1}^{L} m_l v_{i,l}^{n}}{M}, \quad k = 1, 2, \ldots, L, \quad i = 1, 2, \ldots, s \tag{2.43}$$

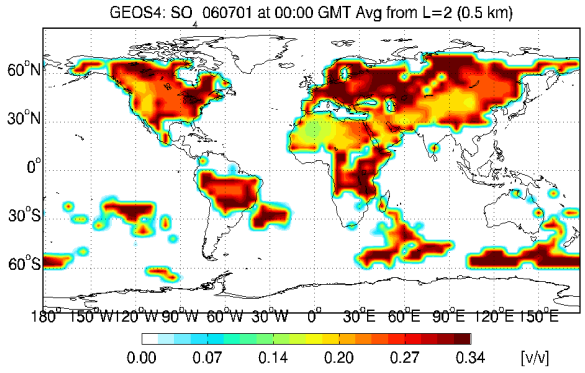
where $v_{i,k}$ is the mixing ratio of tracer i in layer k ($v = c/\rho$, ρ is the density of air), m_l is the air mass in a single layer l, M is the total air mass in the boundary layer column, and L is the number of layers in the boundary layer.



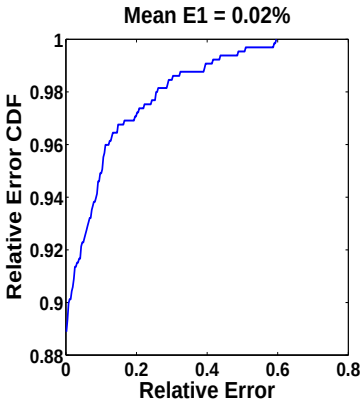
(a) Finite difference approximation $dC_{SO_{x_5}}/dC_{SO_{x_2}}$ at level 5, 00:00 GMT 07/07/2006.



(b) Scatter plot of finite difference approximation versus adjoints.



(c) Adjoint dSO_{x_5}/dSO_{x_2} at level 2, 00:00 GMT 07/01/2006.



(d) Finite difference approximation versus adjoints relative error cumulative distribution function plot.

Figure 2.5: Comparison between adjoint and central finite difference approximation generated by running GEOS-Chem v7 adjoint, **turbulent mixing** only simulation for 6 days from 00:00 GMT 07/01/2006 to 00:00 GMT 07/07/2006, for SOx concentrations in GEOS-Chem between level 2 and level 5.

Equation (2.43) can be written in the matrix form as follows

$$\begin{bmatrix} v_{i,1} \\ \vdots \\ v_{i,L} \end{bmatrix}^{n+1} = \begin{bmatrix} \frac{m_1}{M} & \cdots & \frac{m_L}{M} \\ \vdots & \ddots & \vdots \\ \frac{m_1}{M} & \cdots & \frac{m_L}{M} \end{bmatrix} \cdot \begin{bmatrix} v_{i,1} \\ \vdots \\ v_{i,L} \end{bmatrix}^{n} \tag{2.44}$$

In the above equation, the matrix elements are independent of the concentration field

and therefore a small perturbation in the initial state vector will be propagated through the tangent linear model of the same structure as (2.45). Using the duality principle (2.23), the discrete adjoint of the turbulent mixing method is given as

$$\begin{bmatrix} \lambda_{\nu_{i,1}} \\ \vdots \\ \lambda_{\nu_{i,L}} \end{bmatrix}^{n} = \begin{bmatrix} \frac{m_1}{M} & \cdots & \frac{m_L}{M} \\ \vdots & \ddots & \vdots \\ \frac{m_1}{M} & \cdots & \frac{m_L}{M} \end{bmatrix}^{\mathrm{T}} \cdot \begin{bmatrix} \lambda_{\nu_{i,1}} \\ \vdots \\ \lambda_{\nu_{i,L}} \end{bmatrix}^{n+1} \tag{2.45}$$

The validation set up for turbulent mixing adjoint is similar to convection since with all other processes switched off, the mixing algorithm preserves the mass in the vertical column from ground level to the top planetary boundary layer which is ranges anywhere from 0.1 Km up to 3 Km. To perform the finite difference calculations, we started with initial concentration vector and brought in a perturbation in trace gas SOx at each horizontal grid point at level 2 at 00:00 GMT on July 1, 2006 and measured the changes in the concentration of the same species at level 5 at 00:00 GMT on July 7, 2006, at the end of a 6-day period.

Comparing Figure 2.5(a) which represents the central finite difference approximations of $dC_{SO_{x_5}}/dC_{SO_{x_2}}$ computed at level 5 against the adjoint values 2.5(c) calculated at level 2, it is evident that the turbulent mixing adjoint works well. Due to the reason that turbulent mixing algorithm (2.43) is linear and the constructed adjoint is discrete, the agreement between the adjoint values and their finite difference approximations is very high. Figures 2.5(b) and 2.5(d) reflect that the two entities agree with a mean relative difference of 0.02% where 100% of the points are within 0.7% of their relative differences.

2.8 Emission and Dry Deposition Adjoint Models

This section discusses the implementation of emission and dry deposition adjoints in GEOS-Chem via symbolic preprocessing.

We next consider the adjoints of emission and dry deposition. In GEOS-Chem, emission and dry deposition are handled through chemistry via fake chemical equations. The rate coefficients for these processes are calculated separately and then attached to the chemistry reaction rates. The adjoints of these subroutines are scaled and are calculated using the adjoint integrator. The adjoint integrator provides adjoints with respect to the rates which are then multiplied with the individual rates and accumulated over time. As discussed in [Henze et al. , 2007], KPP does not directly implements the adjoint code for emission or dry deposition, however, it generates required underlying routines $dFun\_dRcoeff$ and $dJac\_dRcoeff$ that provide the gradients of function and Jacobian with respect to the reaction rate coefficients. Since the underlying mechanism for emission and dry deposition are similar, we will be discussing the adjoint construction only for emissions.

The rate of change in the state vector $c(t,x)$ due to emissions is governed by

$$\frac{dc_k}{dt} = E_k, \quad k = 1, 2, \ldots, s \tag{2.46}$$

where c_k represents the concentration values of species k and E_k is the rate at which species k is entered into the system via emission. The continuous adjoint of the above equation is derived as

$$\lambda_E = \int \lambda_c dt \tag{2.47}$$

The above indicates that the emission adjoint is required to be calculated through the sensitivity of discrete chemical solver itself. The scaled emission sensitivities generated through the chemical adjoint solver needs to be multiplied with emission rate coefficients and be accumulated over time to generate the adjoint of emission. The discrete adjoint equation for scaled emissions can be derived from the s-stage Rosenbrock formulation in a similar fashion as (2.36)

$$\lambda_e^n = \lambda_e^{n+1} + \sum_{i=1}^{s} (J_e(t^n, c^n) \times k_i)^{\mathrm{T}} \cdot u_i + \sum_{i=1}^{s} f_e(T_i, C_i) \cdot u_i \tag{2.48}$$

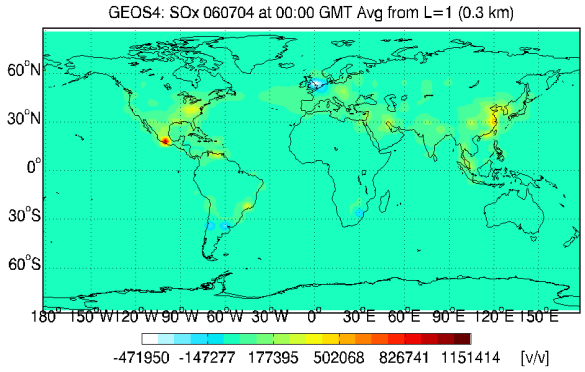
For emissions, the function derivative is simply identity matrix and the Jacobian derivative is zero. Therefore the discrete adjoint of scaled emissions is given by

$$\lambda_e^n = \lambda_e^{n+1} + \sum_{i=1}^{s} I \cdot u_i \tag{2.49}$$

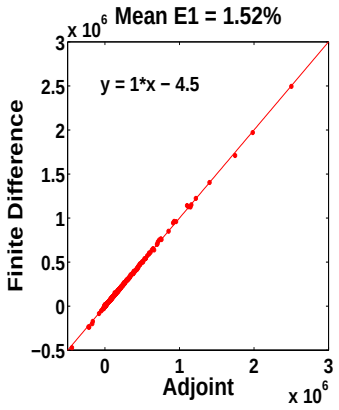
Since the adjoints of emission and dry deposition are calculated through the chemical solver adjoint, we switch off all the processes except chemistry, emission and dry deposition for adjoint validation tests. We will be presenting validation results only for the adjoint of emission since the underlying mechanism for both emission and dry deposition are very similar.

In order to perform the finite difference calculations, an emission species is chosen and the rate at which this species is added to the system through emission is perturbed using a scaling factor in each grid box at the ground level. Similar to chemistry, with all transport processes switched off, mass in each grid box is conserved. The changes therefore are measured for another species, of which the earlier species is a precursor, in each grid box at gournd level at the end of the simulation window.

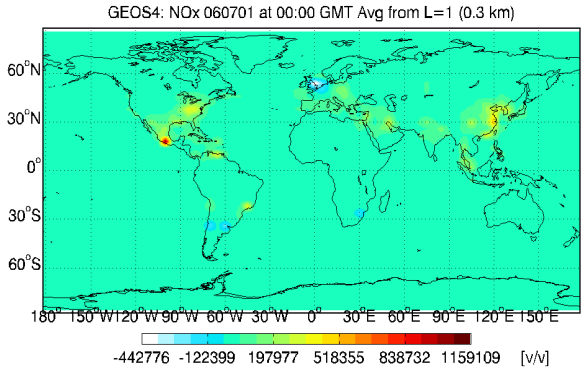
In Figure 2.6, we present the adjoint validation results where we introduce a perturbation in the emission rate for anthropogenic (caused by Human) NOx species every hour starting from 00:00 GMT on July 1, 2006 till 00:00 GMT, July 4, 2006. A central finite difference approximation is then performed for SOx concentrations at the end of the



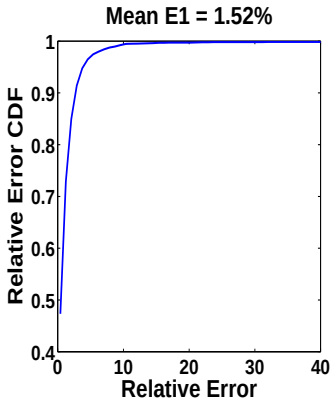
(a) Finite difference approximation dSO_x/dNO_x at 00:00 GMT 07/04/2006.



(b) Scatter plot of finite difference approximation versus adjoints.



(c) Adjoint dSO_x/dNO_x at 00:00 GMT 07/01/2006.



(d) Finite difference approximation versus adjoints relative error cumulative distribution function plot.

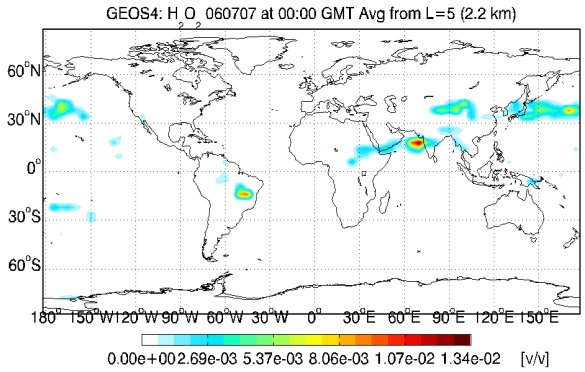
Figure 2.6: Comparison between adjoint and central finite difference approximation generated by running GEOS-Chem v7 adjoint, **emission and dry deposition** only simulation for 3 days from 00:00 GMT 07/01/2006 to 00:00 GMT 07/04/2006, for SOx with respect to NOx concentrations at ground level.

3-day simulation period for each grid box at the ground level. A visual inspection of subfigures 2.6(a) and 2.6(c) reflect that the emission adjoint values agree well with their finite difference approximations. For more rigorous check, we provide a direct point-to-point comparison scatter plot and a relative error cumulative distribution function plot. Subfigures 2.6(b) and 2.6(d) show that the adjoint variables and their finite difference approximations agree with a mean relative error of 1.52% where 96% of the points are within 5% of their relative errors.

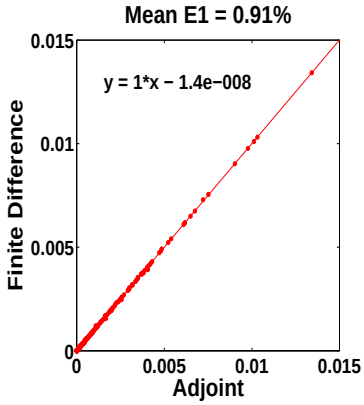
2.9 Wet Deposition Adjoint Model

This section discusses the implementation of wet deposition adjoint in GEOS-Chem.

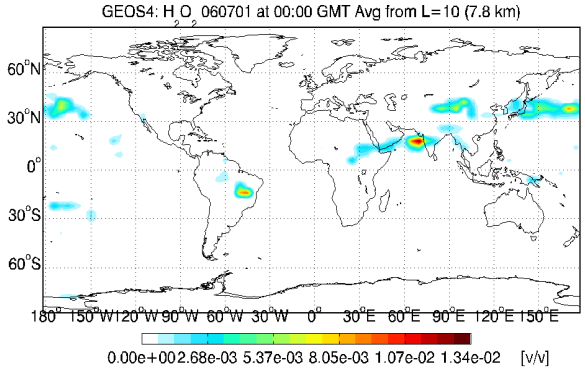
Wet deposition in GEOS-Chem computes the downward mass flux due to washout and rainout of aerosols and soluble tracers. In the current research work, since we are not focusing on aerosols, we will be deriving adjoint only for soluble tracers.



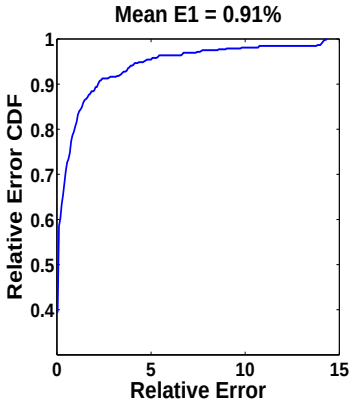
(a) Finite difference approximation $dC_{H2O2_5}/dC_{H2O2_{10}}$ at level 5, 00:00 GMT 07/07/2006.



(b) Scatter plot of finite difference approximation versus adjoints.



(c) Adjoint $dH2O2_5/dH2O2_{10}$ at level 10, 00:00 GMT 07/01/2006.



(d) Finite difference approximation versus adjoints relative error cumulative distribution function plot.

Figure 2.7: Comparison between adjoint and central finite difference approximation generated by running GEOS-Chem v7 adjoint, **wet deposition** only simulation for 6 days from 00:00 GMT 07/01/2006 to 00:00 GMT 07/07/2006, for H2O2 concentrations in GEOS-Chem between level 5 and level 10.

The wet removal algorithm implemented in GEOS-Chem is a first-order process with

discrete forward model equation

$$c_k^{n+1} = c_k^n e^{-r_k \Delta t} \tag{2.50}$$

where c_k is the concentration value of species k and r_k is the rate at which species k is washed out of the system. Since this loss rate is independent of any forward model variables, the tangent linear model has the same structure as (2.50). Therefore, the discrete adjoint for wet deposition process is given as

$$\lambda_k^n = \lambda_k^{n+1} e^{-r_k \Delta t} \tag{2.51}$$

The adjoint of wet deposition is validated using a column testing procedure since the removal of soluble tracers due to washout and rainout takes place only in the vertical column and with all other processes switched off, the mass is conserved for that column. Consider two levels L and H, with $H > L$. We calculate the derivative (dC_{S_L}/dC_{S_H}) of the concentration of a tracer species S at a lower level at the final time with respect to concentration of the same species at a higher level at the initial time.

For finite difference approximation calculations, we introduce perturbations in the concentration field for hydrogen peroxide (H2O2) in each grid box at level 10 of GEOS-Chem at 00:00 GMT on July 1, 2006 and the changes are measured at level 5 after a period of 6 days at 00:00 GMT, July 7, 2006. Presented in Figure 2.7 are the validation result plots where subfigures 2.7(a) and 2.7(c) provide a visual conformation between the adjoint values and their finite difference approximations, while subfigures 2.7(b) and 2.7(d) reflect that the two entities agree well with a mean relative difference of 0.91% where 95% of the points are within 5% of their relative differences.

2.10 Stratosphere-Troposphere Ozone Exchange Adjoint Model

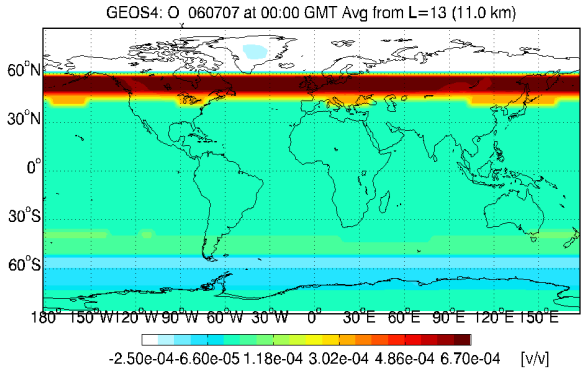
This section discusses the implementation of cross tropopause flux adjoint in GEOS-Chem.

The linearized ozone (linoz) scheme [McLinden et al., 2000] is a stratospheric ozone chemistry mechanism for atmospheric models that focuses on the troposphere and was developed with primary goals of accurate calculation of the cross tropopause flux, and reasonable representation of the ozone gradients near the tropopause. This scheme is simple and computationally efficient. In this method the ozone chemical tendency is expressed as a linear function of ozone, temperature, and the overhead ozone column. The linearizations are performed about an observed climatological state for a standard set of latitudes, months, and altitudes. The 24-hour, zonal-mean ozone photochemical tendency is calculated at each time step for each stratospheric grid box from these monthly

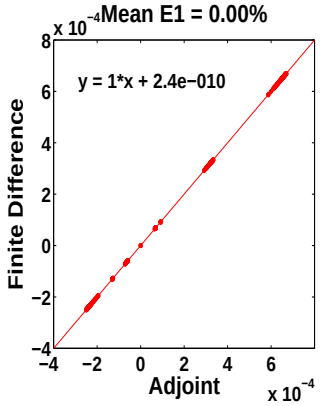
varying coefficients. The change in ozone with time due to local chemistry is given by

$$\frac{df}{dt} = (P - L)[f, T, C_{O3}] \tag{2.52}$$

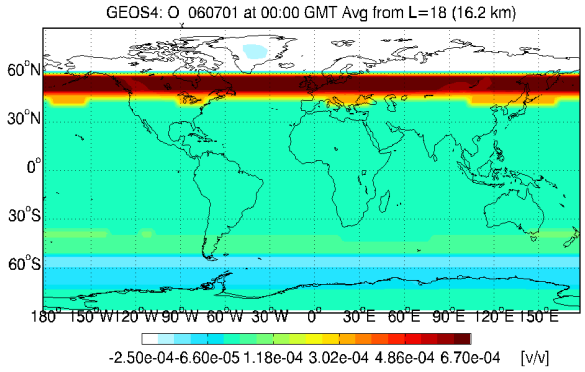
where $(P - L)$ represents the ozone tendency (in units of ppmv/s), the square brackets denote a functional dependence, f is the ozone mixing ratio, T is temperature, and C_{O3} is the column ozone above the point under consideration.



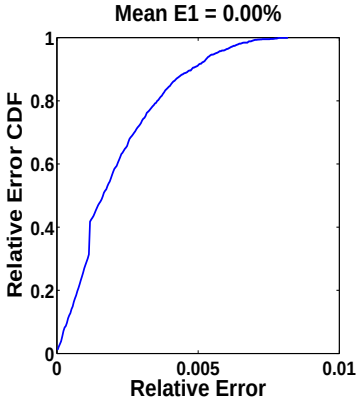
(a) Finite difference approximation $dC_{O_{x_13}}/dC_{O_{x_18}}$ at level 13, 00:00 GMT 07/07/2006.



(b) Scatter plot of finite difference approximation versus adjoints.



(c) Adjoint dO_{x_13}/dO_{x_18} at level 18, 00:00 GMT 07/01/2006.



(d) Finite difference approximation versus adjoints relative error cumulative distribution function plot.

Figure 2.8: Comparison between adjoint and central finite difference approximation generated by running GEOS-Chem v7 adjoint, **linoz** only simulation for 6 days from 00:00 GMT 07/01/2006 to 00:00 GMT 07/07/2006, for ozone concentrations in GEOS-Chem between level 13 and level 18.

The adjoint of the linearized ozone (linoz) scheme is developed through TAMC and integrated into the GEOS-Chem model adjoint code. Since the linoz algorithm is linear and the adjoint calculations required only meteorological field parameters, there was no

checkpointing of forward model variables required. With all the processes switched off, the flux exchanges took place only among the vertical columns conserving the mass of each column. Therefore, in order to test the accuracy of the developed adjoint, a column testing is performed similar to the turbulent mixing case. As the scheme evolves changes only in ozone concentrations, the tracer under consideration is Ox.

Presented in Figure 2.8 are the plots showcasing various means of validating the adjoint of linoz scheme. A central difference approximation to the adjoint values was generated through perturbations in ozone concentrations in each grid point at level 18 (upper tropospheric boundary) of GEOS-Chem at 00:00 GMT on July 1, 2006 and changes in ozone concentrations were measured in all the grid points at level 13 at 00:00 GMT, July 7, 2006. Subfigures 2.8(a) and 2.8(c) provide a visual conformation between the linoz adjoint values against their finite difference approximations.

The band structure of adjoint variable values shows strong linearity property of the scheme where the rate at which ozone is exchanged are very high at $60^{\circ}N$ and $60^{\circ}S$ in the upward and downward directions respectively. Subfigures 2.8(b) and 2.8(d) reflect that the adjoint values agree with their finite difference approximations almost completely with 100% of the points within 0.01% of their relative differences.

The current ozone studies among atmospheric science groups focus mainly on tropospheric ozone concentrations due to man made and naturally occurring activities at lower and mid tropospheric regions owing to the lack of methods that capture the upper tropospheric dynamics on a global scale. Although not discussed here due to the scope of this dissertation, with the development of an adjoint of linoz scheme integrated with a global chemistry transport model such as GEOS-Chem, an effort is in progress in quantifying the amount of ozone in lower and mid troposphere that is a result of exchanges at the upper tropospheric boundary especially considering the fact that the tracer is available in abundance near stratospheric boundary.

2.11 Validation and Performance of the GEOS-Chem Adjoint Model

This section presents the validation of the GEOS-Chem adjoint via careful comparisons against finite difference calculations. The computational performance of the model is also discussed.

The newly developed GEOS-Chem adjoint model is well structured and follows the coding style provided in the GEOS-Chem users manual `http://www-as.harvard.edu/chemistry/trop/geos/doc/man`. The structure of adjoint code is kept similar to the forward model code so that users find it easier to follow. For each science process, all the forward and related adjoint subroutines are kept in the same module file for the ease of

debugging. To handle checkpointing, an additional module file CHECKPOINT_MOD.F has been provided. In addition, various subroutines to perform observation and background cost function calculations, define adjoint variables, include satellite observations are provided through separate files.

The standard GEOS-Chem adjoint package (GCv7_ADJ) is available for download from our project website `http://people.cs.vt.edu/~asandu/Public/GCv7_ADJ`. Users have been provided with nine modes of scientific application frameworks built on top of the original GEOS-Chem code. The source files for each of these modes are in separate directories. FWD_SMV mode is the forward GEOS-Chem code that is available from the Harvard's website. This mode uses the SMVGEAR integrator for chemistry calculations. FWD_KPP is equivalent to FWD_SMV except it uses KPP for chemistry, providing users a suite of fast and highly accurate integrators to choose from. FD_TEST is the finite difference testing module which users can choose to validate existing and newly built adjoint subroutines. ADJ_SENST module provides the infrastructure for adjoint sensitivity analysis studies. The results provided in the next section (Section 2.13) are generated using this module. Rest of the code modules are designed of perform data assimilation experiments using three different assimilation techniques. While, SubOptimalKF uses suboptimal Kalman Filter approach to provide an analysis field that approximately represents the true future state, 3D-Var and 4D-Var modules are designed to perform three dimensional and four dimensional variational data assimilations respectively. All these code modules are capable of integrating real data wherever appropriate, with routines to interface Tropospheric Emission Spectrometer satellite data already included in the code. Users can choose one of these module options by simply (un)commenting the mode option in the run script.

We next discuss the implementation aspects of GEOS-Chem model adjoint code and provide an insight into its computational cost in direct comparison to the forward model code. The forward model code of GEOS-Chem v7 distributed by Harvard's atmospheric modeling group is written in Fortran with options to run the code on SGI/Origin, COMPAQ/Alpha, IBM AIX, Sun Studio and Sun/SPARC, and on Linux using Intel Fortran and PGI compilers. The adjoint of GEOS-Chem has been developed and tested extensively for Intel Fortran compilers. There are a total of 227 files in the original distribution package of GEOS-Chem with several hundred lines of codes on an average in each file. Also provided is a run directory with necessary tracer information, restart files that define the initial conditions for simulations, a file named "input.geos" to let users choose which processes to run and for what time period, and scripts to compile and run the code. The code has a "main.f" driver file that reads all the information provided in "input.geos" file and sets the flags accordingly that decides the scenario to run and meteorological data to read. The driver also calls the science process routines every dynamic time step.

The GEOS-Chem adjoint code is written on top of original distribution package and utilizes the same run directory with an addition of "$RunDir/adjoint" directory to

store intermediate dependent variable values from the forward run (checkpoint files), "$RunDir/opt" directory to store result files and an "input_adj.geos" file that provides settings related to paths for observation data, grid resolution for adjoint code and few settings related to particular experiments. All the scientific application modes in the adjoint package that utilize GEOS-Chem model adjoint such as FD_TEST, ADJ_SENT and 4D-Var have a "*modename*_driver.f" file that calls the forward and backward subdriver files to perform adjoint based calculations. In the FD_TEST and ADJ_SENST modes, forward and adjoint models are called only once, while in 4D-Var, several iterations of forward and backward integrations are performed as part of data assimilation procedure.

Recall equations (2.19) and (2.25) that described the evolution of adjoint variables in time. It is evident from these equations that the adjoint variables are integrated backwards in time. If we zoom further into the backward integration, we will notice that the adjoint of the science processes in the backward mode are invoked in the reverse order of the forward integration mode as presented in equations (2.22) and (2.24). Provided in Figure 2.9 is the call flow of science processes in the forward and adjoint mode of GEOS-Chem.

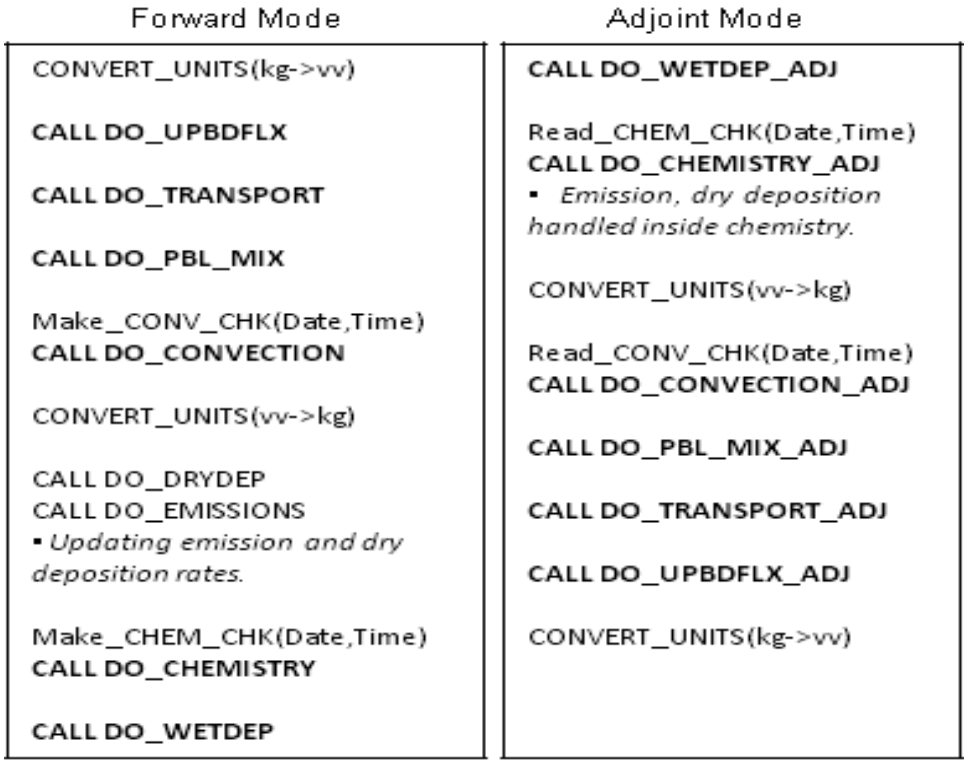
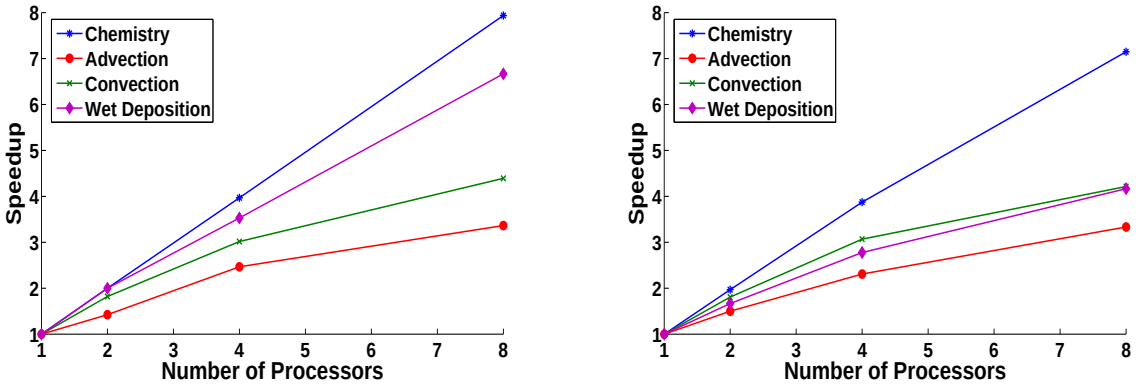


Figure 2.9: GEOS-Chem forward and adjoint function call flows. The adjoint of science processes are called in reverse order in the adjoint mode. Make_*_CHK() are subroutines to create and Read_*_CHK() to read checkpoint files as per the arguments, date and time.

The original GEOS-Chem code is programmed parallel for shared memory systems. One of the main challenges in developing the adjoint model was to parallelize this model completely. For chemistry adjoint we used THREADPRIVATE variables to allow multiple threads to execute the KPP chemistry routines for different grid cells in parallel. Emission and dry deposition adjoints are handled through chemistry. Advection adjoint being continuous derives its parallelism from the forward code. For the adjoints of convection, wet deposition, turbulent mixing and linoz routines, we created OpenMP parallel versions taking care of the thread shared and private variables. The new GEOS-Chem adjoint code is completely parallel and has been tested for consistency against the serial version.



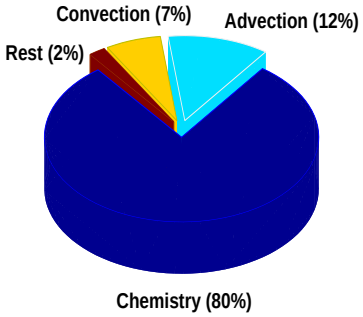
(a) Speedup graph of science processes in forward mode.

(b) Speedup graph of science processes in adjoint mode.

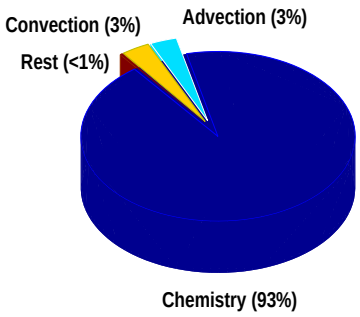
Figure 2.10: Speed up graphs of parallel versions of science processes in the forward and adjoint modes of GEOS-Chem run using 1, 2, 4 and 8 processors averaged over a 6 hour simulation starting at 00:00 GMT July 1, 2006.

We present in Figure 2.10, speed up graphs for parallel versions of science processes in the forward and adjoint modes of GEOS-Chem. The timings were calculated by performing simulations over a 6-hour period starting on July 1, 2006 using one, two, four and eight processors on a machine with same configuration as described towards the end of Section 2.4.1. The plots indicate that the speed up for advection and convection adjoints resemble their forward counterparts. Although not evident till four processors, the speed up for adjoint of chemistry looses its linearity when higher number of processors are used. To understand this phenomenon, we look further into how emission and dry deposition are calculated through chemistry in the forward and adjoint modes. In the forward mode, the emission and dry deposition rates are computed in a separate file "setemis.f" and are then attached to the tracer reaction rate coefficient array. This array is fed to the chemistry solver which is running in parallel in each grid box. In the backward mode, the scaled emission and dry deposition adjoints are generated through the chemical adjoint integrators incurring cost of updating the two variables on top of

tracer adjoint calculations. In addition, the generated scaled adjoints are then multiplied with their respective reaction rates within the chemistry OpenMP parallel loop to generate actual adjoints of emissions and dry deposition. In the case of wet deposition, the plot is a bit deceiving. The speed up in the forward mode is definitely more linear than its adjoint, however, the total computational time in the forward mode for one processor is three times higher than the adjoint mode calculation. The reason could be assigned to the fact that in the forward mode, there are certain parameters calculated which are checkpointed and used in the adjoint mode saving recalculations. It could also be deduced from this observation that the underlying wet deposition algorithm is not completely parallel which is the less time taking than the completely parallel parameter calculation part.



(a) Percentage distribution of times spent by processes in forward mode.



(b) Percentage distribution of times spent by processes in adjoint mode.

Figure 2.11: Distribution plots of wall clock times spent in each science process in the forward and adjoint mode of GEOS-Chem run using a single processor and averaged over a 6 hour simulation starting at 00:00 GMT on July 1, 2006.

As pointed out earlier in the case of wet deposition, speed up graphs provide information only about how the computational cost for a parallel code decreases with increase in the number of processors. A completely parallel code has a perfectly linear speed up graph. However, it is incapable of providing a comparison of actual time spent in each process listed in Figure 2.10. Therefore, in Figure 2.11 we provide a time distribution plot representing the percentage of wall clock times spent in each of the science processes for GEOS-Chem forward and model adjoint codes run on a single processor. The data used to generate this plot is the single processor case of Figure 2.10. Although not visible in the speedup graph, the increase in the percentage of time spent in the chemistry process in backward integration is fairly large as compared to its forward mode equivalent. It is still not possible to deduce from this plot how much actual time was spent in each process, however, it certainly provides an insight into how total forward or backward integration times are divided among several processes.

We next provide in Table 2.1 the wall clock times spent by each process in the forward and backward mode. Quite visible here, the adjoint of chemistry spends a significant amount of time in calculating the scaled emission and dry deposition adjoints. This additional calculation scales well up to four processors, however, takes a dip for higher number of processors as seen in Figure 2.10. It is important to note that we have not provided the timings for other processes such as turbulent mixing and linoz since their computational costs are fairly negligible.

Table 2.1: Timing results for individual science processes in GEOS-Chem forward and adjoint model runs using single processor from a 6-hour simulation starting 00:00 GMT July 1, 2006.

Process Type	Time in Forward Mode (Sec)	Time in Adjoint Mode (Sec)
Chemistry	24.6	88.7
Advection	3.7	3.0
Convection	2.0	3.2
Wet Deposition	0.6	0.2

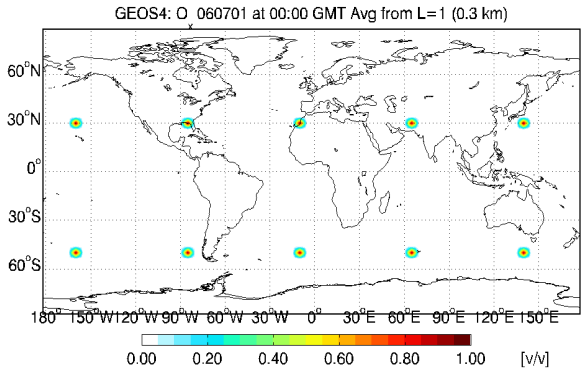
Table 2.2 presents the wall clock time for a 24-hour simulation using GEOS-Chem forward code using SMVGEAR and KPP chemistry solvers. The reason that the computational time is higher in the case of KPP could be attributed to the fact that we have set moderate tolerance levels in KPP to achieve higher accuracies leading to a decrease in the step size and more calculations, and to another fact that KPP invests starting few hours called "spin up time" where it calculates the optimal step sizes for future usage. In longer period simulations, KPP surpasses SMVGEAR with a good margin. Also presented is the time it takes to perform a combined forward and adjoint simulation for one day. Owing in parts to the significant time spent in chemistry adjoint calculations and to the large amount of checkpoint files written ($\sim$500MB per simulation hour) and read, the total time it takes to complete a combined run is about five times of the free model run as opposed to the ideal value of two times.

Table 2.2: Timing results for GEOS-Chem free model run using SMVGEAR and KPP chemistry solvers and a combined forward and adjoint model run from a 24 hour simulation starting 00:00 GMT July 1, 2006.

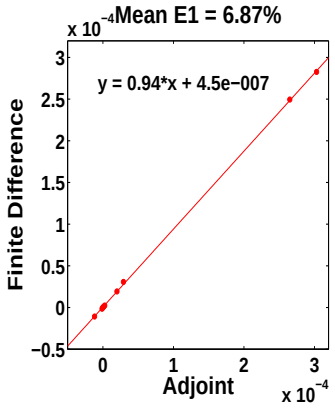
Experiment Description	Real Time
Free model run, SMVGEAR chemistry solver	2 min 50 sec
Free model run, KPP chemistry solver	3 min 18 sec
Combined forward and adjoint model run	16 min 50 sec

Validating the full model adjoint of GEOS-Chem is a difficult task as it includes the intricacies involved with advection adjoint validation with an added complexity due

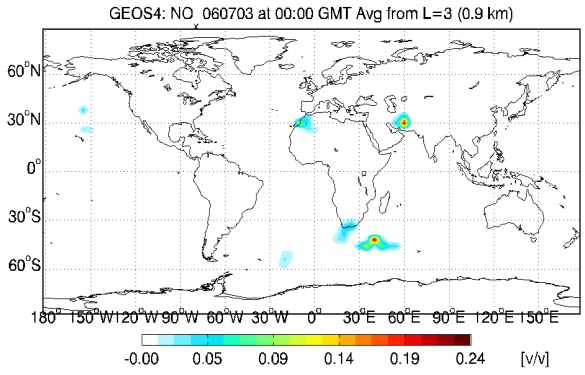
to chemical reactions and various other processes acting on the state vector. To test the accuracy of the full model adjoint, we choose to approximate through central finite differences the sensitivity of Ox in receptor locations at ground level at 00:00 GMT on July 3, 2006 with respect to perturbations introduced in the NOx tracer concentration at level 3 at 00:00 GMT on July 1, 2006 at carefully chosen starting locations. This set up captures the flux exchanges in the vertical columns, chemical reactions in respective grid boxes and horizontal transport in the longitudinal-latitudinal directions.



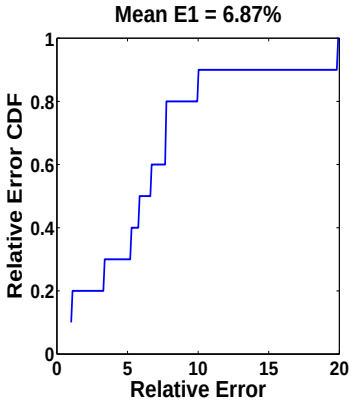
(a) Finite difference approximation $dOx_{rl_1}/dNOx_{sl_3}$ at 00:00 GMT 07/03/2006.



(b) Scatter plot of finite difference approximation versus adjoints.



(c) Adjoint $dOx_{rl_1}/dNOx_{sl_3}$ at 00:00 GMT 07/01/2006.



(d) Finite difference approximation versus adjoints relative error cumulative distribution function plot.

Figure 2.12: Comparison between adjoint and central finite difference approximation generated by running GEOS-Chem v7 model adjoint with all the processes for 2 days from 00:00 GMT 07/01/2006 to 00:00 GMT 07/03/2006, for Ox with respect to NOx concentrations for a set of carefully chosen receptors. rl_1 and sl_3 represent the receptor locations at ground level and starting locations at level 3 respectively.

Figures 2.12(a) and 2.12(c) showcase the receptor and starting locations. Adjoint variable for Ox tracer at ground level were initialized with unit concentration at ten receptor lo-

cations evenly spaced by 20 grid points in both longitude and latitude directions starting at grid numbers 5 and 11 respectively. Full model adjoint was applied to integrate the adjoint vector backwards to the start of the simulation 00:00 GMT July 1, 2006. Looking at the spread in the NOx adjoint variable at level 3 at this time, 10 starting locations were selected for which central finite difference calculations were performed. Figures 2.12(b) and 2.12(d) provide a direct comparison of adjoint values with their finite difference approximations at the ten locations. The mean relative difference between the two entities is 6.87% where all the points were within 20% of their relative difference. The lower order of accuracy for full model adjoint falls on the continuous adjoint for advection and the inability to determine exactly the sensitive locations in the flux spread. However, considering that all the science process adjoints showed good agreement when compared with their finite difference approximations and full model adjoint is within the permissible level of mismatch, we would like to conclude that we have successfully implemented an adjoint of GEOS-Chem v7.

2.12 Validation and Performance of the CMAQ Adjoint Model

This section presents the validation of the CMAQ adjoint by comparisons against finite differences, and by analyzing the computed adjoint solution.

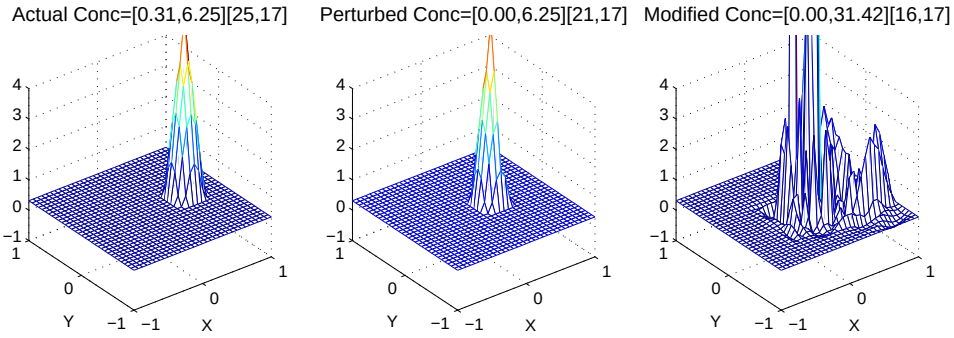
The Models-3 Community Multiscale Air Quality (CMAQ) is a multi-pollutant, multi-scale air quality model that contains state-of-science techniques for simulating all atmospheric and land processes that affect the transport, transformation and deposition of atmospheric pollutants and/or their precursors on both regional and urban scales [Byun and Ching , 1999].

The adjoint of CMAQv4.3 was developed by Amir Hakami and of CMAQv4.5 was developed at Virginia Tech. The CMAQv4.5 adjoint was designed to carry out data assimilation and sensitivity studies with respect to tracers, emissions and boundary conditions using newer algorithms and more recent data. CMAQv4.3 adjoint was built on an older version but also included aerosol adjoints. Later an updated adjoint code was released that encompassed capabilities of both the versions. We will be discussing here only about the contributions made to the CMAQv4.5 developments.

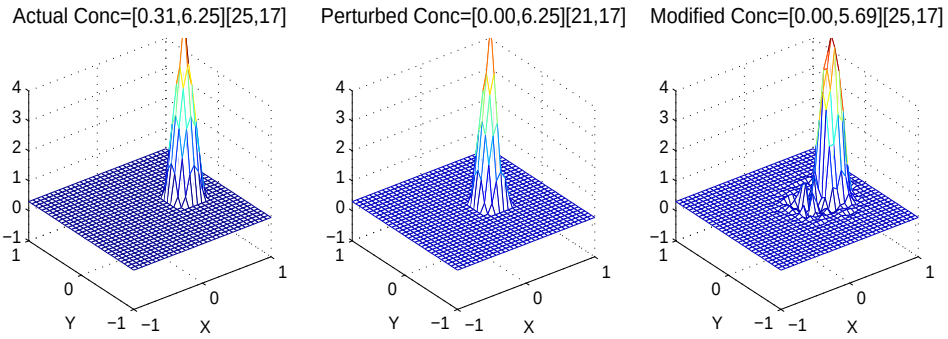
The underlying governing equation for CMAQ is defined exactly by (2.1a). The adjoint code construction for CMAQ was similar to that of the GEOS-Chem except that we considered both the continuous (by hand) and discrete adjoints (through TAMC) for advection process. When tested against finite difference approximations for a small simulation interval of 8 hours, the agreement with the approximations was higher for discrete adjoints as compared to their continuous counterparts. However, before we

integrated the discrete adjoint of advection into the full model adjoint, we conducted a twin data assimilation experiment for both the versions. The term "twin" suggests that the assimilation routine will be using an artificial observation data. Data assimilation is discussed extensively in Chapter 3, however, we will be providing the definition here.

Variational data assimilation attempts to find the control variable values (e.g., initial conditions) which minimize the discrepancy between model forecast and observations; the minimization is subject to the governing dynamic equations, which are imposed as strong constraints (2.8). In other words, given the model forecast, the observations and the mismatch information in the form of cost function and gradients, variational data assimilation generates an initial analysis field that when evolved through the model, best fits the observation data.



(a) Profile recovery using **discrete** adjoint of advection.



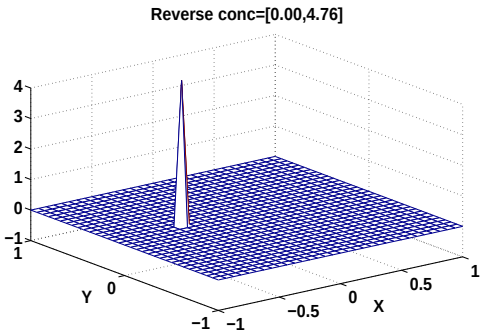
(b) Profile recovery using **continuous** adjoint of advection.

Figure 2.13: Demonstration of the effects of discrete versus continuous adjoints of advection on data assimilation in the recovery of a perturbed cone profile. Total 12 l-BFGS optimization iterations were performed.

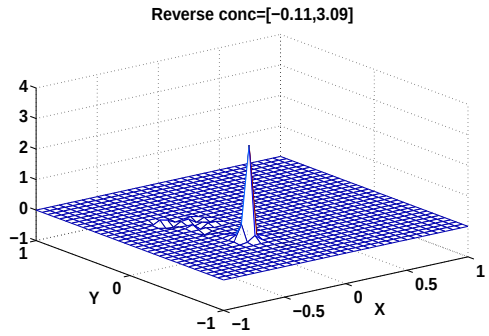
For our twin experiment set up, we considered a three dimensional cone profile with peak amplitude value of 6.25 measured with respect to the Z=0 X-Y plane. The base of the cone was located on the Z=0.31 plane. The ranges [-1,1] and [-1,1] on X and Y axes were divided into 32x32 grid with the peak of the cone located at (25,17) with respect

to [-1,-1] point on X-Y coordinate. This cone profile was treated as the artificial observation. A perturbation was then introduced in the original profile shifting the location of the peak to (21,17) and the base to Z=0 plane. This perturbed profile was considered the initial forecast profile. The optimization routine used was limited memory bound-constrained BFGS [Zhu et al. , 1994] which was supplied with the initial forecast profile, the artificial observation, and the cost function and gradient information calculated using advection adjoint only.

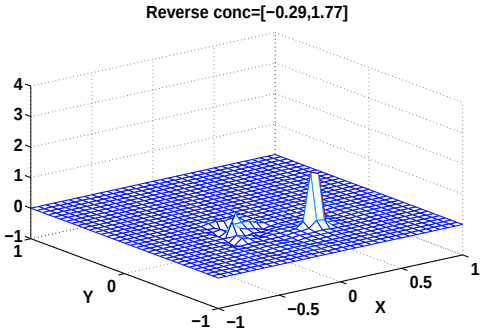
Figure 2.13 showcases the artifacts in assimilation due to the discrete adjoint of full monotonicity constrained Piecewise Parabolic Method. While Figure 2.13(b) shows that the assimilation using continuous adjoint recovered the original cone profile almost perfectly, in the discrete adjoint case in Figure 2.13(a), the oscillations at the lower boundaries of the cone due to discontinuities being differentiated paralyzes the convergence of the optimization routine. This reasoning was deduced after we conducted another simple experiment where we started with a cone profile with base at Z=0 plane and peak amplitude 4.76 located at (9,14), and then let it evolve in time using discrete advection adjoint routine for a period of 24 hours.



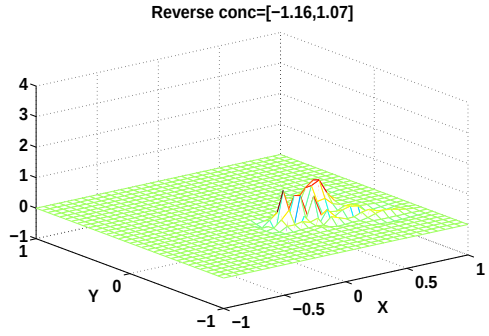
(a) Profile at the initial time.



(b) Profile at the end of 8 hours.



(c) Profile at the end of 16 hours.



(d) Profile at the end of 24 hours.

Figure 2.14: Evolution of a cone profile backwards in time using discrete adjoint of advection in CMAQ over a period of 24 hours.

Figure 2.14 shows that the discrete adjoint causes formation of spikes at the cone bound-

aries by the end of 8 hours and starts overwhelming the actual profile by the end of 24 hours. Earlier, when the validation was conducted only for 8 hours, this problem was unnoticeable, however it surfaced in the data assimilation since the gradients generated through discrete adjoints contained some erroneous values. To avoid such situations, we performed longer simulations to validate the adjoints: 24 hours for regional CMAQ model and 2 to 6 days in global GEOS-Chem model.

CMAQv4.5 forward model code comprises of vertical advection (Z direction), horizontal advections directions (X-Y directions), vertical diffusion, horizontal diffusion and chemistry. The underlying mechanism for advection and diffusion remain the same in both horizontal and vertical directions. Tables 2.3 and 2.4 present the call flow of science processes in CMAQ forward and adjoint model. The decouple and couple procedures in the adjoint mode are applied inside the horizontal and vertical diffusion adjoint routines respectively and hence do not appear in the adjoint call flow.

Table 2.3: SCIPROC Subroutines

Subroutine	**Purpose**
VDIFF()	vertical diffusion
COUPLE()	coupling concentrations for mass conserved processes
XADV()	horizontal advection in X-direction
YADV()	horizontal advection in Y-direction
ZADV()	vertical advection
HDIFF()	horizontal diffusion
DECOUPLE()	decoupling mass conserved concentrations
CHEM()	performs chemical operations

Table 2.4: SCIPROC_ADJ Subroutines

Subroutine	**Purpose**
CHEM_ADJ()	chemical process discrete adjoint
HDIFF_ADJ()	horizontal diffusion discrete adjoint
ZADV_ADJ()	vertical advection continuous adjoint
YADV_CAD()	H-advection cont adjoint in Y-direction
XADV_CAD()	H-advection cont adjoint in X-direction
VDIFF_ADJ()	vertical diffusion discrete adjoint

In order to test the accuracy of the adjoint of CMAQ model, we use central finite difference validation procedure similar to GEOS-Chem adjoint validation case. However, we transform the equation (2.32) into a relatively simpler form to provide scalar quantities

for comparison. Essentially we use

$$\mathcal{J}(c^0 + \delta c^0) - \mathcal{J}(c^0 - \delta c^0) = (\lambda^0)^{\mathrm{T}} \cdot (2\delta c^0) \tag{2.53}$$

where $(\lambda^0)^{\mathrm{T}} \cdot (2\delta c^0)$ is the dot product between the adjoint and the perturbation vector at the initial time, resulting in a scalar quantity.

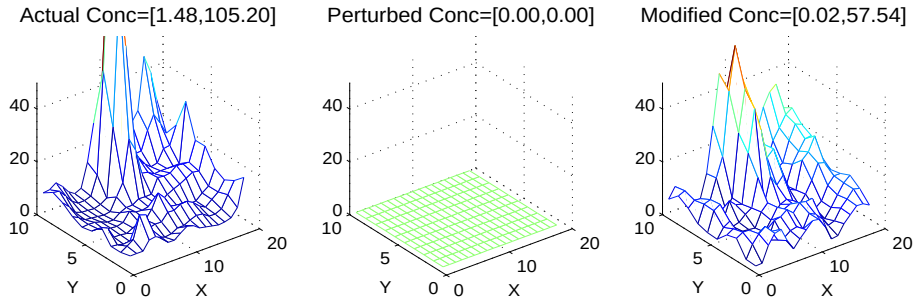
As presented in Table 2.4, we have implemented discrete adjoint for chemistry and diffusion while continuous adjoint for advection. Provided in Table 2.5 are the validation results for individual science processes and full model adjoint of CMAQ. The results suggest that the adjoint model for CMAQ is quite accurate. Except for horizontal advection, adjoints for all the science processes are within 1% of relative error. The continuous adjoint of advection does very well in the vertical direction. Overall relative error of 4.3% is within the admissible range to carry out data assimilation tests.

Table 2.5: CMAQ Adjoint Validation Results

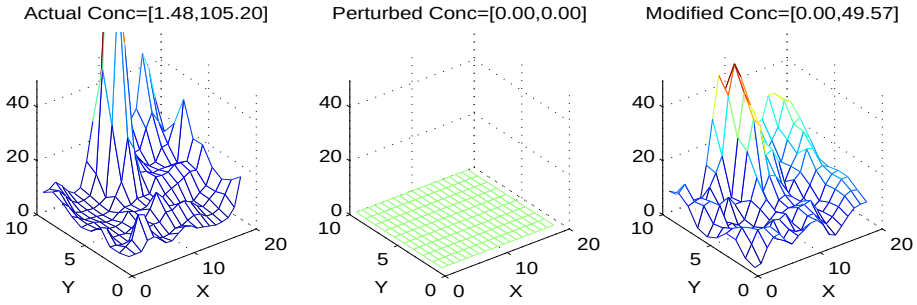
Process	$\mathcal{J}(c^0 + \delta c^0) - \mathcal{J}(c^0)$	$(\lambda^0[m])^{\mathrm{T}} \cdot \delta c^0[m]$	Relative Difference
HDIFF	2.4512E-008	2.4620E-008	0.4376%
HADV	1.8438E-008	1.9746E-008	6.6265%
ZADV	2.4445E-008	2.4307E-008	-0.5701%
VDIFF	1.4462E-008	1.4547E-008	0.5837%
CHEM	1.5801E-008	1.5916E-008	0.7244%
ALL	6.4055E-007	6.1417E-007	-4.2955%

The above results are encouraging, however, the scalar verification technique as described by equation (2.53) does not provide information about the agreement among individual points. We would like to argue that if the overall relative difference percentage is low there can not be too many points with large differences. In addition, the chemistry adjoint has been generated through KPP; accuracy of which could be seen in Figure 2.2(b) and in many published articles [Sandu et al. , 2003, 2005a; Henze et al. , 2007; Hakami et al. , 2007]. The adjoint of advection has been tested separately through twin experiment approach, results of which are shown in Figure 2.13. We now provide a twin experiment test for vertical diffusion adjoint to confirm its accuracy.

Presented in Figure 2.15 are the assimilation results for both continuous and discrete adjoints of vertical diffusion. We first ran the forward CMAQ model for 1 hour to generate a plume of concentration in a 10km x 20Km grid represented by "Actual Conc". This profile is treated as artificial observation and the initial condition ("Perturbed Conc") is taken to be a zero vector. These informations accompanied with the cost function and its gradient values calculated through vertical diffusion adjoint are fed to the l-BFGS subroutine which generates an analysis profile that best fits the observation data. It is evident from the plots that both continuous and discrete adjoints of vertical diffusion



(a) Profile recovery using **discrete** adjoint of vertical diffusion.



(b) Profile recovery using **continuous** adjoint of vertical diffusion.

Figure 2.15: Demonstration of the effects of discrete versus continuous adjoints of vertical diffusion on data assimilation in the recovery of a perturbed initial condition. Total 12 l-BFGS optimization iterations were performed.

work well, with discrete adjoint performing better. Hence, we have implemented the discrete adjoint for diffusion in the CMAQv4.5 adjoint. With all the validation results provided above showcasing the accuracy of the developed adjoint, we would like to conclude that we have successfully implemented an adjoint of CMAQv4.5.

2.13 Adjoint Sensitivity Analysis

This section presents sensitivity analysis results for test cases based on real data.

Forward sensitivity analysis is a source-oriented approach, that is, a perturbation in a certain parameter at the initial time is propagated forward in time. On the other hand the adjoint sensitivity analysis approach is receptor-oriented. The causes of a perturbation in an output variable or metric are traced backward in time. Therefore, adjoint analysis is efficient in calculating sensitivities of a few output variable or metrics with respect to a large number of (input) parameters. To demonstrate this, let us rewrite equation (2.26)

as follows

$$\lambda_i^0 = \frac{\partial \mathcal{J}(c^0)}{\partial p_i}, \quad 1 \leq i \leq m \tag{2.54}$$

where m is the total number of parameters including all the grid points and number of species. λ_i^0 is the i^{th} component of the adjoint vector at initial time and represents the gradient of the cost function with respect to parameter p_i. For simplicity, we consider these parameters to be the initial conditions $p = c^0$. Applying chain rule, the adjoint sensitivity of the cost function with respect to each parameter can be given by

$$\lambda^0 = \left(\frac{\partial c^1}{\partial c^0}\right)^{\mathrm{T}} \cdot \left(\frac{\partial c^2}{\partial c^1}\right)^{\mathrm{T}} \cdots \left(\frac{\partial c^N}{\partial c^{N-1}}\right)^{\mathrm{T}} \cdot \frac{\partial \mathcal{J}(c^0)}{\partial c^N}, \quad \left(\frac{\partial c^n}{\partial c^{n-1}}\right)^{\mathrm{T}} = \mathcal{M}'^{*}_{t^n \rightarrow t^{n-1}} \tag{2.55}$$

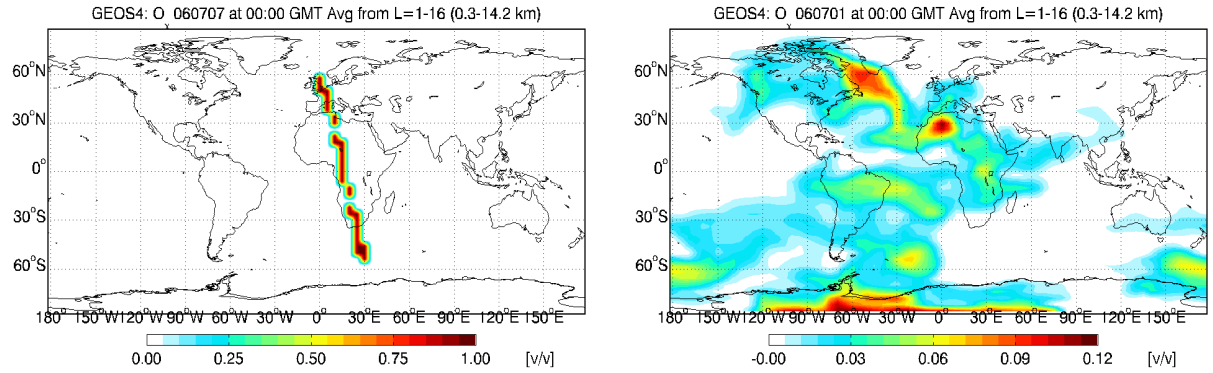
Thus, if we define a cost function at the final time, the gradients of this cost function with respect to concentration values at the final time is multiplied successively from right to left by $\mathcal{M}'^{*}_{t^n \rightarrow t^{n-1}}, n = N, N-1, \ldots, 1$. Each matrix-vector multiplication corresponds to one step of the adjoint model. It is important to note that, regardless of the number of parameters m, the above process needs to performed only once and is therefore very efficient.

Ozone is an important constituent of Earth's atmosphere with its abundance in the upper troposphere stratosphere providing a protective shield from harmful which absorbs the high energy UV-B and UV-C rays, thus preventing the disintegration of DNA molecules and supporting the existence of life. However, ozone present in mid to low troposphere is a pollutant, a powerful oxidizing agent leading to destruction of tissues, damaging fibers and creating breathing problems. We study here the sources and sinks of ozone precursors that lead to a certain amount of ozone over a particular region at a given time. In particular we study the sensitivity of ozone with respect to tracers, total emissions and dry depositions and individual emissions on a global scale in case of GEOS-Chem and with respect to tracers on a regional scale in case of CMAQ. Considering the efficiency of the adjoint sensitivity, all these informations are obtained in a single model adjoint run.

2.13.1 Sensitivity analysis with GEOS-Chem

We consider a possible Tropospheric Emission Spectrometer (TES, `http://tes.jpl.nasa.gov`) satellite trajectory. A detailed description on TES is presented in Chapter 3. Here we use only the coordinates of the trajectory of this instrument, not the profile retrieval data. To conduct the sensitivity analysis study, we start with the GEOS-Chem model forecast of chemical state of the atmosphere at 00:00 GMT on July 1, 2006. The forward model was let evolve this initial concentration field in time for 6 days until 00:00 GMT, July 7, 2006. At this time, a simple cost function was considered in such a way that the adjoint variable vector (receptor) was zero for all the species over all grid

points except for ozone which was initialized with unit concentration (in volume mixing ratio, [v/v]) underneath the trajectory from ground level up to 16 levels (14Km), which is the average height of troposphere. A snapshot of the adjoint variable vector at 00:00 GMT on July 7, 2006 is presented in Figure 2.16(a). Model adjoint was then applied on this adjoint vector backwards in time disintegrating each species into its precursors following the chain rule equation (2.55). We present in Figure 2.16(b), the state of adjoint variable for ozone at 00:00 GMT on July 1, 2006. This plot shows the global distribution of changes in ozone concentrations on July 1, 2006 that contributed to 1 [v/v] of change in ozone on July 7, 2006.



(a) dOx/dOx at 00:00 GMT, July 7, 2006.

(b) dOx/dOx at 00:00 GMT, July 1, 2006.

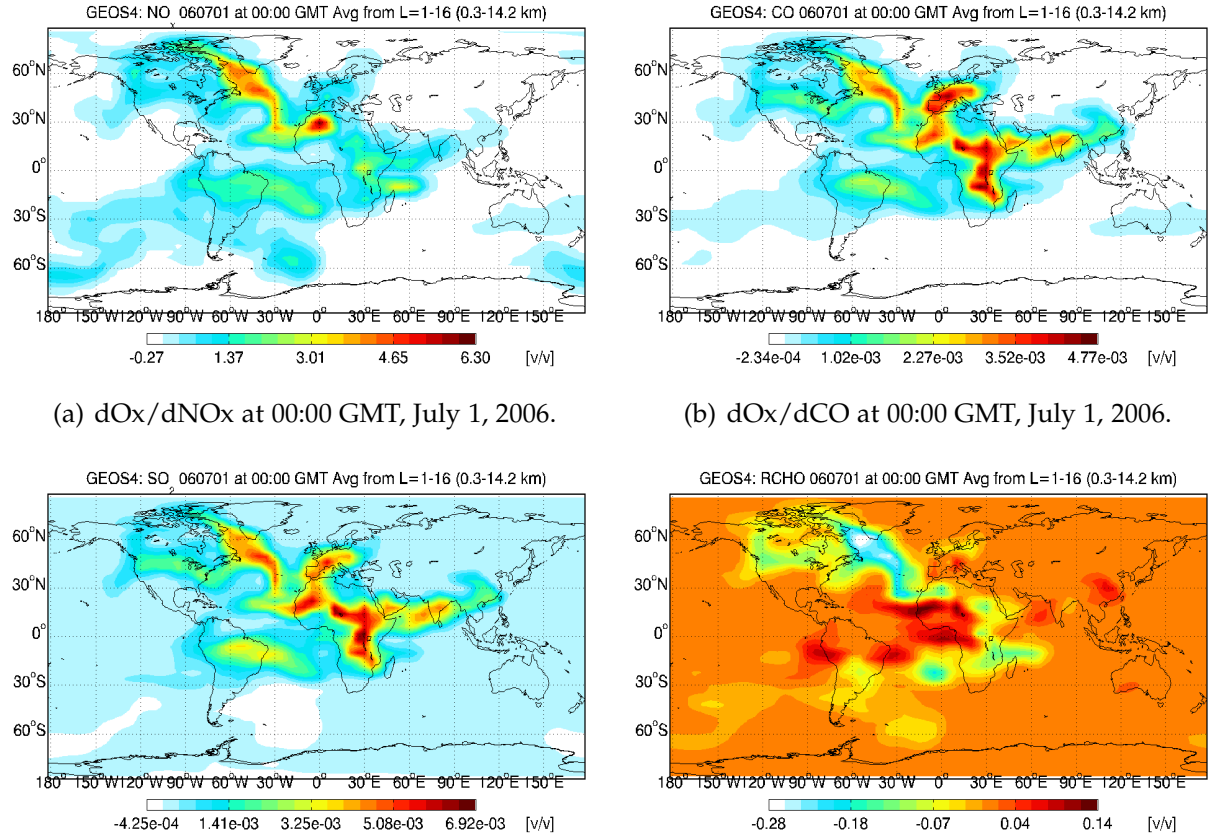
Figure 2.16: Sensitivity of ozone with respect to itself from a 6-day backward integration of ozone adjoint variable initialized with unit concentration at 00:00 GMT on July 7, 2006 underneath a possible TES trajectory upto 16 GEOS-Chem levels.

It is interesting to note that in Figure 2.16(b), the southern polar region contributed majorly to our sensitivity measurements. The reason could be attributed to the fact that the troposphere is quite thin in these regions and the abundant stratospheric ozone might be playing a role here.

Tracer Concentrations

We next examine the sensitivity of ozone with respect to its trace gas precursors mainly nitrogen/nitric oxide (NOx), carbon monoxide (CO), sulfur dioxide (SO2) and aldehyde (R-CHO).

Figure 2.17(a) reflects that changes in NOx concentration on July 1, 2006 primarily over north eastern Africa and North Atlantic oceanic region contributed to 1 [v/v] of change in Ox concentrations on July 7, 2006. Influence areas showcasing the changes in carbon monoxide and sulfur dioxide concentrations as presented in Figures 2.17(b) and 2.17(c)



(a) dOx/dNOx at 00:00 GMT, July 1, 2006.

(b) dOx/dCO at 00:00 GMT, July 1, 2006.

(c) dOx/dSO2 at 00:00 GMT, July 1, 2006.

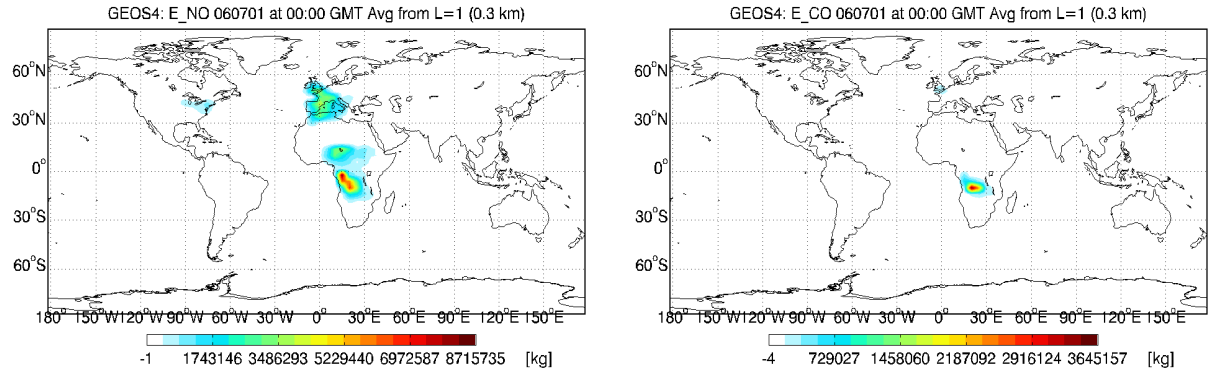
(d) dOx/dRCHO at 00:00 GMT, July 1, 2006.

Figure 2.17: Sensitivity of ozone with respect to trace gas precursors nitrogen oxide, carbon monoxide, sulfur dioxide and aldehyde from a 6-day backward integration of ozone adjoint variable initialized with unit concentration at 00:00 GMT on July 7, 2006 underneath a possible TES trajectory upto 16 GEOS-Chem levels.

respectively, have similar profiles with major contributions originating in North Atlantic Ocean, north east and mid Africa, western Europe, Middle East, and India. The aldehyde case is interesting where both negative and positive changes are substantial. While a dip in the aldehyde concentrations were noticed over regions including North Atlantic Ocean, South Atlantic Ocean, mid and southwest Africa and Indian Ocean, a rise in its concentrations were seen over north east and west South America, mid Atlantic Ocean, western Europe, Indian Ocean and as far as western China and south Australia.

Total Emissions

Emissions are major natural and man made sources of pollutants that are injected into the atmosphere influencing its chemical constituency. Though emissions heavily impact the local regional atmospheres creating health hazards for life forms using it, due to horizontal and vertical transport processes and complex chemical reactions, the impact could be noticed across continents as well. We study here the changes in total emissions at the ground level mainly of nitric oxide and carbon monoxide that influence the tropospheric ozone concentrations.



(a) dOx/dE_{NO} at 00:00 GMT, July 1, 2006.

(b) dOx/dE_{CO} at 00:00 GMT, July 1, 2006.

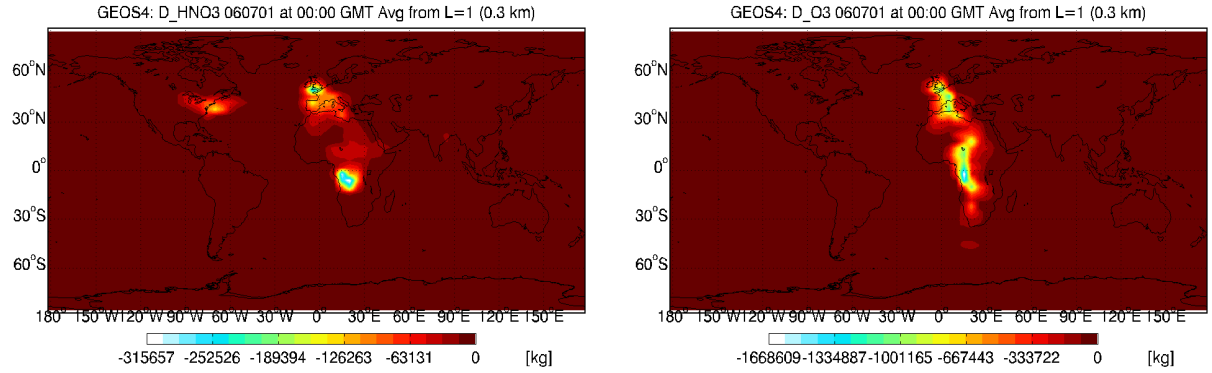
Figure 2.18: Sensitivity of tropospheric ozone with respect to total nitric oxide and carbon monoxide emissions at ground level from a 6-day backward integration of ozone adjoint variable initialized with unit concentration at 00:00 GMT on July 7, 2006 underneath a possible TES trajectory upto 16 GEOS-Chem levels.

Presented in Figures 2.18(a) and 2.18(b) are the location of the sources and changes in the amount of total NOx and CO emissions on July 1, 2006 contributing to 1 [v/v] change in tropospheric ozone on July 7, 2006. The CO emission sources were located mainly in Congo, Zaire, Zambia and Angola in Africa and United Kingdom in Europe, while NOx emission sources were located at most of the western Europe, north and midwest Africa including most of the CO emission locations and as far as east coast of United States. The reason for such high emission rates in Africa are contributed to the biomass and biofuel being used as the main source of energy. It will be visible clearly when we further disintegrate these total emissions into their individual sources.

Notice the large emission values in the colorbars as compared to the tracer concentrations. The reason is that emission adjoints are calculated in [kg] by multiplying the scaled emission sensitivities with the respective emission rates in [kg/s], while tracer concentrations are presented in volume mixing ratios [v/v].

Total Dry Depositions

We next study about the impact of perturbations in an important natural process which in conjunction with wet deposition helps regulate the concentration of pollutants by sedimenting them. We consider primarily the changes in the dry deposition of a water soluble species: nitric acid (HNO3) and a powerful oxidant: ozone (O3).



(a) dOx/dD_{HNO3} at 00:00 GMT, July 1, 2006.

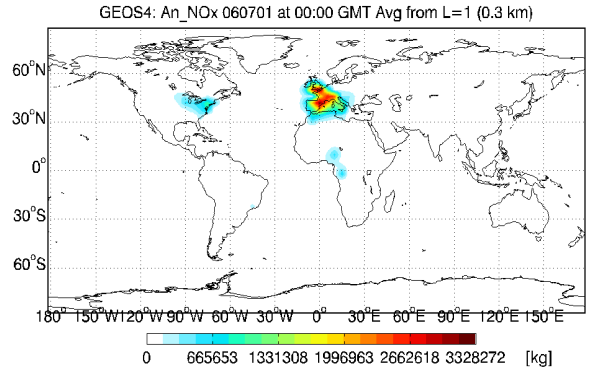
(b) dOx/dD_{O3} at 00:00 GMT, July 1, 2006.

Figure 2.19: Sensitivity of tropospheric ozone with respect to total nitric acid and ozone dry depositions at ground level from a 6-day backward integration of ozone adjoint variable initialized with unit concentration at 00:00 GMT on July 7, 2006 underneath a possible TES trajectory upto 16 GEOS-Chem levels.

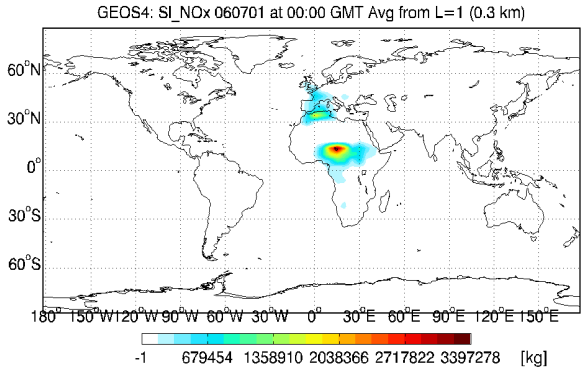
Figures 2.19(a) and 2.19(b) present the sink locations and the amount of change in the HNO3 and O3 being deposited on July 1, 2006 that contributed to 1 [v/v] change in the ozone concentrations after 6 days. As noted in the science encyclopedia `http://science.jrank.org`, the wet deposition is a slow process, however, dry deposition rates could be as large as point source emissions. The negative values of concentrations in deposition case represent the exit of depositing species from the atmosphere and entering the ecosystem depending on three important factors: meteorological variables (wind speed, temperature, atmospheric stability, and humidity), surface variables (surface aerodynamic roughness and structure, pH, surface charge, hydrophobicity, porosity), and properties of the depositing material (chemical reactivity, solubility, diameter, surface charge, and shape) [Cohen , 1998]. Large amount of dry deposition along the line of trajectory indicates that it a sensitive sink area. However, more interesting is to notice the cross continental sink for HNO3 located at North Atlantic east coast of the United States.

Individual Source Emissions

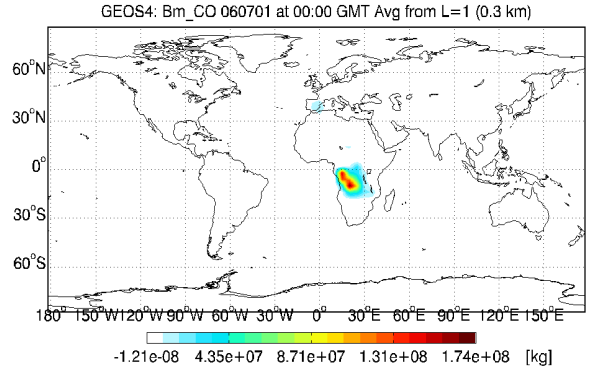
Individual source emissions are most insightful of all the previously presented sensitivities. The ability of GEOS-Chem adjoint to provide this information opens new doors to conduct research in this field and make policies to regulate individual sources especially those that are influenced by human activities. We consider four important sources of emissions: anthropogenic NOx, soil NOx, biomass burning CO and biofuel burning CO.



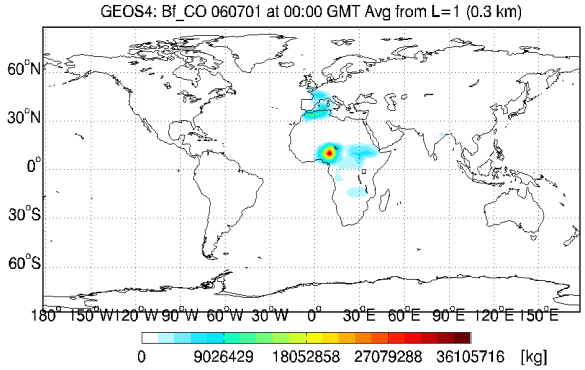
(a) dOx/dAn_{NOx} at 00:00 GMT, July 1, 2006.



(b) dOx/dSl_{NOx} at 00:00 GMT, July 1, 2006.



(c) dOx/dBm_{CO} at 00:00 GMT, July 1, 2006.



(d) dOx/dBf_{CO} at 00:00 GMT, July 1, 2006.

Figure 2.20: Sensitivity of tropospheric ozone with respect to anthropogenic NOx, soil NOx, biomass burning CO and biofuel burning CO emissions at ground level from a 6-day backward integration of ozone adjoint variable initialized with unit concentration at 00:00 GMT on July 7, 2006 underneath a possible TES trajectory upto 16 GEOS-Chem levels.

We first take a look at the anthropogenic NOx emission sources, which in GEOS-Chem are mostly accounted for burning fossil fuel. Presented in Figure 2.20(a), the anthropogenic NOx emission sources that contributed to 1 [v/v] change in Ozone on July 7,

2006 along TES trajectory are located in midwest Africa, western Europe, and as far as east coast of United States. The western Europe mainly Spain, France, Austria, Germany and UK contributed heavily, probably because of the fossil fuels being burned through automobiles and fuels used to provide electricity and heating for homes. An assessment report on NOx emissions published by European Environment Agency using 2007 dataset suggests that road transport contributed 33.4%, energy industries 34% and Agriculture 2%. This suggests that a massive 67% of total NOx emissions in Europe were due to fossile fuel burning. A similar reasoning could be attributed to the U.S. east coast. However, in case of Africa, the fossil fuel burning is majorly accredited to the oil spill. Nigeria and Cameroon suffered the worst land oil spill disaster accumulated over last 50 years. Due to unavailability of adequate equipments and dispersants, people simply burn the oil injecting several tons of NOx, CO, SO2 and many other hydrocarbons.

We next study the soil NOx emissions produced by soil microbes during nitrification and denitrification processes. This emission is controlled by a suite of environmental variables including inorganic nitrogen availability, water-filled pore space, and soil temperature, and could be significant in agricultural regions using nitrogen based fertilizers [Hall et al. , 1996]. Figure 2.20(b) reflects the amount and sources of soil NOx emissions with major contributions along the dry and humid savanna agricultural regions in Africa, while north Algeria, Tunisia in Africa and some western Europian countries also contributed to the cause. It is interesting to note that soil NOx emissions could be as large as anthropogenic NOx emissions.

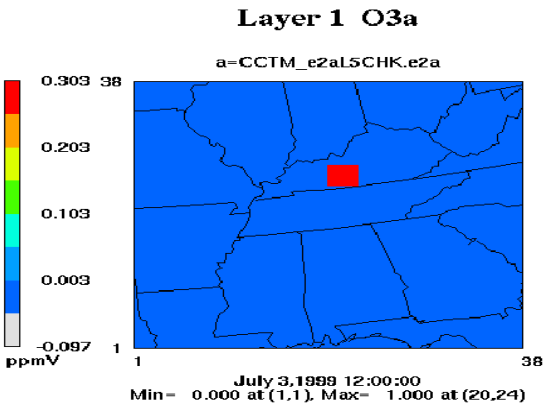
Figures 2.20(c) and 2.20(d) showcase the changes in CO emissions due to biomass and biofuel burning on July 1, 2006 at ground level that contributed to the changes in tropospheric ozone concentrations on July 7, 2006. Biomass burning includes burning of vegetation mostly human-initiated to clear land fields in developing countries as well as natural and lightning-induced fires. Biofuel burning is similar to biomass in that it includes processed fuel such as dried firewoods, agricultural residues, animal wastes, and charcoal in developing countries, while biodiesel, vegetable oil, ethanol, methane in developed countries. Not surprisingly, mid and southwest African regions contribute maximum to the biomass burning, while mid and noth African regions contribute highest to biofuel burning. A report released by the U.S. Energy Information Administration (`http://www.eia.doe.gov/emeu/cabs/chapter7.html`) suggests that Africa is the world's largest consumer of biofuel energy (firewood, agricultural residues, animal wastes, and charcoal), calculated as a percentage of overall energy consumption. On the other hand, due to lack of technology, forests and plane lands are simply burnt to create agricultural fields and create housing.

2.13.2 Sensitivity analysis with CMAQ

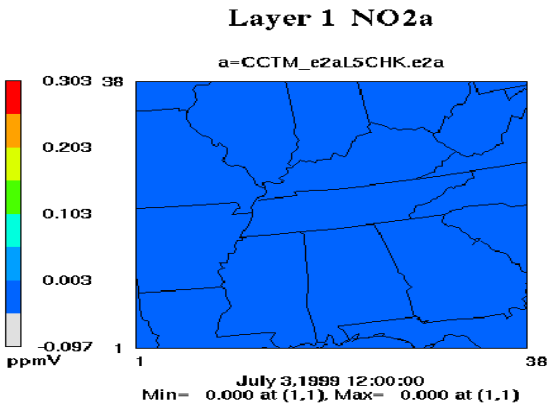
In this section, we will provide a simple demonstration of the adjoint sensitivity capability in CMAQ. In this test setup, a forward run with the original concentration values is performed for 36 hrs starting at 00:00 GMT on July 2, 1999. At the end of the forward mode, adjoint variable for all other species except O3 is set to zero at all the grid points, where adjoint variable for O3 is initialized at the ground level with unit concentration [ppmV] over a 2x2 grid in southern region of Kentucky, US. Snapshots of the adjoint variable for ozone and nitrogen dioxide are presented in Figures 2.21(a) and 2.21(b) respectively, while mathematically it could be represented as

$$\lambda^{final} = 0.0, \quad \lambda^{final}(20:22, 24:26, 1, O3) = 1.0 \tag{2.56}$$

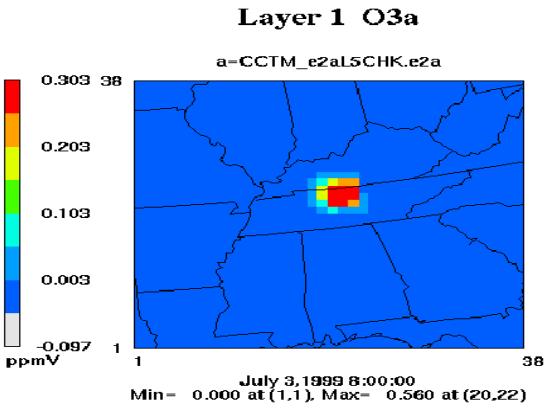
In case of GEOS-Chem, we presented the sensitivity of ozone with respect to several parameters at the initial time, however, here we will be presenting how the adjoint sensitivities evolve in time. Presented in Figure 2.21 are a series of plots that showcase the adjoint sensitivity of ozone with respect to itself and a gas-phase precursor NO2, over a period of 16 hours backwards in time. Figures 2.21(e),2.21(c) and 2.21(a) in that order, provide information on how changes in ozone concentrations at 20:00 GMT July 2, 08:00 GMT July 3 and 12:00 GMT July 3, 2006 contributed individually to a 1 [ppmV] change in the ozone concentrations at 12:00 GMT on July 3, 2006. Similarly, Figures 2.21(f),2.21(d) and 2.21(b) provide information on changes in NO2 concentrations at three different times showcasing their contributions. It is interesting to see how adjoint sensitivities of ozone with respect to nitrogen dioxide evolve in time during backward integration due to chemical adjoint process and spread due to adjoint of transport processes.



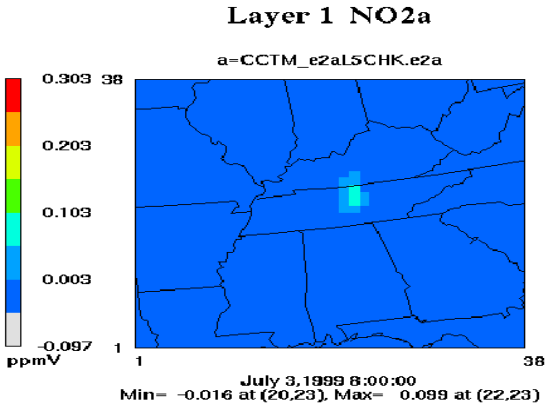
(a) dO3/dO3 at T=36hrs



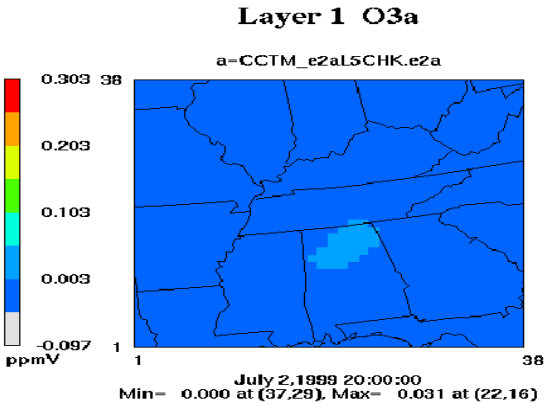
(b) dO3/dNO2 at T=36hrs



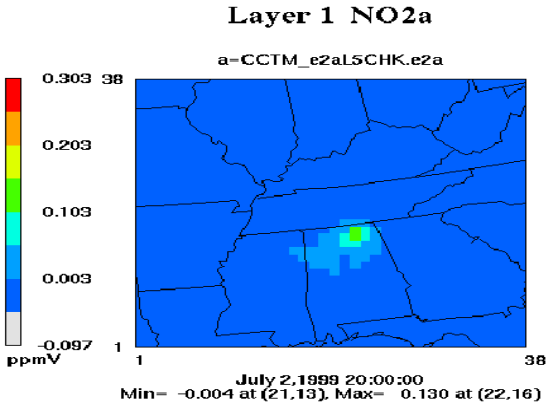
(c) dO3/dO3 at T=32hrs



(d) dO3/dNO2 at T=32hrs



(e) dO3/dO3 at T=20hrs



(f) dO3/dNO2 at T=20hrs

Figure 2.21: Sensitivity of ozone with respect to nitrogen dioxide trace gas and itself at ground level from a 16-hour backward integration of ozone adjoint variable initialized with unit concentration on 00:00 GMT July 2, 1999 over a 2x2 grid in southern Kentucky, USA.

2.14 Conclusions

Adjoint models are essential tools for sensitivity analysis studies and for providing gradients in the solution of inverse problems. The construction of an adjoint model is an extremely difficult, labor intensive, and error prone task.

This chapter describes the development of adjoints for two chemical transport models: Harvard's GEOS-Chem, the most widely used global CTM, and Environmental Protection Agency's regional CMAQ, the most widely used regional CTM.

We employ a variety of strategies for the construction of adjoint models: symbolic preprocessing, derivation of adjoint differential equations and their numerical solution, and automatic differentiation. This makes the resulting adjoint code truly hybrid. The chemical adjoint is discrete and implemented using symbolic preprocessing. The advection adjoint is continuous. Convection and wet-deposition adjoints are discrete and constructed using automatic differentiation software together with manual coding. Emission and dry deposition adjoints are treated as pseudo chemical equations. The stratosphere-troposphere exchange adjoints are implemented using forward source analysis and hand coding. A two level checkpointing scheme is used to balance the storage space and the extra time spent in repeated calculations. The adjoint model inherits the parallelization structure of the forward model.

The adjoint models are validated by comparing their solutions against finite difference approximations of the sensitivities. Adjoint sensitivity analysis results are shown for test cases based on real data.

The source codes of the adjoint models developed here are publicly available [`http://people.cs.vt.edu/~asandu/Public/GCv7_ADJ/`,`http://people.cs.vt.edu/~asandu/Software/CMAQ_ADJ/CMAQ_ADJ.html`] These adjoint frameworks are currently used by the GEOS-Chem and the CMAQ communities to perform sensitivity analyses and data assimilation for improved initial tracer concentrations, boundary conditions, and emission sources.

Chapter 3

Atmospheric Data Assimilation with GEOS-Chem: a Comparison Between Variational and Suboptimal Kalman Filter Approaches

Abstract

Chemical data assimilation is a process of optimally combining observations of reality with imperfect model predictions to produce a better estimate of the chemical state of the atmosphere. Models are imperfect and cannot capture exactly the complex dynamics of the atmosphere. As a result there is always a mismatch between model generated forecasts and reality. Data assimilation is the procedure to combined data from observations with model predictions to obtain a more accurate representation of the state of the atmosphere.

Two families of data assimilation methods are currently widely used: variational and Kalman filter (KF). The variational approach is based on control theory, and formulates data assimilation as a minimization problem of a cost functional that measures the mismatch between model prediction and observations. The Kalman filter approach is rooted in statistical estimation theory and provides the analysis covariance together with the best state estimate. Suboptimal Kalman filters employ different approximations of the covariances in order to make the computations feasible with large models. Each family of methods has both merits and drawbacks.

This chapter compares several data assimilation methods used for global chemical data assimilation. Specifically, improved estimates of summertime global ozone distribution are obtained through the assimilation of ozone profile retrievals from Tropospheric Emis-

sion Spectrometer (TES) into GEOS-Chem using variational and suboptimal Kalman filter (KF) approaches. The resulting analyses are compared against an independent data set provided by ozonesonde measurements to assess the effectiveness of each assimilation method. The smallest differences between mean analysis profiles and ozonesonde measurements are obtained with four dimensional variational data assimilation with window lengths between 5 days and 2 weeks. The two sequential assimilation approaches (three dimensional variational and suboptimal KF), although derived differently, provide similar ozone estimates.

3.1 Introduction

There has been a lot of interest in the study of tropospheric ozone distribution lately [Li et al., 2005; Cooper et al., 2006, 2007; Hudman et al., 2007] due to its hazardous properties. Despite attempts to quantify and characterize its distribution through sophisticated chemical transport and general circulation models [Horowitz et al., 2003; Horowitz, 2006; Stevenson et al., 2006], the findings from these studies vary significantly due to the strong variability in ozone lifetimes and uncertainties in determining the amount of ozone lost through dry deposition, entered through upper troposphere-stratosphere exchanges, or evolved due to chemical reactions of trace gas and emission precursors.

Data assimilation has been used to improve initial conditions, emissions, and boundary values. Besides the initial conditions, improvements in boundary values lead to improved air quality forecasts. Considerable experience with data assimilation have been accumulated in the field of numerical weather prediction [Daley, 1991; Courtier et al., 1998; Rabier et al., 2000; Kalnay, 2002; Navon, 2009]. In this work we focus on chemical data assimilation, i.e., on assimilation of observations of pollutant levels in the atmosphere. Chemical data assimilation poses specific challenges related to the multiphysics nature of the system, the stiffness of chemical kinetic equations, the sparseness of chemical observations, and the uncertainty in the levels of anthropogenic and natural pollutants emitted into the atmosphere.

Previous studies have employed various approaches to assimilating observations of trace gases for improved tropospheric chemistry representations. The base concepts of the variational approach to chemical data assimilation, and the construction of adjoint chemical transport models are discussed in detail in [Sandu et al. , 2005a; Carmichael et al., 2008]. 3D-Var was first used by [Derber et al., 1991; Parrish and Derber, 1992] and later applied by most of the meteorological centers [Courtier et al., 1998; Cohn et al., 1998; Gauthier et al., 1999a]. A study on ozone improvement using 3D-Var assimilation is presented in [Bei et al., 2008]. Adjustment of gas phase chemical tracer initial conditions has been studied in [Chai et al., 2007; Zhang et al., 2008]. Adjustment of pollutant emissions through 4D-Var chemical data assimilation has been discussed in [Chai et al.,

2009]. Data assimilation studies involving particle measurements to improve aerosol fields have been performed in [Hakami et al. , 2005; Henze et al., 2009]. Suboptimal Kalman filters have been employed successfully for chemical data assimilation [Menard et al., 2000; Lamarque et al., 2002; Segers et al., 2005; Clark et al., 2006; Pierce et al., 2007; Parrington et al., 2009]. The use of the ensemble Kalman filter (EnKF) in chemical data assimilation has been studied in [Constantinescu et al., 2007b,c].

A discussion on relationship between optimality of variational data assimilation and Kalman filters is presented in [Li and Navon, 2001]. [Houtekamer, 2005] compared the quality of background statistics in 3D-Var and EnKF using radiance observations from satellite, while, [Laroche et al., 2005] compared the characteristics 3D-Var and 4D-Var introduced in the operational suite of the Canadian Meteorological Center (CMC). [Constantinescu et al., 2007c] and [Wu et al., 2008] compare the performances of EnKF with 4D-Var for chemical transport models on a regional scale using ground-level ozone measurements, while, [Geer et al., 2006] provides an intercomparison of tropospheric ozone estimates obtained through 3D-Var, 4D-Var and Kalman filter assimilation systems for both chemical transport and global circulation models as part of the Assimilation of ENVIAST Data (ASSET) project.

The Tropospheric Emission Spectrometer (TES) [Beer et al., 2001] is the first dedicated infrared instrument from which information of the global and vertical distribution of tropospheric ozone can be retrieved. [Parrington et al., 2009] provided the first set of results from the assimilation of vertical profiles of ozone from TES into the GEOS-Chem using suboptimal Kalman filter. We have developed 3D-Var and 4D-Var data assimilation capabilities for GEOS-Chem v7. The goal of this chapter is to provide the first direct comparison of global tropospheric ozone distribution estimated through 3D-Var, 4D-Var and suboptimal KF assimilation systems showcasing the potential of TES profile retrievals. The assessment of analyses generated through different assimilation systems are on the similar lines of [Geer et al., 2006; Parrington et al., 2009].

This work is significantly different from [Wu et al., 2008] and implicitly addresses the shortcomings of this chapter. First of all we study the global ozone distribution assimilating satellite observations into GEOS-Chem as compared to the ozone forecasts over western Europe through assimilation of observations from ground stations into Polair3D [Boutahar et al., 2004]. We evaluate the quality of tropospheric ozone analyses estimated by our assimilation systems through ozonesonde measurement data, an independent observation dataset not used in assimilation, while [Wu et al., 2008] has a forecast scoring scheme, where the scores are calculated as root mean square of the differences between analyses and assimilated observations. Not surprisingly, their optimal interpolation (OI) scheme fetched the best overall score while 4D-Var performed average. Such a scoring scheme does not provide a fair assessment of the performance of an assimilation system since there is no provision to adjudge whether observations are erroneous. In addition, the average performance of their 4D-Var assimilation could be attributed to the limitations of their study as they adjust only the initial conditions while boundary conditions

remain unchanged, making the assimilation ineffective especially for long range simulations. We believe this plays a significant role in the decrease of their 4D-Var performance with increase in assimilation window lengths, although they have attributed it solely to the model errors. Such a situation does not arise in our case as global assimilations are not restricted by any horizontal boundaries.

This chapter is structured as follows. Section 3.2 provides the mathematical overview of how observations are integrated into the model in different data assimilation systems. Section 3.3 discusses the characteristics of background error covariance matrices used in this study. Section 3.4 provides a brief overview of the global chemical transport model (GEOS-Chem) and its adjoint development. A description of the TES instrument, its observation operator and profile retrieval formulation is provided in Section 3.5. Section 3.6 details the experimental settings, computational costs and assessment of tropospheric ozone estimates through different assimilation systems. Summary and points of future work are discussed in Section 3.7.

3.2 Chemical data assimilation

Data assimilation combines the following three sources of information.

1. The apriori, or background state $\mathbf{x}^b$ represents the best estimate of the true state $\mathbf{x}^t$ available before any measurements are taken. This estimate is assumed unbiased, and the random background (estimation) errors ε^b are typically assumed to have a normal probability density with a background error covariance matrix $\mathbb{B}$

$$\varepsilon^b = \mathbf{x}^b - \mathbf{x}^t \in \mathcal{N}(0, \mathbb{B}). \tag{3.1}$$

2. The model encapsulates our knowledge about physical and chemical laws that govern the evolution of the system. The model evolves an initial state $\mathbf{x}_0 \in \mathbb{R}^n$ at the initial time t_0 to future state values $\mathbf{x}_i \in \mathbb{R}^n$ at future times t_i,

$$\mathbf{x}_i = \mathcal{M}_{t_0 \to t_i}(\mathbf{x}_0). \tag{3.2}$$

   The size of the state space in realistic chemical transport models is very large. For example, a GEOS-Chem simulation at the $2^\circ x 2.5^\circ$ horizontal resolution has $n \in \mathcal{O}(10^8)$ variables.

3. Observations $\mathbf{x}_i^{\text{obs}} \in \mathbb{R}^m$ of the state are taken at times t_i, $1 = 1, \cdots, N$

$$\mathbf{x}_i^{\text{obs}} = \mathcal{H}(\mathbf{x}_i) + \varepsilon_i^{\text{obs}}. \tag{3.3}$$

   The observation operator $\mathcal{H}$ maps the state space onto the observation space. In many practical situations $\mathcal{H}$ is a highly nonlinear mapping (as is the case, e.g., with

satellite observation operators). Usually the observations are sparsely distributed, and the number of observations is small compared to the dimension of the state space, $m \ll n$.

The observations are corrupted by measurement and representativeness errors $\varepsilon_i^{\text{obs}}$. The observation errors at each time are assumed to be independent of background errors, and independent of the observation errors at other times. They are typically assumed to have a normal distribution with mean zero and covariance $\mathbb{R}_i$,

$$\varepsilon_i^{\text{obs}} \in \mathcal{N}\left(0, \mathbb{R}_i\right). \tag{3.4}$$

Based on these three sources of information data assimilation computes the posterior estimate $\mathbf{x}^a$ of the true state; $\mathbf{x}^a$ is called the "analysis".

Variational methods solve the data assimilation problem in an optimal control framework [Sasaki, 1958; LeDimet and Talagrand, 1986; Courtier and Talagrand , 1987; Lions, 1971]. Specifically, they attempt to find the control variable values (e.g., initial conditions) which minimize the discrepancy between the model forecast and observations; the minimization is constrained by the governing dynamic equations. In this discussion, for simplicity of presentation, we focus on discrete models (in time and space) where the initial conditions are the control variables.

3.2.1 Three dimensional variational (3D-Var) data assimilation

In the 3D-Var data assimilation the observations (6.3) are considered successively at times $t_1, \cdots, t_N$. The background state (i.e., the best state estimate at time t_i) is given by the model forecast, starting from the previous analysis (i.e., best estimate at time t_{i-1}):

$$\mathbf{x}_i^b = \mathcal{M}_{t_{i-1} \to t_i}\left(\mathbf{x}_{i-1}^a\right).$$

The discrepancy between the model state $\mathbf{x}_i$ and observations at time t_i, together with the departure of the state from the model forecast $\mathbf{x}_i^b$, are measured by the 3D-Var cost function:

$$\mathcal{J}\left(\mathbf{x}_i\right) = \frac{1}{2}\left(\mathbf{x}_i - \mathbf{x}_i^b\right)^T \mathbb{B}^{-1}\left(\mathbf{x}_i - \mathbf{x}_i^b\right) + \frac{1}{2}\left(\mathcal{H}(\mathbf{x}_i) - \mathbf{x}_i^{obs}\right)^T \mathbb{R}_i^{-1}\left(\mathcal{H}(\mathbf{x}_i) - \mathbf{x}_i^{obs}\right) \tag{3.5}$$

While in principle a different background covariance matrix should be used at each time, in practice the same matrix is re-used throughout the assimilation window. The 3D-Var analysis is computed as the state which minimizes (6.8)

$$\mathbf{x}_i^a = \arg\min\ \mathcal{J}\left(\mathbf{x}_i\right). \tag{3.6}$$

Typically a gradient-based numerical optimization procedure is employed to solve (6.7). The gradient $\nabla\mathcal{J}$ of the cost function (6.8) is

$$\nabla\mathcal{J}\left(\mathbf{x}_i\right) = \mathbb{B}^{-1}\left(\mathbf{x}_i - \mathbf{x}_i^b\right) + \left(\mathcal{H}'(\mathbf{x}_i)\right)^T \mathbb{R}_i^{-1}\left(\mathcal{H}(\mathbf{x}_i) - \mathbf{x}_i^{obs}\right) \tag{3.7}$$

Note that the gradient requires to computation of the linearized observation operator $\mathcal{H}'$ about the current state.

Preconditioning is often used to improve convergence of the numerical optimization problem (6.7). A change of variables is performed, for example, by shifting the state and scaling it with the square root of covariance:

$$\widehat{\mathbf{x}}_i = \mathbb{B}^{1/2} \left(\mathbf{x}_i - \mathbf{x}_i^b \right) , \tag{3.8}$$

The optimization is then carried out on the new variables $\widehat{\mathbf{x}}_i$.

3.2.2 Four dimensional variational (4D-Var) data assimilation

In strongly-constrained 4D-Var data assimilation all observations (6.3) at all times $t_1, \cdots, t_N$ are simultaneously considered. The control parameters are the initial conditions $\mathbf{x}_0$; they uniquely determine the state of the system at all future times via the model equation (6.6).

The discrepancy between model predictions and observations at all future times $t_1, \cdots, t_N$, together with the departure of the initial state from the background state, are measured by the 4D-var cost function:

$$\mathcal{J}\left(\mathbf{x}_0\right) = \frac{1}{2} \left(\mathbf{x}_0 - \mathbf{x}_0^b \right)^T \mathbb{B}^{-1} \left(\mathbf{x}_0 - \mathbf{x}_0^b \right) + \frac{1}{2} \sum_{i=1}^{N} \left(\mathcal{H}(\mathbf{x}_i) - \mathbf{x}_i^{obs} \right)^T \mathbb{R}_i^{-1} \left(\mathcal{H}(\mathbf{x}_i) - \mathbf{x}_i^{obs} \right) \tag{3.9}$$

Note that the departure of the initial conditions from the background is weighted by the inverse background covariance matrix, $\mathbb{B}^{-1}$, while the differences between the model predictions $\mathcal{H}(\mathbf{x}_i)$ and observations $\mathbf{x}_i^{obs}$ are weighted by the inverse observation error covariances, $\mathbb{R}_i^{-1}$.

The 4D-Var analysis is computed as the initial condition which minimizes (6.10) subject to the model equation constraints (6.6)

$$\mathbf{x}_0^a = \arg\min \ \mathcal{J}\left(\mathbf{x}_0\right) \quad \text{subject to} (6.6). \tag{3.10}$$

The model (6.6) propagates the optimal initial condition (6.10) forward in time to provide the analysis at future times, $\mathbf{x}_i^a = \mathcal{M}_{t_0 \to t_i} \left(\mathbf{x}_0^a\right)$.

The optimization problem (6.11) is solved numerically using a gradient-based technique. The gradient of (6.10) reads

$$\nabla \mathcal{J}\left(\mathbf{x}_0\right) = \mathbb{B}^{-1} \left(\mathbf{x}_0 - \mathbf{x}_0^b \right) + \sum_{i=1}^{N} \left(\frac{\partial \mathbf{x}_i}{\partial \mathbf{x}_0} \right)^T \left(\mathcal{H}'(\mathbf{x}_i) \right)^T \mathbb{R}_i^{-1} \left(\mathcal{H}(\mathbf{x}_i) - \mathbf{x}_i^{obs} \right) \tag{3.11}$$

The 4D-Var gradient requires not only the linearized observation operator $\mathcal{H}'$, but also the transposed derivative of future states with respect to the initial conditions. The 4D-Var gradient can be obtained effectively by forcing the adjoint model with observation increments, and running it backwards in time. The construction of an adjoint model requires considerable effort, time, and know-how.

3.2.3 Suboptimal Kalman filter

The suboptimal Kalman filter is a sequential data assimilation approach [Khattatov et al., 2000] in which corrections in the concentration state vector are performed as soon as observations become available. Similar to 3D-Var, for every observation window starting with a model forecast state (x^f), this technique provides an expected analysis state (x^a) that reduces the discrepancy between model predictions and what is observed. The generated analysis state vector could be expressed as

$$\log \mathbf{x}^a = \log \mathbf{x}^f + K\left(\widehat{\mathbf{z}} - \mathcal{H}\left(\mathbf{x}^f\right)\right) \tag{3.12}$$

where K is the Kalman gain matrix, $\mathcal{H}$ is the observation operator defined in equation (6.3), and $\hat{z}$ is the ozone profile retrievals from TES as described in equation (3.20). The analysis state is calculated in natural logarithm of volume mixing ratio (log VMR) at each observation grid point since the TES profile retrievals are in log VMR. An exponential operator and a linear interpolation operator based on pressure is then applied to this logarithm of analysis state in succession to regain the actual analysis state in GEOS-Chem grid domain. The points which do not lie on the observation grid remain unaffected by the assimilation.

The observation operator $\mathcal{H}$ that transforms higher resolution model state to the TES profile vertical grid (observation grid) domain is expressed by equation (3.21). The Kalman gain matrix K is defined as

$$K = \mathbb{P}^f \mathcal{H}^{\mathrm{T}} \left(\mathcal{H}\mathbb{P}^f \mathcal{H}^{\mathrm{T}} + \mathbb{R}\right)^{-1} \tag{3.13}$$

where $\mathbb{P}^f$ is the forecast error covariance matrix and $\mathbb{R}$ is the observation error covariance matrix associated with the TES profile retrievals defined in equation (3.4). If a diagonal or block-diagonal forecast error covariance matrix $\mathbb{P}^f$ is used in equation (3.13), the analysis state generated through equation (3.12) is suboptimal. A description on the structure of $\mathbb{P}^f$ is provided in Section 3.3.

At each observation window, along with the analysis state, an analysis error covariance matrix $\mathbb{P}^a$ is also calculated as

$$\mathbb{P}^a = (I - K\mathcal{H})\,\mathbb{P}^f \tag{3.14}$$

where I is the identity matrix. There are multiple ways in which this analysis covariance matrix is made available to the next observation window, however, here we will be considering transporting it as a passive tracer following [Menard et al., 2000].

3.3 Background error variance specification

We consider a diagonal background error covariance matrix ($\mathbb{B}$) in all our variational data assimilation experiments for simplicity. The initial variances (the diagonal entries of the $\mathbb{B}$ matrix) are constructed from the average background concentrations $\mathbf{x}_0^B$ on each of the $Nlev$ vertical layers

$$\mathbb{B} = \begin{bmatrix} \mathbb{B}^{(0)} & 0 \dots & 0 \\ 0 & \mathbb{B}^{(1)} \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & 0 \dots & \mathbb{B}^{(Nlev)} \end{bmatrix} \tag{3.15}$$

where

$$\mathbb{B}^{(\ell)} = \begin{bmatrix} \sigma_\ell^2 & 0 \dots & 0 \\ 0 & \sigma_\ell^2 \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & 0 \dots & \sigma_\ell^2 \end{bmatrix}_{dim \times dim} , \quad dim = Nlon \cdot Nlat , \tag{3.16}$$

with

$$\sigma_\ell = \frac{\alpha_{rel}}{dim} \sum_{i=1}^{Nlon} \sum_{j=1}^{Nlat} \mathbf{x}_0^B(i, j, \ell, s_{O3}) , \quad \ell = 1, \cdots, Nlev , \quad s_{O3} = \text{index of ozone} \tag{3.17}$$

The relative uncertainty level in the background initial conditions is taken to be 50%, i.e., $\alpha_{rel} = 0.5$.

The forecast error covariance matrix $\mathbb{P}^f$ used in our suboptimal Kalman filter approach is diagonal. The initial forecast error is assumed to be 50% of the initial forecast field that is supposed to capture the representativeness error as well. In matrix form, $\mathbb{P}_0^f$ could be represented as

$$\mathbb{P}_0^f = \begin{bmatrix} \mathbb{P}_0^{f^{(0)}} & 0 \dots & 0 \\ 0 & \mathbb{P}_0^{f^{(1)}} \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & 0 \dots & \mathbb{P}_0^{f^{(Nobs)}} \end{bmatrix} \tag{3.18}$$

where $Nobs$ is the number of observation grid points in TES retrieval domain. The initial forecast error covariance matrix block corresponding to each observation grid point is

given as

$$\mathbb{P}_0^{f^{(i)}} = \alpha_{rel} \cdot \begin{bmatrix} \mathbf{x}_0^f(i,1,s_{O3}) & 0 & \dots & 0 \\ 0 & \mathbf{x}_0^f(i,2,s_{O3}) & \dots & 0 \\ \vdots & \ddots & & \vdots \\ 0 & 0 & \dots & \mathbf{x}_0^f(i,Nret,s_{O3}) \end{bmatrix}_{Nret \times Nret} \quad , \quad i = 1,2,\dots,Nobs \tag{3.19}$$

where $Nret$ is the number of vertical TES profile retrieval levels. Although the initial forecast error covariance matrix $\mathbb{P}^f$ and all analysis $\mathbb{P}^a$s henceforth are diagonal and there are no horizontal correlations being accounted for, the averaging kernels in the observation operator of TES as defined in equation (3.21) provide vertical correlations when operated on $\mathbb{P}^f$ through equation (3.13). A detailed discussion on how to efficiently extend the background error covariance matrices to non-diagonal forms that capture spatial error correlations is provided in [Singh et al., 2010].

3.4 GEOS-Chem

In this chapter we specifically consider GEOS-Chem (`http://acmg.seas.harvard.edu/geos`), a global three-dimensional chemical transport model (CTM) driven by assimilated meteorological fields from Goddard Earth Observing System(GEOS-4) at the NASA Global Modeling and Assimilation Ofice (GMAO). It is being widely used by research groups world-wide for performing global atmospheric chemistry studies. The model along with comparison of model predictions with observations was first described in [Bey et al., 2001]. GEOS-Chem accounts in detail for emissions from both natural and anthropogenic sources, for gas phase chemistry, aerosol processes, long range transport of pollutants, troposphere-stratosphere exchanges, etc. A description on the underlying governing equation is provided in Chapter 2. Anthropogenic emissions are obtained from the Global Emissions Inventory Activity (GEIA) [Benkovitz et al., 1996] while lightning NOx source emissions are estimated using [Price and Rind, 1992], based on deep convective cloud top heights provided with the GMAO meteorological fields. Biomass burning emissions are based on [Duncan et al., 2003] while biofuel emissions are from [Yevich and Logan, 2003]. The meteorological fields have a horizontal resolution of 1° along latitude and 1.25° along longitude with 55 vertical levels, and a temporal resolution of 6 hrs (3 hrs for surface fields).

The GEOS-Chem Adjoint system (`http://wiki.seas.harvard.edu/geos-chem/index.php/GEOS-Chem_Adjoint`) has been developed through a joint effort of groups at Virginia Tech, University of Colorado, Caltech, Jet Propulsion Laboratory, and Harvard [Henze et al. , 2007; Singh et al., 2009a,b; Eller et al. , 2009]. The system can perform adjoint sensitivity analyses and chemical data assimilation. Inverse modeling

studies with GEOS-Chem v6 adjoint are presented in [Kopacz et al. , 2009; Henze et al., 2009]. Detailed discussion on the construction and validation of GEOS-Chem v7 adjoint is provided in Chapter 2. Also provided are the adjoint sensitivity results of ozone with respect to trace gas, emission and dry deposition precursors using full model adjoint.

3.5 Tropospheric Emission Spectrometer (TES) observations

The Troposphere Emission Spectrometer (TES) [Beer et al., 2001] is a high-resolution imaging infrared Fourier-transform spectrometer, launched aboard the NASA EOS Aura satellite on 14 July 2004 (`http://tes.jpl.nasa.gov`). The Aura satellite is in a polar Sun-synchronous orbit with a repeat cycle of 16 days. The instrument utilizes a nadir-viewing geometry and an instrument field-of-view at the surface of 8 Km $\times$ 5 Km to observe spectral radiances in the range 650-3050 cm^{-1} at an apodized spectral resolution of 0.1 cm^{-1}. It operates in a global survey mode, in which the observations are spaced about 220 Km along the orbit track, and in a step-and-stare mode, in which the observations are spaced every 30 Km along the orbit track.

Vertical profiles of chemical concentrations are retrieved from the radiance measurements on a vertical grid of 67 levels with a discretization of approximately 1 Km per level although the vertical resolution of the profiles are much coarser. The retrieval is an off-line inversion process based on a Bayesian framework that solves a constrained nonlinear least squares problem [Bowmann et al., 2006]. In this work we assimilate the retrieved ozone vertical profiles. Figure 3.1 shows the location of TES profiles for two days.

A-priori information about the vertical concentration profile of the species of interest is needed to solve the retrieval inverse problem (the prior information does not come from the measurement). Let $\mathbf{x}^{\text{prior}}$ be the prior vertical ozone concentration profile (in volume mixing ratio units), and let $\mathbf{z}^{\text{prior}} = \log \mathbf{x}^{\text{prior}}$. Let $\mathbf{z}^{\text{radiance}}$ (=$\log \mathbf{x}^{\text{true}}$) be the atmospheric profile as resulting directly from the radiances.

The vertical ozone profile retrieval can be expressed according to the formula

$$\widehat{\mathbf{z}} = \mathbf{z}^{\text{prior}} + A\,\left(\mathbf{z}^{\text{radiance}} - \mathbf{z}^{\text{prior}}\right) + G\,\eta, \quad \widehat{\mathbf{x}} = \exp\left(\widehat{\mathbf{z}}\right). \tag{3.20}$$

Here A is the averaging kernel matrix, G is the gain matrix, and η is the spectral measurement error (assumed to have mean zero and covariance S_η). More details can be found in [Bowman et al., 2002; Jones et al., 2003; Worden et al., 2004]. The averaging kernels give the sensitivity of the retrieved state to the true state of the atmosphere. The trace of the averaging kernel matrix gives a measure of the number of independent pieces of information available in the measurements, more commonly referred to as the degrees of freedom for signal (DOFS) [Rodgers, 2000].

The corresponding TES observation operator 6.3 is linear with respect to the logarithm of the concentrations, but nonlinear with respect to the concentration profile:

$$\mathcal{H}(\mathbf{x}) = \mathbf{z}^{\text{prior}} + A\left(\log(L(\mathbf{x})) - \mathbf{z}^{\text{prior}}\right) \tag{3.21}$$

where L is an interpolation operator that transforms $\mathbf{x}$ from the GEOS-Chem N-level vertical grid to the TES profile retrieval P-level grid.

For this reason several chemical data assimilation studies based on TES retrieved profiles [Jones et al., 2003; Bowmann et al., 2006; Parrington et al., 2009] have opted to perform the suboptimal Kalman filtering step in the logarithm of the concentrations:

$$\log \mathbf{x}^a = \log \mathbf{x}^f + K\left(\widehat{\mathbf{z}} - \mathcal{H}\left(\mathbf{x}^f\right)\right)$$

For variational data assimilation the forcing calculation is carried out in concentrations. For this reason, an adjoint of the observation operator needs to be derived to update the gradients as described in equations (3.7) and (6.12)

$$(\mathcal{H}'(\mathbf{x}))^T \cdot v = \left(\frac{\partial}{\partial \mathbf{x}}(A \log(L(\mathbf{x})))\right)^T \cdot v = \left(\frac{\partial L}{\partial \mathbf{x}}\right)^T \cdot \begin{pmatrix} (L\mathbf{x})_0^{-1} & 0 & \cdots & 0 \\ 0 & (L\mathbf{x})_1^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & (L\mathbf{x})_P^{-1} \end{pmatrix} \cdot A^T \cdot v$$

Here, $(\mathcal{H}'(\mathbf{x}))^T$ is a matrix and $v = \mathbb{R}^{-1}\left(\mathcal{H}(\mathbf{x}) - \mathbf{x}^{obs}\right)$. The TES averaging kernel A is usually a non-symmetric matrix, and the result of $A^T \cdot v$ is fed to the interpolation operator to construct the diagonal matrix with the i-th element being $1/(L\mathbf{x})_i$. The term $(\partial L/\partial \mathbf{x})^T$ is the adjoint of the interpolation operator and brings entities from the TES profile retrieval domain back to the GEOS-Chem model domain.

NOTE: [Worden et al., 2007] was the first to report that TES ozone retrievals are biased high in the upper troposphere, compared to ozonesonde profiles, while [Nassar et al., 2008] reported that this bias could be as high as 10% throughout the troposphere. We removed this bias as estimated by [Nassar et al., 2008] before assimilating the data.

3.6 Numerical experiments

For numerical experiments, we employ GEOS-Chem v7-04-10 adjoint code package [Singh et al., 2009b], capable of performing both 3D-Var and 4D-Var data assimilations with real data. It also incorporates suboptimal Kalman filter approach of data assimilation developed by Mark Parrington [Parrington et al., 2009]. We assimilate Tropospheric Emission Spectrometer (TES) satellite ozone profile retrievals into the GEOS-Chem model and

validate the generated analyses through an independent observation dataset provided by direct ozone profile measurements from ozonesondes. The numerical optimization method used in all variational experiments is the limited memory bound-constrained BFGS [Zhu et al., 1997]. This quasi-Newton approach has become the "gold standard" in solving large scale chemical data assimilation problems [Sandu et al. , 2005a].

3.6.1 Experimental setting

Simulations with GEOS-Chem v7 adjoint can be carried out at $4^\circ \times 5^\circ$ and $2^\circ \times 2.5^\circ$ resolutions. We have used $4^\circ \times 5^\circ$ resolution in all our experiments. There are 46 x 72 latitude-longitude grid boxes at this resolution, and 55 vertical levels; near the equator and at ground level each grid box covers an area of about 400 Km $\times$ 500 Km. The current GEOS-Chem model does not capture well the dynamics of the upper troposphere and of the stratosphere. Therefore, we performed data assimilation for only the first 23 model levels (for up to about 50 hPa). In addition, as TES is designed specifically for measurements in the tropospheric region, we do not trust the observations completely above 23 levels. This trust level is entered through the observation error covariances in the following manner

```
TRUST_FACTOR = TRUST_FACTOR/100d0
k = 1
do while(obspress(k,n)>geos_plevs(NLEVS))
k = k+1
end do
inv_tes_cov(1:nret,k:nret) = TRUST_FACTOR*inv_tes_cov(1:nret,k:nret)
inv_tes_cov(k:nret,1:k) = TRUST_FACTOR*inv_tes_cov(k:nret,1:k)
```

where, NLEVS is 23 and obspress(k,n)>geos_plevs(NLEVS) provides the profile retrieval level (k) that maps to 23 levels in GEOS-Chem. The TRUST_FACTOR parameter is set through the run script. For all data assimilation experiments, we have set it to 20%.

The 3D-Var data assimilation experiments were performed for a period of two weeks in the month of August 2006, starting at 00:00(GMT) on August 1st. The TES satellite data was read once every 4 simulation hours; the observation operator called at model time t (hours) reads in all the measurements collected within the interval $t-2$ (hours) to $t+2$ (hours). 3D-Var data assimilation treats all observations in this interval as instantaneous, and assimilates them in the same optimization run. In all our 3D-Var experiments, we performed 8 iterations per analysis since the cost function decreased significantly within the first few iterations. It is important to note that 3D-Var does not involve any model adjoint calculations; gradients require only the adjoint of the observation operator. The optimization adjusts ozone concentrations. The generated analysis profile at the end of

each observation window is evolved through the forward model that becomes the initial condition for the next observation window. It is also important to mention here that a new background error covariance matrix (3.15) is constructed for every observation window.

The set up for data assimilation using suboptimal Kalman filter is quite similar to 3D-Var where we assimilated TES profile retrievals into GEOS-Chem over a two week period from 00:00 GMT on August 1, 2006 to 00:00 GMT on August 15, 2006. Observations were read every 4 hours and analysis states were generated for each observation window through the sequential update formula (3.12).

The 4D-Var data assimilation experiments were performed for two different assimilation window lengths to adjudge if model errors hamper the quality of assimilations in GEOS-Chem involving longer assimilation windows; 4D-Var is strongly constrained by the forward model equation (6.11). Starting at 00:00 GMT on August 1, 2006, the first assimilation window is considered to be of five days while the second window is of two weeks. All the three assimilation systems had the same initial conditions to start with and were generated through a free GEOS-Chem model run. There were 12 optimization iterations performed in order to improve the ozone initial condition. Each iteration during 4D-Var assimilation includes a forward model and a backward model adjoint run. TES satellite profile retrievals were read every 4 hours during the model adjoint run, and the cost function and adjoint gradients accumulated the impact of all 4 hour data sets throughout the assimilation window. Contrary to the sequential 3D-Var and suboptimal KF where analysis states were generated every observation window, in 4D-Var the analysis is generated only at the initial time which accounts for the mismatch between observations and model predictions over all the observations in the assimilation window.

3.6.2 Computational costs

3D-Var and suboptimal KF frameworks are built on top of GEOS-Chem v7 package which uses Sparse Matrix Vectorized GEAR (SMVGEAR) solver for chemistry. However, to construct the adjoint of chemistry required by the 4D-Var, we interfaced Kinetic PreProcessor (KPP) library with GEOS-Chem. As pointed out in [Henze et al. , 2007], the computational cost of Rosenbrock solver increases significantly with the tolerance levels; higher tolerances use smaller internal time steps requiring more computation. In our experiments, we have set RTOL=10^{-3} and ATOL=10^{-2} to achieve moderate to high accuracy.

Table 3.1 provides a comparison of the computational costs of different data assimilation systems and the cost of free running model for a 24 hour simulation. All the simulations are performed on a Dell Precision T5400 workstation with 2 quadcore Intel(R) Xeon(R) processors with clock speed 2.33GHz and a RAM of 16GB shared between the two processors.

Table 3.1: Timing results for GEOS-Chem free model runs using SMVGEAR and KPP chemistry, suboptimal Kalman filter, 3D-Var and 4D-Var data assimilations with diagonal background error covariance matrix for a 24 hour simulation starting 00:00 GMT August 1, 2006.

Experiment Description	CPU Time
Free model run, SMVGEAR chemistry solver	2 min 50 sec
Free model run, KPP chemistry solver	3 min 18 sec
Suboptimal Kalman filter with diagonal $\mathbb{P}^f$	3 min 08 sec
3D-Var with diagonal $\mathbb{B}$	3 min 57 sec
4D-Var with diagonal $\mathbb{B}$ (per model run)	16 min 51 sec

Suboptimal Kalman filter is less expensive than 3D-Var since it generates the analysis through the single update formula (3.12), while 3D-Var requires a few iterations before the optimization routine could generate a stable optimal analysis field. This is true however as long as the forecast error covariance matrix is diagonal. Once we move to non-diagonal matrices, the cost of calculating Kalman gain matrix as described in equation (3.13), would be very high. In case of 3D-Var and 4D-Var, using even full $\mathbb{B}$ matrix adds a minimal cost to the overall simulation since the complete matrix is never constructed; at each step only a matrix vector product is required and efficient techniques are employed to derive the inverse and other powers of $\mathbb{B}$ matrix [Singh et al., 2010]. The 4D-Var assimilation is most expensive of all the assimilation systems under consideration. The reason is attributed to the fact that a single 4D-Var iteration performs both the forward and adjoint model runs, where, several variables on which the adjoint equation depends on, are written in checkpoint files in the forward model run. A plot of disk memory requirements versus time is presented in Chapter 6.

3.6.3 Comparison with ozonesonde measurements

In order to assess the quality of analysis fields generated through different assimilation systems, we use ozonesonde profiles measured by the INTEX Ozonesonde Network Study 2006 (IONS-6) (`http://croc.gsfc.nasa.gov/intexb/ions06.html` [Thompson et al., 2007a, 2007b]) for the month of August, assuming that these measurements provide values close to the true state of the atmosphere. There are 418 ozonesondes launched from 22 stations across North America as shown in the Figure 3.1. A detailed description of the number of ozonesondes launched per station with longitude and latitude information can be found in [Parrington et al., 2008]. The ozonesonde observations are not used in data assimilation, and therefore provide an independent data set against which the analysis results are validated. Forecast scoring techniques using assimilated data as described in [Wu et al., 2008; Constantinescu et al., 2007c] do not

provide a fair assessment of the quality of assimilation if the observation measurements involved high observation errors.

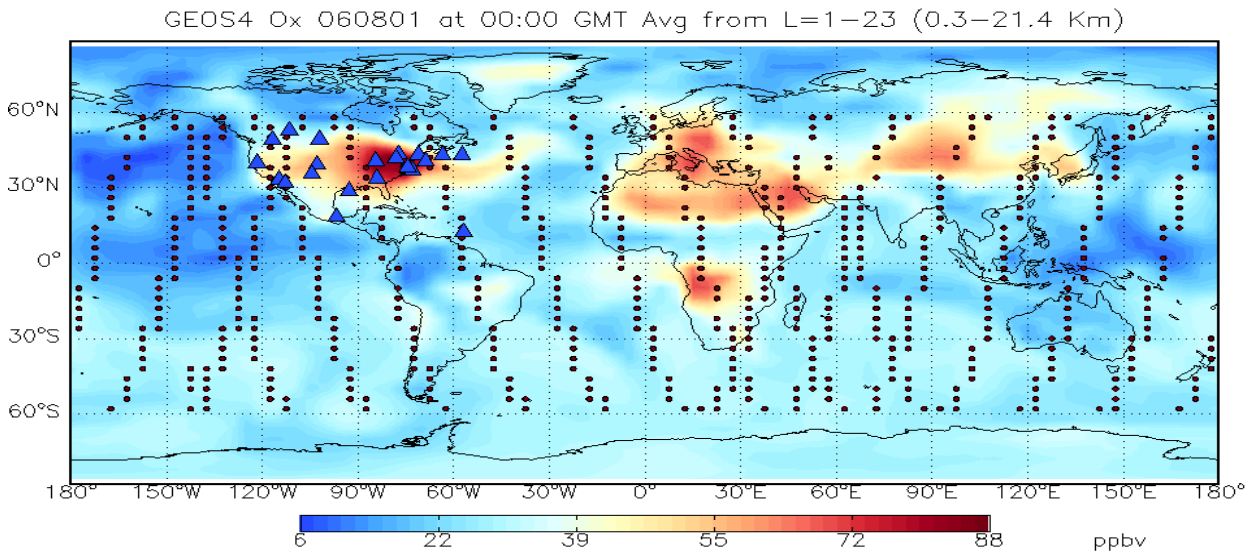


Figure 3.1: Ozonesonde sounding stations (triangles) used during IONS06 campaign and AURA/TES satellite trajectory snapshots (dots) plotted over the global ozone distribution on August 1st, 2006.

We first consider the case where assimilation window length is of 5 days. As per the property of sequential data assimilation, we obtain an analysis field at the end of every observation window that accounts for the mismatch between the model prediction and the observations within that window. However, it is important to note that the model prediction at any observation window incorporates implicitly the corrections from all previous observations. Thus, as we move forward in time, the analysis field agrees better with the true state of the atmosphere. 4D-Var on the other hand accumulates the forcing due to mismatch between model forecast and observations throughout the assimilation window to produce an initial condition that when evolved forward in time through the model, will best fit the observations. Therefore, in case of sequential assimilation approaches, to obtain a stable analysis state that resembles the true chemical state of the atmosphere at a particular instant, we need to start the simulation days or months in prior. In many cases, there are no prior observations or meteorological fields available. 4D-Var is advantageous in such situations as it provides the best estimate using only the observations available in the assimilation window under consideration.

We present in Figure 3.2, a comparison of analysis profiles obtained from different assimilation systems, and free GEOS-Chem model run against ozonesonde measurement data. The plots provide an assessment of the quality of tropospheric ozone as estimated

by suboptimal Kalman filter, 3D-Var and 4D-Var approaches, and in turn reflect the impact of TES profile retrievals on these assimilation systems. The left panel is the plot of pressure level against ozone estimates averaged over all ozonesonde locations for all ozonesonde measurements available in the 5-day assimilation window. The center panel provides the relative difference of mean ozone estimates from assimilation systems and free model run against mean ozonesonde measurements. Since the mean values do not clearly reflect the accuracy of each point, we provide in rightmost panel, the standard deviation of absolute values of the differences between ozone predictions and ozonesonde measurements.

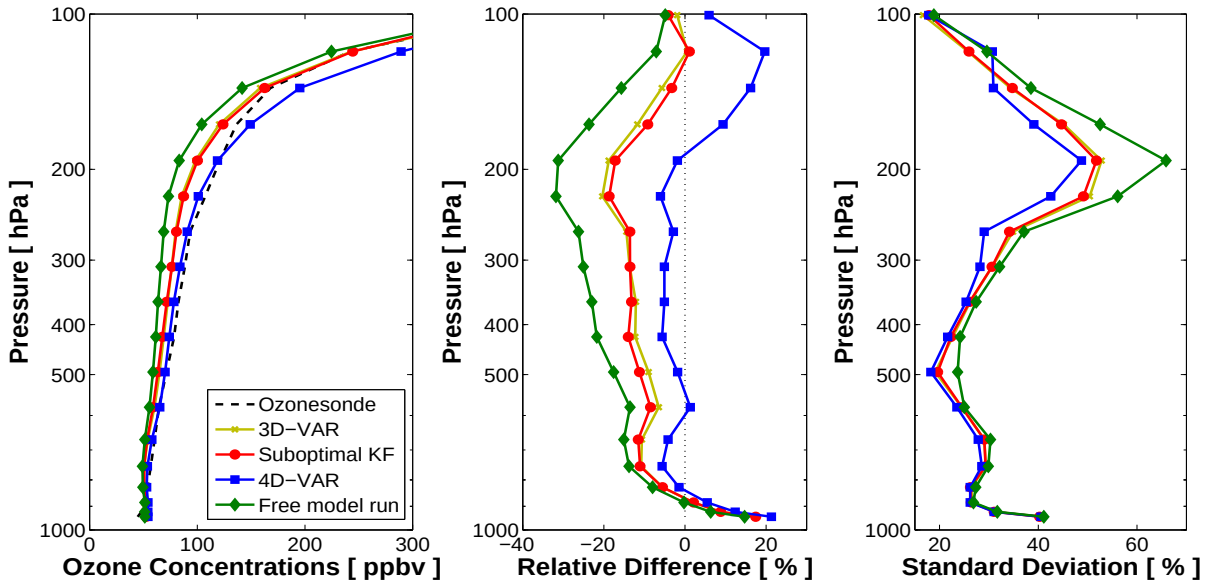
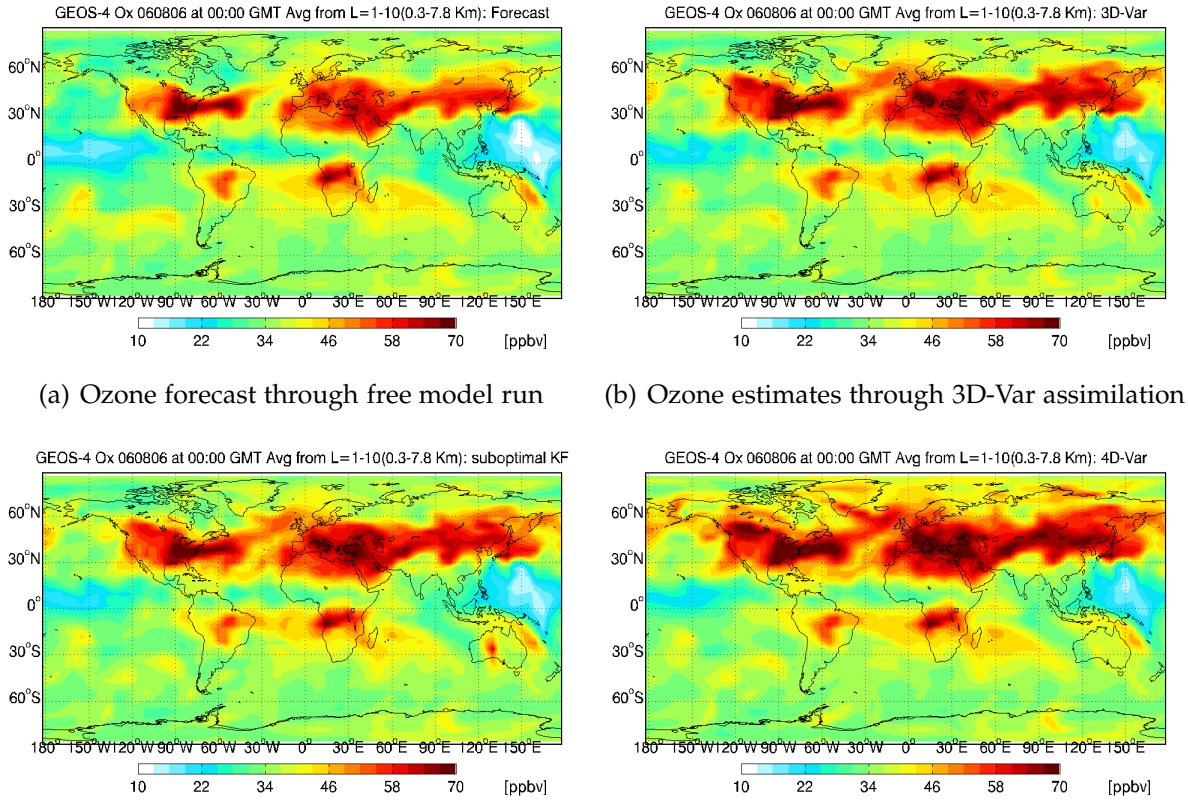


Figure 3.2: The impact of ozone profile retrievals from TES on data assimilation systems for GEOS-Chem. Left panel: mean ozone concentrations at ozonesonde locations for 3D-Var, 4D-Var, suboptimal KF analyses and free model trajectories. Center panel: relative mean errors of predicted ozone concentrations with respect to ozonesonde measurements. Right panel: standard deviation of absolute values of errors with respect to ozonesonde measurements. The data is averaged over all ozonesonde launches. These plots were generated from 5 days simulation from 00:00 GMT August 1, 2006 to 00:00 GMT August 6, 2006 and compared against ozonesonde data available for the month of August.

It is evident that 4D-Var provided the best estimate for lower and mid troposphere. The relative difference between the mean ozone analysis field and the ozonesonde measurements were decreased to less than 4% up to 180 hPa as compared to 5-20% in cases of suboptimal KF and 3D-Var. The overestimation of ozone in the upper troposphere by 4D-Var is intriguing and could be attributed to the lack of sensitive observations at

higher levels. A detailed analysis on the information brought in by TES profile retrievals into the 4D-Var assimilation system at different pressure levels is provided in Chapter 5.



(a) Ozone forecast through free model run

(b) Ozone estimates through 3D-Var assimilation

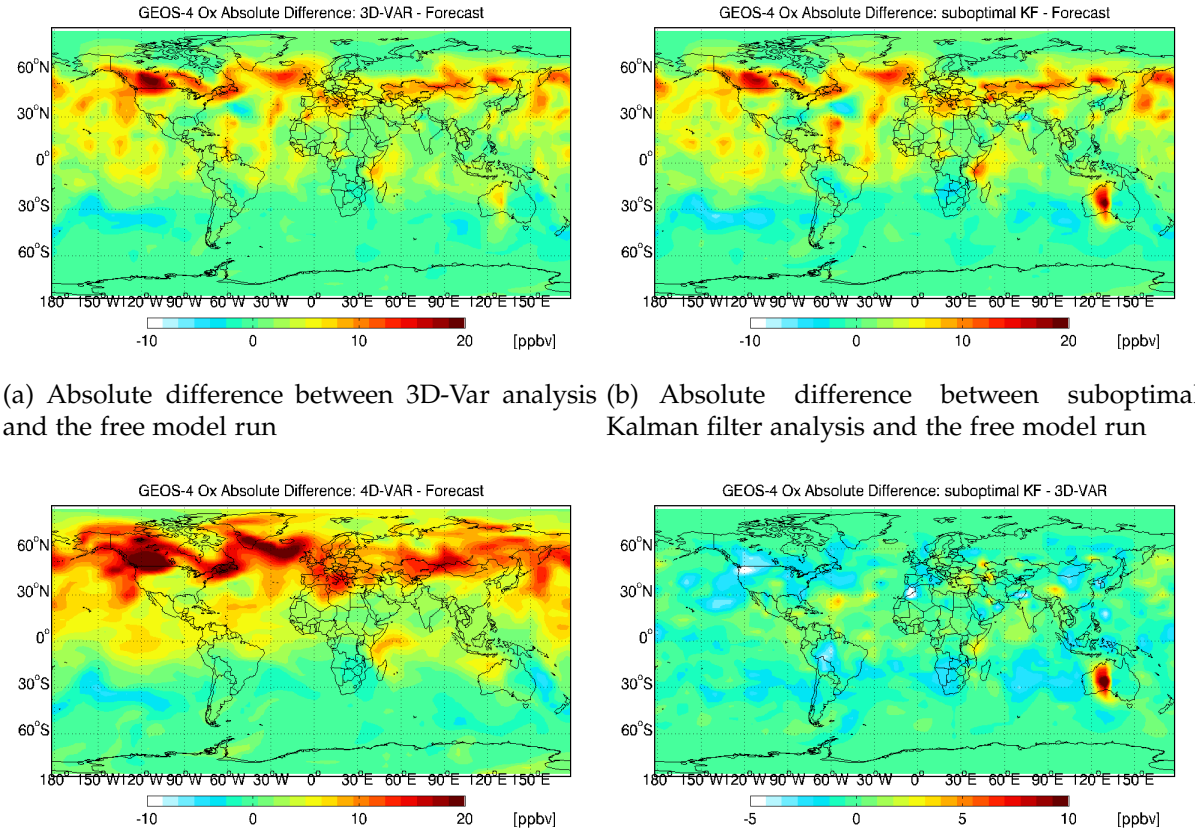
(c) Ozone estimates through suboptimal Kalman filter

(d) Ozone estimates through 4D-Var assimilation

Figure 3.3: Global ozone distribution at 00:00 GMT on August 6, 2006 averaged over the first 10 GEOS-Chem vertical levels. Panels (a)-(d): Global tropospheric ozone estimates provided by free model run and suboptimal KF, 3D-Var, and 4D-Var data assimilation systems from a 5-day simulation.

Figure 3.3 provides the global tropospheric ozone distribution as estimated by GEOS-Chem free model run and different assimilation systems. The ozone concentration values are averaged over 10 GEOS-Chem levels (about 370 hPa) for each longitude-latitude grid point on the horizontal domain. Due to abundance of ozone in higher layers, if more levels were included in the plots, the contribution of each assimilation system is subdued and the plots look quite similar. All the assimilation systems seem to have caused an increase in the tropospheric ozone as compared to the model forecast with 4D-Var bringing the highest amount. The gain seems to be prominent in the 30° N to 60° N latitude region in case of suboptimal KF and 3D-Var, while it is extended up to 90°

N in case of 4D-Var. For a clear demonstration of these changes, we provide in Figure 3.4, the plots of differences in the tropospheric ozone estimates through free model run and different assimilation systems.



(a) Absolute difference between 3D-Var analysis and the free model run

(b) Absolute difference between suboptimal Kalman filter analysis and the free model run

(c) Absolute difference between 4D-Var analysis and the free model run

(d) Absolute difference between suboptimal Kalman filter and the 3D-Var analyses

Figure 3.4: Differences in global ozone concentrations at 00:00 GMT on August 6, 2006, the end of 5-day simulation, averaged over first 10 GEOS-Chem vertical levels. Panels (a)-(c): Differences between suboptimal KF, 3D-Var, and 4D-Var analysis fields and the model forecast (solution without data assimilation). Panel (d): Difference between suboptimal KF and 3D-Var analysis fields.

In Figure 3.4, panels (a) and (b) show that the structure of corrections in the ozone concentrations through 3D-Var and suboptimal KF data assimilation are quite similar. The reason behind such a structure is that these sequential algorithms bring in instantaneous corrections based solely on the mismatch between the model predictions and the observations in an observation window (analysis cycle). The localized corrections here are mostly along the Aura satellite orbit. Panel (c) on the other hand showcases the smoother correction profile of 4D-Var. In each 4D-Var optimization iteration, the

cost function and gradients are accumulated for all the observation windows where the adjoint variable (gradient) is flown backwards in time as governed by the model adjoint equation. The corrections brought in by the optimization routine therefore are no more localized. We also plot the difference in the analysis fields obtained by 3D-Var and suboptimal KF showcasing their close resemblance (panel (d)). Interestingly, there seems to be a localized overcorrection in the mid west Australian region by the suboptimal Kalman filter.

We next consider simulations with assimilation window length of 2 weeks. A longer assimilation window provides an insight into how ozone estimates due to assimilation evolve with time and if the corrections maintain structures similar to 5-day case. It also helps adjudge if model errors in GEOS-Chem cause any degradation in the assimilation systems, especially the strongly constrained 4D-Var. Similar to Figure 3.2, we present in

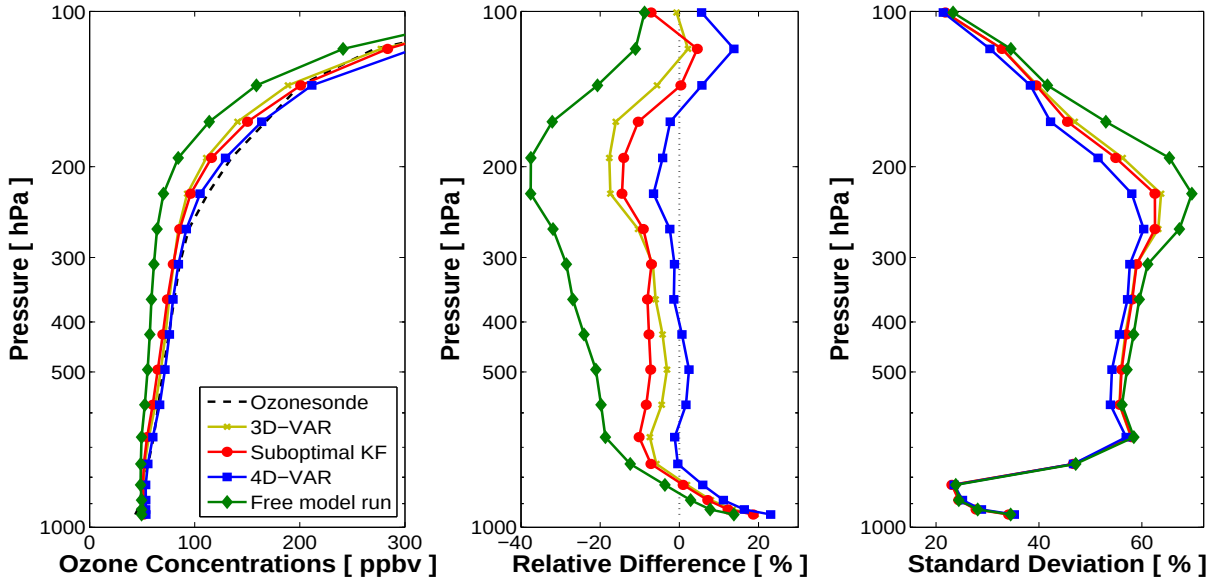
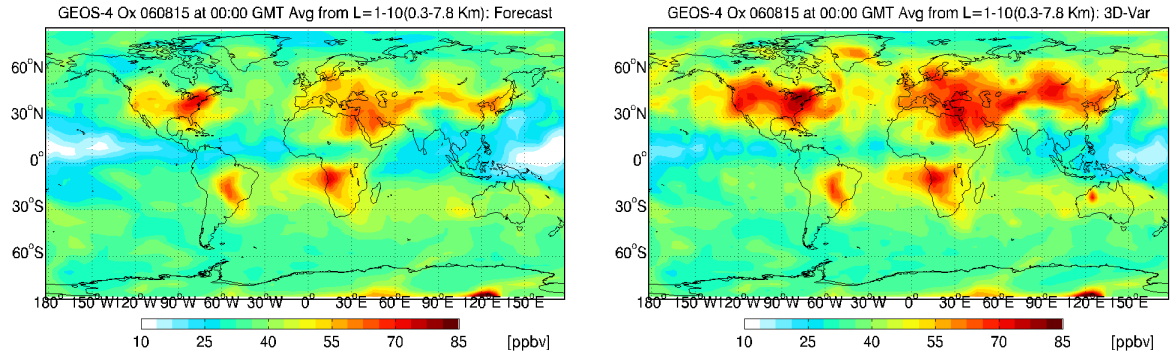


Figure 3.5: The impact of ozone profile retrievals from TES on data assimilation systems for GEOS-Chem. Left panel: mean ozone concentrations at ozonesonde locations for 3D-Var, 4D-Var, suboptimal KF analyses and free model trajectories. Center panel: relative mean errors of predicted ozone concentrations with respect to ozonesonde measurements. Right panel: standard deviation of absolute values of errors with respect to ozonesonde measurements. The data is averaged over all ozonesonde launches. These plots were generated from 5 days simulation from 00:00 GMT August 1, 2006 to 00:00 GMT August 6, 2006 and compared against ozonesonde data available for the month of August.

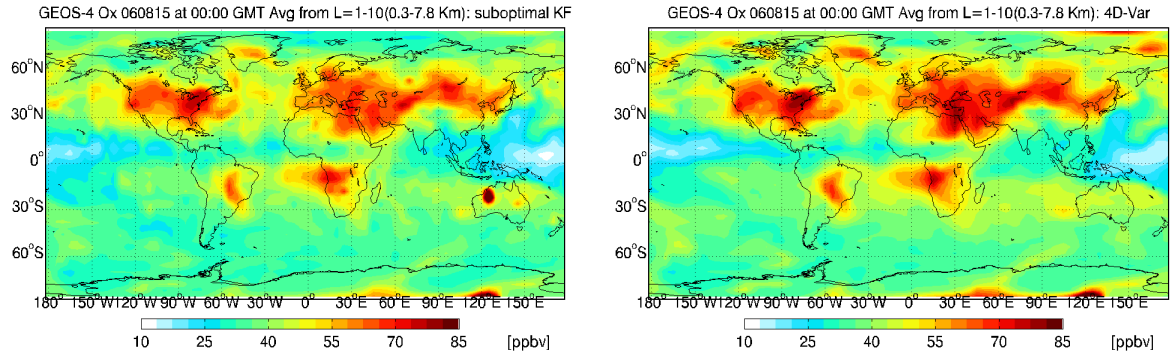
Figure 3.5, a comparison of analysis profiles obtained from different assimilation systems against ozonesonde measurement data. The plots reflect that the accuracy of suboptimal

Kalman filter and 3D-Var assimilations start to differ with longer assimilation window. While suboptimal KF underestimates ozone concentrations in the lower and mid troposphere, it performs better than 3D-Var in the mid and upper tropospheric region. 4D-Var still provided the best ozone estimate of all the assimilation systems, and, unlike the 5 days assimilation window length case, it performed well in the upper tropospheric region as well except near the stratospheric boundary. Panel (c) suggests that the standard deviation of 4D-Var analysis from the ozonesonde measurements stayed the least among all the assimilation systems. The relative difference between the mean ozone analysis field and the ozonesonde measurements were decreased to less than 4% upto 150 hPa as compared to 4-16% in cases of suboptimal KF and 3D-Var. With longer assimilation window, all the assimilation systems seem to have benefited from more (meaningful) observations being brought in.



(a) Ozone forecast through free model run

(b) Ozone estimates through 3D-Var assimilation

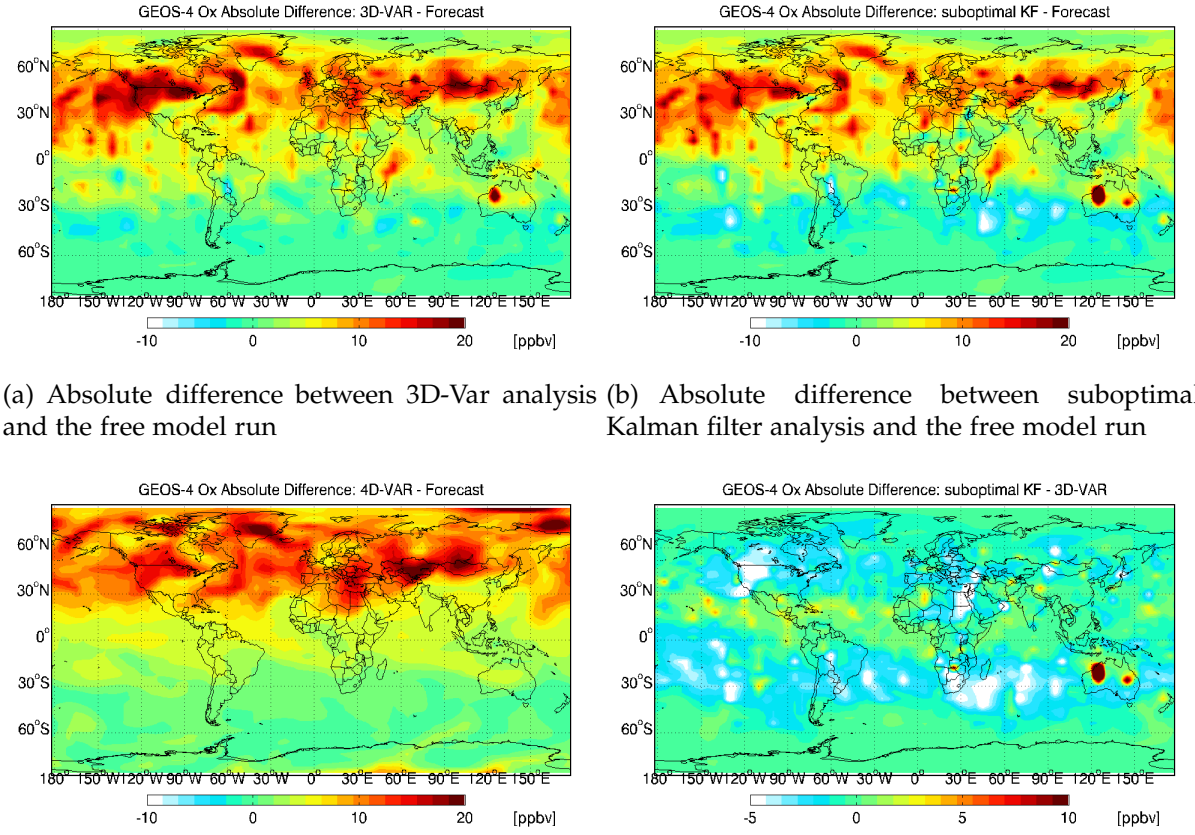


(c) Ozone estimates through suboptimal Kalman filter

(d) Ozone estimates through 4D-Var assimilation

Figure 3.6: Global ozone distribution at 00:00 GMT on August 6, 2006 averaged over the first 10 GEOS-Chem vertical levels. Panels (a)-(d): Global tropospheric ozone estimates provided by free model run and suboptimal KF, 3D-Var, and 4D-Var data assimilation systems from a 5-day simulation.

Figure 3.6 provides the global tropospheric ozone distribution as estimated by GEOS-Chem free model run and different assimilation systems. The underestimation of ozone by suboptimal KF in the lower and mid troposphere as noted earlier in Figure 3.5 is quite visible here. Similar to the 5 days assimilation window case, 4D-Var causes the maximum increase in the tropospheric ozone.



(a) Absolute difference between 3D-Var analysis and the free model run

(b) Absolute difference between suboptimal Kalman filter analysis and the free model run

(c) Absolute difference between 4D-Var analysis and the free model run

(d) Absolute difference between suboptimal Kalman filter and the 3D-Var analyses

Figure 3.7: Differences in global ozone concentrations at 00:00 GMT on August 6, 2006, the end of 5-day simulation, averaged over first 10 GEOS-Chem vertical levels. Panels (a)-(c): Differences between suboptimal KF, 3D-Var, and 4D-Var analysis fields and the model forecast (solution without data assimilation). Panel (d): Difference between suboptimal KF and 3D-Var analysis fields.

Figure 3.7 showcases the structure of corrections in model predicted ozone through different assimilation systems. The localized correction structure in 3D-Var and suboptimal KF cases still persists with longer assimilation window. 4D-Var provides a better correction distribution with significant increase in ozone concentrations in the 30° N to 90° N latitude region. Interestingly the localized correction in the mid west Australian region

which was not visible in the 3D-Var case for 5 days assimilation window case, seems to be prominent in longer assimilation, while, in case of suboptimal KF it seems to have accentuated.

Contrary to what was observed in [Wu et al., 2008] for the 4D-Var assimilation in Polair3D case where the accuracy of the ozone estimates decreased with increase in the assimilation window length, our findings show that with increase in the assimilation window, the 4D-Var system performs even better. It seems ingesting meaningful observations keep the effect of model errors from compromising the quality of assimilation. There is however one case where the accuracy of ozone estimates decrease with increase in assimilation window length for 4D-Var and that is when the model adjoints are inaccurate. We have studied this case in detail in Chapter 6 and have utilized inaccurate gradients to work towards our benefit in terms of memory and computational costs.

3.7 Conclusions

We have successfully implemented 3D-Var and 4D-Var data assimilation frameworks into GEOS-Chem v7 adjoint package to carry out assimilations using real data. The current study involved estimation of global tropospheric ozone distribution. We provide the first set of results of direct comparison between 3D-Var, 4D-Var and suboptimal Kalman filter data assimilation systems. If using diagonal background error covariance matrix, suboptimal KF is computationally least expensive followed by 3D-Var. Memory and computational costs for 4D-Var is highest as it requires checkpointing dependent variables and an additional adjoint model run every iteration. The quality of estimated ozone from different assimilation systems were verified using ozonesonde measurements, an independent data set.

Two different assimilation window lengths were considered. All the three assimilation systems seem to have benefited from TES profile retrievals. Sequential assimilation methods, 3D-Var and suboptimal KF seem to perform similar for smaller assimilation window decreasing the relative difference between mean analysis and ozonesonde measurements to about 5-20%. 4D-Var on the other hand brought down this difference to less than 4% for upto 180 hPa. For larger assimilation window length, the sequential approaches seem to differ a bit with suboptimal KF underestimating the ozone concentrations in the lower and mid troposphere, however performing better than 3D-Var in the mid and upper troposphere. The relative difference was brought down to 4-16% by sequential approaches while to less than 4% up to 150 hPa by 4D-Var. The structure of corrections in ozone concentration due to sequential assimiation methods seem to be localized along the satellite orbit, while smoother and well distributed in case of 4D-Var. The latitude region 30° N to 60° N seem to have gained maximum from all the assimilation systems extending upto 90° N in case of 4D-Var.

The developed data assimilation frameworks and comparison results would enable users of GEOS-Chem to obtain better estimates for all available trace gases at surface, tropospheric and stratospheric levels, depending on the memory and computational requirements. The framework is currently built to assimilate TES profile retrievals, however, it could easily be extended to use data from any instrument. Another interesting idea is to carry out assimilation with respect to surface emissions; a good estimate of emission rates is significantly important for policy making. Some other interesting and completely new applications involving variational data assimilation methods are studied in the next chapters of this dissertation.

Chapter 4

Construction of Non-diagonal Background Error Covariance Matrices for Global Chemical Data Assimilation

Abstract

It is widely accepted that a key ingredient for successful data assimilation is a realistic estimation of the background error distribution. Particularly important is the specification of the background error covariance matrix, which contains information about the magnitude of the background errors and about their correlations. As models evolve toward finer resolutions, the use of diagonal background covariance matrices is increasingly inaccurate, as they captures less of the spatial error correlations. This chapter discusses an efficient computational procedure for constructing non-diagonal background error covariance matrices which account for the spatial correlations of errors. The benefits of using the non-diagonal covariance matrices for variational data assimilation with chemical transport models are illustrated.

4.1 Introduction

Chemical data assimilation attempts to optimally use noisy observations of reality along with imperfect model predictions to produce a better estimate (in some optimal sense) of the chemical state of the atmosphere. This improved estimate state better defines the spatial and temporal fields of key chemical components in relation to their sources and sinks. This information is critical for improved studies of the atmospheric composition. Chemical data assimilation could also, in principle, improve estimates of emission in-

ventories, of model boundary conditions, or of important model parameters like wet deposition velocities or photolysis rates.

A realistic estimation of the background error distribution is key to a successful data assimilation. Particularly important is the specification of the background error covariance matrix, which contains information about the magnitude of the background errors and about their correlations. Background covariance matrices impact how the information from observations is spread both spatially and among the different types of analysis variables.

The construction of background covariance matrices is challenging due to poorly characterized background errors, and to the very large dimension of the state space of realistic atmospheric models. As a consequence, many chemical data assimilation studies to date have used diagonal background covariance matrices. A popular approach to approximate the background covariance matrix is the NMC method [Parrish and Derber, 1992], in which the differences between several forecasts verifying at the same time are used to approximate the background error. This method has been successfully applied to chemical data assimilation [Chai et al. , 2006]. An alternative approach constructs autoregressive models of background errors based on the short-term linearized model dynamics [Constantinescu et al., 2007a].

A popular ansatz is that the background error correlations decay exponentially in space. This ansatz allows the construction of simple error correlation models, and is the basis of the covariance localization technique used in ensemble Kalman filtering [Gaspari and Cohn, 1999; Ott et al., 2004; Constantinescu et al., 2007b]. Experimental studies with chemical transport models support this assumption; for example, in [Chai et al. , 2006] it has been shown that ozone error correlations decrease follow, on average, an exponential decay curve. Sub-optimal Kalman filters based on covariance models that impose an exponential decay of correlations with distances have been used in the assimilation of chemical constituents [Khattatov et al., 1999; Pierce et al., 2007].

In the troposphere, ozone is an important greenhouse gas and a major pollutant, which adversely impacts air quality. Its distribution is highly heterogeneous, reflecting the combined influence of atmospheric transport and local chemical sources and sinks. Until recently, observations of the three-dimensional structure of tropospheric ozone have been limited. The Tropospheric Emission Spectrometer (TES) satellite instrument, launched in 2004, produced the first continuous, global profile retrievals of tropospheric ozone. Similar observations are now available from other satellite instruments, such as the Infrared Atmospheric Sounding Interferometer (IASI). Assimilating these data into atmospheric models provides a powerful means to obtain an improved understanding of the processes controlling tropospheric ozone. Parrington et al. [2008] was the first to assimilate the TES ozone profile retrievals, but they did not account for horizontal correlations in the background error.

We propose here a computationally efficient approach for constructing (background)

error covariances that account for spatial correlations in both horizontal and vertical directions, and assess its impact on the assimilation of tropospheric ozone profiles from TES. The construction is based but not restricted to the ansatz of exponential decay of error correlations. The correlation lengths in the latitudinal, longitudinal, and vertical directions can be specified according to the application requirements. Due to the large number of state variables an explicit representation of the full background covariance matrix is impractical. The proposed strategy constructs a multi-dimensional correlation matrix from tensor products of one-dimensional correlation matrices. This avoids the explicit construction and storage of full covariance matrices, and allows the needed linear algebra operations to be performed very efficiently.

The chapter is organized as follows. The algorithm for constructing multidimensional covariance matrices is discussed in Section 4.2. Section 4.3 presents assimilation results of TES ozone profiles with the global chemical transport model GEOS-Chem, and illustrates the benefits of nondiagonal covariances in both three and four dimensional variational data assimilation settings. Section 4.4 draws conclusion and provides points of future work.

4.2 Construction of the background error covariance matrix

A correct characterization of the background errors is necessary for obtaining a meaningful analysis, i.e., for the success of the data assimilation procedure. Under the usual assumption that the background errors are normally distributed their probability density is described by the background state $\mathbf{x}^b$ and the background error covariance matrix $\mathbb{B}$. In variational data assimilation both $\mathbf{x}^b$ and $\mathbb{B}$ enter directly into the formulation of the cost function; errors in their specification directly impact the analysis result [Daescu, 2008].

A non-diagonal background error covariance matrix allows the information from local observations to spread out in space to contribute to corrections of state variables in neighboring locations; similarly, it allows observations of certain components of the state vector to contribute to corrections of other components. This spread of information results in a smooth analysis state, and allows different sets of observations to complement each other.

Despite these advantages, most chemical data assimilation studies to date have employed diagonal background covariances. Little work has been devoted to date to modeling off-diagonal terms [Chai et al. , 2006; Constantinescu et al., 2007a]. This is due to a number of practical difficulties that arise in the construction of background covariance matrices. The "true" state is, fundamentally, unknown, and so are the "true" errors;

surrogate states have to be used to mimic forecast errors. Ensembles can be employed to estimate error correlations; however, the number of ensemble members is necessarily very small and only low rank approximations of the covariance matrix can be obtained. Localization is often employed to remove spurious correlations and to improve the rank of the resulting matrix [Gaspari and Cohn, 1999]. The large number of state variables make the construction and storage of full covariance matrices impractical.

We next discuss the proposed approach to constructing a background error covariance matrix $\mathbb{B}$ that accounts for both vertical and horizontal correlations without explicitly constructing the full covariance matrix. We explain the construction of the matrix in the two-dimensional case, i.e., for capturing horizontal correlations; the extensions to correlations in three dimensions and to correlations among multiple state variables are immediate. Our target application is global chemical data assimilation using GEOS-Chem.

Consider a uniform latitude-longitude grid and denote by x the longitude and by y the latitude level. A grid point (x_i, y_j) has longitude coordinate x_i, $i = 1, \ldots, n_x$, and latitude coordinate y_j, $i = 1, \ldots, n_y$. The state vector contains the state values at all gridpoints ordered latitude-first:

$$\Big[(x_1, y_1), \cdots, (x_1, y_{n_y}), (x_2, y_1), \cdots, (x_2, y_{n_y}), (x_{n_x}, y_1), \cdots, (x_{n_x}, y_{n_y})\Big] \tag{4.1}$$

4.2.1 Directional error correlation matrices

The one-dimensional correlation between errors at two locations (x_i, y_k) and (x_j, y_k) situated at the same latitude y_k is modeled as

$$\left(\widetilde{\mathbb{C}}_x^k\right)_{i,j} = \text{corr}\left((x_i, y_k), (x_j, y_k)\right) = e^{-\frac{\text{dist}\left((x_i,y_k),(x_j,y_k)\right)^2}{\ell_x^2}}; \quad i, j = 1, \cdots, n_x; \quad k = 1, \cdots, n_y; \tag{4.2}$$

where ℓ_x is the correlation distance in the longitude direction. For a uniform lat-lon grid the distance between x_i and x_j depends only on $\min(|i-j|, n_x - |i-j|)$. This distance also depends on the y_k; for this reason equation (4.2) defines a different longitudinal correlation matrix $\widetilde{\mathbb{C}}_x^k \in \mathbb{R}^{n_x \times n_x}$ for each latitude y_k. Due to the periodicity along each latitude circle the point x_1 is strongly correlated with both x_2 and x_{n_x}, etc. The periodicity is captured by the distance function in (4.2).

Similarly, the one-dimensional correlation between errors at two locations (x_k, y_i) and

(x_k, y_j) situated at the same longitude x_k is modeled as

$$\left(\widetilde{\mathbb{C}}_y^k\right)_{i,j} = \text{corr}\left((x_k, y_i), (x_k, y_j)\right) = e^{-\frac{\text{dist}\left((x_k,y_i),(x_k,y_j)\right)^2}{\ell_y^2}}, \quad i,j = 1,\cdots,n_y\,; \quad k = 1,\cdots,n_y\,; \tag{4.3}$$

where ℓ_y is the correlation distance in the latitude direction. Equation (4.3) defines a single latitudinal correlation matrix $\widetilde{\mathbb{C}}_y \in \mathbb{R}^{n_y \times n_y}$. For a uniform lat-lon grid this correlation matrix is the same for each longitude x_k; consequently the superscript k is dropped. To simplify the construction the correlations due to the periodicity along a meridional circle are ignored. Otherwise, error correlations across the poles would lead to correlations between errors at all longitudes; such cross-correlations are not captured by (4.3).

The cost function and gradient calculations described in equations (6.8), (3.7), (6.10), and (6.12) require the inverse of the background error covariance matrix; this involves the inverses of the correlation matrices in longitudinal and latitudinal directions. The construction of the directional correlation matrices $\widetilde{\mathbb{C}}_x$ and $\widetilde{\mathbb{C}}_y$ does not guarantee that they are non-singular. To avoid a possible singularity we take a convex combination between the identity matrix and tensor product correlations as follows:

$$\mathbb{C}_x^k = \theta_x\ \mathbb{I}_{n_x \times n_x} + (1 - \theta_x)\ \widetilde{\mathbb{C}}_x^k\,, \tag{4.4}$$

and

$$\mathbb{C}_y^k = \theta_y\ \mathbb{I}_{n_y \times n_y} + \left(1 - \theta_y\right)\ \widetilde{\mathbb{C}}_y^k\,. \tag{4.5}$$

The above procedure brings a shift in the spectrum and ensures the positive definiteness of $\mathbb{C}_x$ and $\mathbb{C}_y$. In all our experiments both θ_x and θ_y are chosen to be equal to 0.2.

The longitudinal correlation matrix between all the points on the two-dimensional grid (4.1) can be constructed from the one-dimensional longitudinal correlation matrices as follows

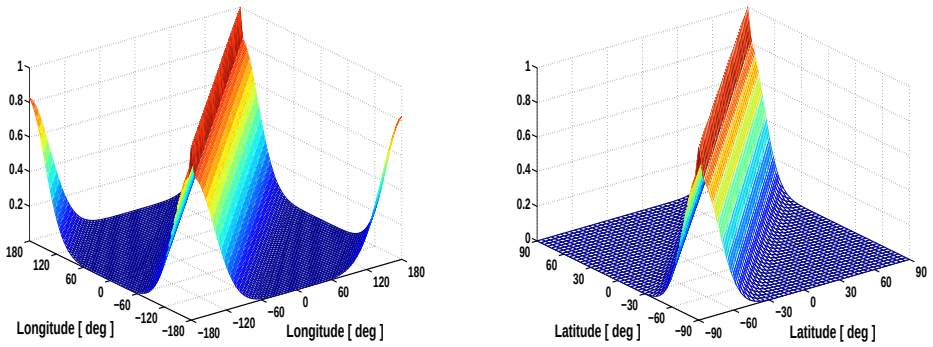
$$\mathbb{C}_x^{1:n_y} \otimes \mathbb{I}_{n_y \times n_y} = \begin{pmatrix} \begin{pmatrix} (\mathbb{C}_x^1)_{1,1} & 0 & \cdots & 0 \\ 0 & (\mathbb{C}_x^2)_{1,1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \left(\mathbb{C}_x^{n_y}\right)_{1,1} \end{pmatrix} & \cdots & \begin{pmatrix} (\mathbb{C}_x^1)_{1,n_x} & 0 & \cdots & 0 \\ 0 & (\mathbb{C}_x^2)_{1,n_x} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \left(\mathbb{C}_x^{n_y}\right)_{1,n_x} \end{pmatrix} \\ \vdots & \ddots & \vdots \\ \begin{pmatrix} (\mathbb{C}_x^1)_{n_x,1} & 0 & \cdots & 0 \\ 0 & (\mathbb{C}_x^2)_{n_x,1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \left(\mathbb{C}_x^{n_y}\right)_{n_x,1} \end{pmatrix} & \cdots & \begin{pmatrix} (\mathbb{C}_x^1)_{n_x,n_x} & 0 & \cdots & 0 \\ 0 & (\mathbb{C}_x^2)_{n_x,n_x} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \left(\mathbb{C}_x^{n_y}\right)_{n_x,n_x} \end{pmatrix} \end{pmatrix}$$

With some abuse of notation we extend the use of the Kronecker product symbol $\otimes$ in the above equation in order to highlight the structure of the two-dimensional longitudinal correlation matrix.

Similarly, the latitudinal correlation matrix between all the points on the two-dimensional grid (4.1) can be constructed from the one-dimensional latitudinal correlation matrices as follows

$$\mathbb{I}_{n_x \times n_x} \otimes \mathbb{C}_y = \begin{pmatrix} \begin{pmatrix} (\mathbb{C}_y)_{1,1} & (\mathbb{C}_y)_{1,2} & \cdots & (\mathbb{C}_y)_{1,n_y} \\ (\mathbb{C}_y)_{2,1} & (\mathbb{C}_y)_{2,2} & \cdots & (\mathbb{C}_y)_{2,n_y} \\ \vdots & \vdots & \ddots & \vdots \\ (\mathbb{C}_y)_{n_y,1} & (\mathbb{C}_y)_{n_y,2} & \cdots & (\mathbb{C}_y)_{n_y,n_y} \end{pmatrix} & \cdots & \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix} \\ \vdots & \ddots & \vdots \\ \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix} & \cdots & \begin{pmatrix} (\mathbb{C}_y)_{1,1} & (\mathbb{C}_y)_{1,2} & \cdots & (\mathbb{C}_y)_{1,n_y} \\ (\mathbb{C}_y)_{2,1} & (\mathbb{C}_y)_{2,2} & \cdots & (\mathbb{C}_y)_{2,n_y} \\ \vdots & \vdots & \ddots & \vdots \\ (\mathbb{C}_y)_{n_y,1} & (\mathbb{C}_y)_{n_y,2} & \cdots & (\mathbb{C}_y)_{n_y,n_y} \end{pmatrix} \end{pmatrix}$$

The structure of the one-dimensional correlation matrices is represented in Figure 4.1. The longitudinal correlation $\mathbb{C}_x^k$ is represented at latitude $y_k = 20\,^\circ N$; note that due to the periodicity along each latitude circle not only the elements near the diagonal, but also the elements in the corners of the matrix have non-zero values. The latitudinal correlation $\mathbb{C}_y$ does not account for periodicity (along each meridian the grids 1 and n_y correspond to the South and to the North pole, respectively).



(a) Longitudinal correlation matrix $\mathbb{C}_x^k$ at latitude $y_k = 20\,^\circ N$

(b) Latitudinal correlation matrix $\mathbb{C}_y$

Figure 4.1: Mesh representation of the one-dimensional longitudinal and latitudinal correlation matrices. The latitude-longitude model grid resolution is $4^\circ \times 5^\circ$ (about 400Km $\times$ 500Km near the equator) and the correlation lengths are ℓ_x = 1500Km and ℓ_y = 1200Km.

Figure 4.2 represents contour lines of the longitudinal correlation $\mathbb{C}_x$ for points at different latitudes. The correlation length ℓ_x is short (top panel), medium (middle panel), and

large (bottom panel). Note that the same correlation length ℓ_x translates into a larger number of correlated grid points at higher latitudes.

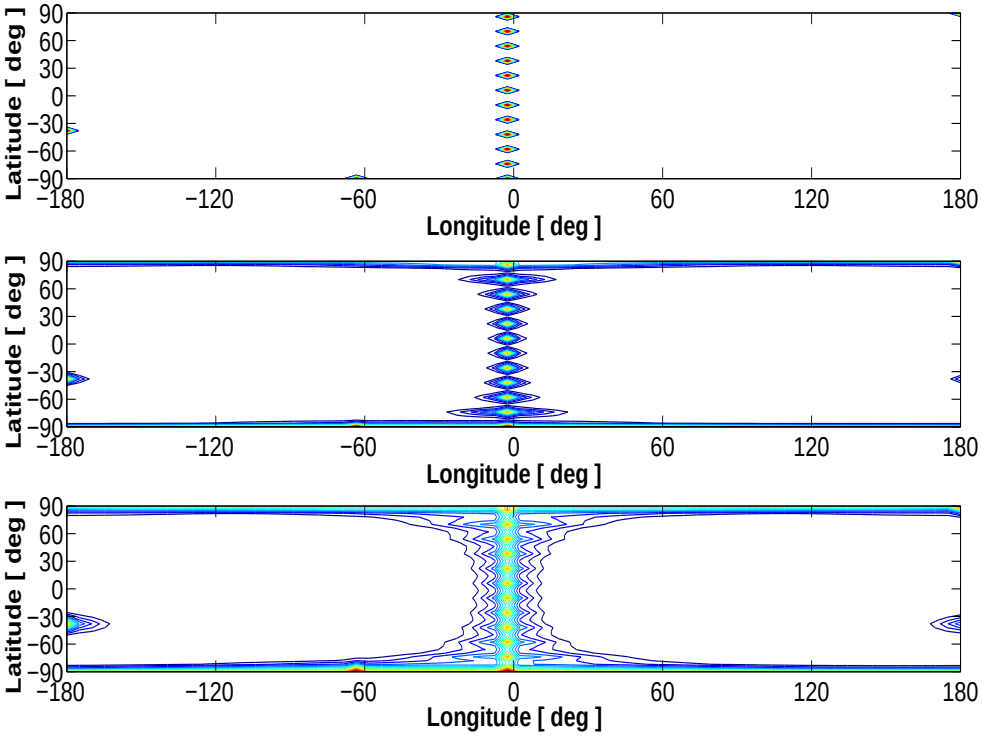


Figure 4.2: Contour lines of the longitudinal correlation $\mathbb{C}_x$ for points at different latitudes. The correlation length ℓ_x is short (top panel), medium (middle panel), and large (bottom panel). Note that the same correlation length ℓ_x translates into a different number of correlated grid points depending on the latitude.

4.2.2 Two-dimensional covariance matrices

Formally the full background error correlation matrix $\mathbb{C} \in \mathbb{R}^{n_x n_y \times n_x n_y}$ (which accounts for both latitudinal and longitudinal correlations) is constructed via the following relation

$$\mathbb{C} = \left(\mathbb{C}_x^{1:n_y} \otimes \mathbb{I}_{n_y \times n_y} \right) \cdot \left(\mathbb{I}_{n_x \times n_x} \otimes \mathbb{C}_y \right). \tag{4.6}$$

Note that this (huge) matrix is never explicitly formed. One needs to form and store only n_y one-dimensional longitudinal correlation matrices (4.2) and a single one-dimensional latitudinal correlation matrix (4.3). Note that the diagonal entries of the tensor product matrix (4.6) are all equal to one. The tensor product matrix (4.6) is not symmetric.

A symmetric version of the two-dimensional correlation matrix can be constructed as follows. Any symmetric positive definite matrix $\mathbb{C}$ has a matrix square root $\mathbb{C}^{1/2}$ such that

$$\mathbb{C} = \mathbb{C}^{1/2}\,\mathbb{C}^{T/2}.$$

The matrix square root is not uniquely defined; in particular it can be symmetric or not depending on the decomposition method used as described in equations (4.10) and (4.11). Let $\left(\mathbb{I}_{n_x\times n_x}\otimes\mathbb{C}_y\right)^{1/2}$ be a square root of the longitudinal correlation matrix. The symmetric two-dimensional correlation matrix can be constructed as:

$$\mathbb{C}^{\text{sym}} = \left(\mathbb{I}_{n_x\times n_x}\otimes\mathbb{C}_y\right)^{1/2}\cdot\left(\mathbb{C}_x^{1:n_y}\otimes\mathbb{I}_{n_y\times n_y}\right)\cdot\left(\mathbb{I}_{n_x\times n_x}\otimes\mathbb{C}_y\right)^{T/2}. \tag{4.7}$$

Let $\sigma_{i,j}$ be the standard deviation of the error at location (x_i, y_j) and

$$\Sigma = \text{diag}_{1\le i\le n_x, 1\le j\le n_y}\,\sigma_{i,j}$$

the diagonal matrix with all standard deviations at all grid points ordered according to (4.1). The two-dimensional covariance matrix is constructed from the correlation matrix (4.6) by scaling it from left and right with the diagonal matrix of standard deviations Σ

$$\mathbb{B} = \Sigma\cdot\mathbb{C}\cdot\Sigma = \Sigma\cdot\left(\mathbb{C}_x^{1:n_y}\otimes\mathbb{I}_{n_y\times n_y}\right)\cdot\left(\mathbb{I}_{n_x\times n_x}\otimes\mathbb{C}_y\right)\cdot\Sigma. \tag{4.8}$$

Similarly, a symmetric version of the covariance matrix can be constructed from the symmetric correlation (4.7) as follows:

$$\mathbb{B}^{\text{sym}} = \Sigma\cdot\mathbb{C}^{\text{sym}}\cdot\Sigma = \Sigma\cdot\left(\mathbb{I}_{n_x\times n_x}\otimes\mathbb{C}_y\right)^{1/2}\left(\mathbb{C}_x^{1:n_y}\otimes\mathbb{I}_{n_y\times n_y}\right)\cdot\left(\mathbb{I}_{n_x\times n_x}\otimes\mathbb{C}_y\right)^{T/2}\cdot\Sigma. \tag{4.9}$$

4.2.3 Efficient covariance matrix function calculations

The symmetric positive definite one-directional longitudinal correlation matrix has a matrix square root $\mathbb{C}_y^{1/2}$. A symmetric square root, the inverse of the symmetric square root, and the matrix inverse can be obtained via the singular value decomposition (SVD)

$$\mathbb{C}_y = \mathbb{U}\,\Theta\mathbb{U}^T,\quad \left(\mathbb{C}_y\right)^r = \mathbb{U}\,\Theta^r\mathbb{U}^T,\quad \text{for } r\in\left\{\frac{1}{2},-\frac{1}{2},-1\right\}. \tag{4.10}$$

while a nonsymmetric square root can be obtained via a Cholesky decomposition

$$\mathbb{C}_y = \mathbb{L}_y\,\mathbb{L}_y^T,\quad \mathbb{C}_y^{1/2} = \mathbb{L}_y. \tag{4.11}$$

By the properties of the Kronecker product we have that the square root, the inverse square root, and the inverse of the two-dimensional longitudinal correlation matrix can

be constructed in terms of the same matrix functions applied to the one dimensional longitudinal correlations:

$$\left(\mathbb{I}_{n_x\times n_x}\otimes\mathbb{C}_y\right)^r=\mathbb{I}_{n_x\times n_x}\otimes\left(\mathbb{C}_y\right)^r\quad\text{for }\ r\in\left\{\frac{1}{2},-\frac{1}{2},-1\right\}. \tag{4.12}$$

Consequently, the symmetric covariance (4.9) can be implemented as

$$\mathbb{B}^{\text{sym}}=\Sigma\cdot\left(\mathbb{I}_{n_x\times n_x}\otimes\left(\mathbb{C}_y\right)^{1/2}\right)\cdot\left(\mathbb{C}_x^{1:n_y}\otimes\mathbb{I}_{n_y\times n_y}\right)\cdot\left(\mathbb{I}_{n_x\times n_x}\otimes\left(\mathbb{C}_y\right)^{T/2}\right)\cdot\Sigma$$

using either of the one-dimensional square roots.

Similarly, different powers of each one-dimensional latitudinal correlation matrix can be obtained via a singular value decomposition:

$$\mathbb{C}_x^{1:n_y}=\mathbb{V}\,\Gamma\,\mathbb{V}^T,\quad\left(\mathbb{C}_x^{1:n_y}\right)^r=\mathbb{V}\,\Gamma^r\,\mathbb{V}^T,\quad\text{for }\ r\in\left\{\frac{1}{2},-\frac{1}{2},-1\right\}.$$

By the properties of the extended Kronecker product we have that the square root, the inverse square root, and the inverse of the two-dimensional latitudinal correlation matrix can be constructed in terms of the same matrix functions applied to the one dimensional latitudinal correlations:

$$\left(\mathbb{C}_x^{1:n_y}\otimes\mathbb{I}_{n_y\times n_y}\right)^r=\left(\mathbb{C}_x^{1:n_y}\right)^r\otimes\mathbb{I}_{n_y\times n_y}\quad\text{for }\ r\in\left\{\frac{1}{2},-\frac{1}{2},-1\right\}. \tag{4.13}$$

We now use these relations to build functions of the covariance matrices. The inverse of the background covariance is needed in the formulation of the variational cost function. The inverse of the non-symmetric covariance (4.8) is

$$\mathbb{B}^{-1}=\Sigma^{-1}\cdot\left(\mathbb{I}_{n_x\times n_x}\otimes\left(\mathbb{C}_y\right)^{-1}\right)\cdot\left(\left(\mathbb{C}_x^{1:n_y}\right)^{-1}\otimes\mathbb{I}_{n_y\times n_y}\right)\cdot\Sigma^{-1}. \tag{4.14}$$

The inverse of the symmetric covariance (4.9) matrix is

$$(\mathbb{B}^{\text{sym}})^{-1}=\Sigma^{-1}\cdot\left(\mathbb{I}_{n_x\times n_x}\otimes\left(\mathbb{C}_y\right)^{-T/2}\right)\left(\left(\mathbb{C}_x^{1:n_y}\right)^{-1}\otimes\mathbb{I}_{n_y\times n_y}\right)\cdot\left(\mathbb{I}_{n_x\times n_x}\otimes\left(\mathbb{C}_y\right)^{-1/2}\right)\cdot\Sigma^{-1}. \tag{4.15}$$

Finally, the symmetric covariance (4.9) has a (non-symmetric) matrix square root

$$(\mathbb{B}^{\text{sym}})^{1/2}=\Sigma\cdot\left(\mathbb{I}_{n_x\times n_x}\otimes\left(\mathbb{C}_y\right)^{1/2}\right)\cdot\left(\left(\mathbb{C}_x^{1:n_y}\right)^{1/2}\otimes\mathbb{I}_{n_y\times n_y}\right). \tag{4.16}$$

This is built out of tensor products involving the square roots of the one-dimensional correlation matrices. The inverse of the square root matrix (4.16) is

$$(\mathbb{B}^{\text{sym}})^{-1/2}=\left(\left(\mathbb{C}_x^{1:n_y}\right)^{-1/2}\otimes\mathbb{I}_{n_y\times n_y}\right)\cdot\left(\mathbb{I}_{n_x\times n_x}\otimes\left(\mathbb{C}_y\right)^{-1/2}\right)\cdot\Sigma^{-1}. \tag{4.17}$$

4.2.4 Efficient linear algebra operations involving the covariance matrix

Matrix vector operations involving $\mathbb{B}$ can be performed effectively by exploiting its structure. Consider a vector of concentrations (or concentration errors) $u_{i,j}$ - indexed by latitude and longitude, but stored as a state vector with the convention (4.1).

Consider the non-symmetric covariance matrix. The covariance matrix-vector product $v = \mathbb{B} \cdot u$ can be computed in stages. Each stage produces a temporary result which is a two-dimensional vector.

Expression	Computation	
$\alpha = \Sigma \cdot u$	$\alpha_{i,j} = \sigma_{i,j}\, u_{i,j},$	for $i = 1, \cdots, n_x, j = 1, \cdots, n_y.$
$\beta = \left(\mathbb{I}_{n_x \times n_x} \otimes \mathbb{C}_y\right) \cdot \alpha$	$\beta_{i,1:n_y} = \mathbb{C}_y \cdot \alpha_{i,1:n_y},$	for $i = 1, \cdots, n_x.$
$\gamma = \left(\mathbb{C}_x^{1:n_y} \otimes \mathbb{I}_{n_y \times n_y}\right) \cdot \beta$	$\gamma_{1:n_x,j} = \mathbb{C}_x^j \cdot \beta_{1:n_x,j},$	for $j = 1, \cdots, n_y.$
$v = \Sigma \cdot \gamma$	$v_{i,j} = \sigma_{i,j}\, \gamma_{i,j},$	for $i = 1, \cdots, n_x, j = 1, \cdots, n_y.$

(4.18)

Similarly the inverse covariance matrix-vector product $v = \mathbb{B}^{-1} \cdot u$ can be computed as follows

Expression	Computation	
$\alpha = \Sigma^{-1} \cdot u$	$\alpha_{i,j} = u_{i,j}/\sigma_{i,j},$	for $i = 1, \cdots, n_x, j = 1, \cdots, n_y.$
$\beta = \left(\mathbb{C}_x^{1:n_y} \otimes \mathbb{I}_{n_y \times n_y}\right)^{-1} \cdot \alpha$	Solve$\mathbb{C}_x^j \cdot \beta_{1:n_x,j} = \alpha_{1:n_x,j},$	for $j = 1, \cdots, n_y.$
$\gamma = \left(\mathbb{I}_{n_x \times n_x} \otimes \mathbb{C}_y\right)^{-1} \cdot \beta$	Solve$\mathbb{C}_y \cdot \gamma_{i,1:n_y} = \beta_{i,1:n_y},$	for $i = 1, \cdots, n_x.$
$v = \Sigma^{-1} \cdot \gamma$	$v_{i,j} = \gamma_{i,j}/\sigma_{i,j},$	for $i = 1, \cdots, n_x, j = 1, \cdots, n_y.$

(4.19)

Similar procedures can be developed for the symmetric covariance time vector products.

The square root (4.16) times vector product $v = \left(\mathbb{B}^{\text{sym}}\right)^{1/2} u$ is computed as:

Expression	Computation	
$\alpha = \left(\left(\mathbb{C}_x^{1:n_y}\right)^{1/2} \otimes \mathbb{I}_{n_y \times n_y}\right) \cdot u$	$\gamma_{1:n_x,j} = \mathbb{C}_x^j \cdot \beta_{1:n_x,j},$	for $j = 1, \cdots, n_y$
$\beta = \left(\mathbb{I}_{n_x \times n_x} \otimes \left(\mathbb{C}_y\right)^{1/2}\right) \cdot \alpha$	$\beta_{i,1:n_y} = \mathbb{C}_y^{1/2} \cdot \alpha_{i,1:n_y},$	for $i = 1, \cdots, n_x$
$v = \Sigma \cdot \beta$	$v_{i,j} = \sigma_{i,j}\, \gamma_{i,j},$	for $i = 1, \cdots, n_x, j = 1, \cdots, n_y.$

(4.20)

All the above implementations are based on repeated operations involving the one-dimensional covariance matrices and their square roots. These operations are very efficient since all the linear algebra operations (matrix-vector multiplication, SVD, Cholesky factorization, the solution of linear systems) are performed on small dimensional matrices ($n_x \times n_x$ or $n_y \times n_y$).

4.3 Numerical experiments

We employ GEOS-Chem v7-04-10 adjoint code [Singh et al., 2009b], capable of assimilating Tropospheric Emission Spectormeter (TES) ozone profile retrievals into the model through 3D-Var and 4D-Var data assimilation systems. The generated analyses were compared against direct ozone profile measurements from Ozonesondes. A detailed discussion on GEOS-Chem and its adjoint construction is provided in Chapter 2, while details about the TES instrument and ozonesonde measurements could be found in Chapter 3.

4.3.1 Experimental setting

We have used $4^\circ \times 5^\circ$ resolution in all our experiments. There are 46 x 72 latitude-longitude grid boxes, and 55 vertical levels at this resolution. The data assimilation was performed for only the first 23 model levels (for up to about 50 hPa). The code to calculate full rank covariance matrices was interfaced with the data assimilation systems, and the dependent parameters such as correlation lengths were adjusted through a run script. In order to demonstrate the benefits of including spatial correlations through background error covariance matrices, we used correlations lengths of 0 Km, 500 Km, $1,000$ Km, and $1,500$ Km.

The 3D-Var data assimilation experiments were performed over the months of July and August 2006, starting at 00:00(GMT) on July 1st, while 4D-Var assimilation experiments were performed over a 5 day assimilation window starting at 00:00(GMT) on August 1st, 2006 and ending at 00:00(GMT) on August 6th of the same year. The TES satellite data was read once every 4 simulation hours. A detailed description on the set up of these assimilation systems and the diagonal background error covariance matrices is provided in Chapter 3.

Computational costs

As described in Section 4.1, the construction of the background error covariance matrix $\mathbb{B}$ impacts the result of the data assimilation. If one considers no correlation among differ-

ent model grid points, or among different chemical species, $\mathbb{B}$ turns out to be diagonal. However, such approximations are inaccurate as the ozone errors are highly correlated spatially [Constantinescu et al., 2007a,c,d] and correlated to errors in other chemical species; this inter-species correlation is not discussed in this work. In Section 4.2, we have introduced an efficient methodology to construct a non-diagonal background error covariance matrix, $\mathbb{B}$. Its inverse, $\mathbb{B}^{-1}$, needed in 3D-Var (6.8) and in 4D-Var (6.10) cost function formulations, can be obtained either via a Cholesky decomposition or via a singular value decomposition. (Note that by the "computation of the inverse" we mean the solution of a linear system).

Table 4.1 illustrates the computational cost of data assimilation compared to the cost of free running model for a 24 hour simulation. All the simulations are performed on a Dell Precision T5400 workstation with 2 quadcore Intel(R) Xeon(R) processors with clock speed 2.33GHz and a RAM of 16GB shared between the two processors.

Table 4.1: Timing results for GEOS-Chem free model run, 3D-Var and 4D-Var data assimilations with diagonal and non-diagonal $\mathbb{B}$ for a 24 hour simulation starting July 1st, 2006.

Experiment Description	CPU Time
Free model run, SMVGEAR chemistry solver	2 min 50 sec
Free model run, KPP chemistry solver	3 min 18 sec
3D-Var with diagonal $\mathbb{B}$	3 min 57 sec
3D-Var with non-diagonal $\mathbb{B}$, Cholesky	4 min 00 sec
3D-Var with non-diagonal $\mathbb{B}$, SVD	9 min 38 sec
4D-Var with diagonal $\mathbb{B}$ (per model run)	16 min 51 sec
4D-Var with non-diagonal $\mathbb{B}$, Cholesky (per model run)	16 min 51 sec

Performing 3D-Var with a non-diagonal error covariance matrix whose inverse is computed by Cholesky decomposition is only about 1.4% more expensive than the 3D-Var with a diagonal covariance matrix, and this does not vary with changes in correlation lengths. The inverse of $\mathbb{B}$ calculation via the Cholesky decomposition is considerably more efficient than the calculation via a singular value decomposition, as expected. The 4D-Var assimilation is more expensive than the 3D-Var. The use of the non-diagonal $\mathbb{B}$ (with Cholesky decomposition) in 4D-Var causes a minimal to zero increase in the computational time when compared to the diagonal $\mathbb{B}$ case.

4.3.2 Impact of non-diagonal background error covariance in 3D-Var assimilation

We first compare the tropospheric ozone concentrations generated through 3D-Var assimilations using various correlation lengths against the ozonesonde observations. The left panel of Figure 4.3 shows the forecast, the analysis and the ozonesonde ozone concentrations averaged over all ozonesonde launches in August 2006. The model ozone fields are interpolated to the space-time location of each ozonesonde launch for comparison. The center panel of Figure 4.3 shows the mean relative errors in model predicted ozone concentrations (the relative differences between the forecast/analysis profiles and the ozonesonde profiles), averaged over all ozonesonde launches. The rightmost panel provides an estimate of the variability of ozonesonde against the variability of ozone concentration predicted through different assimilation techniques.

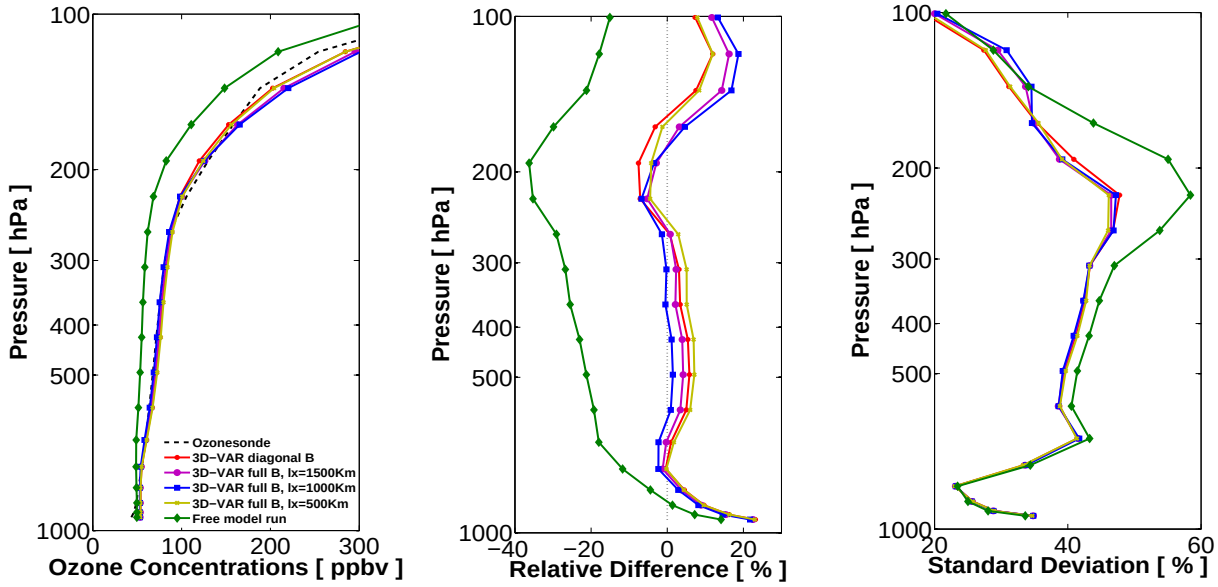


Figure 4.3: The impact of non-diagonal background error covariances in 3D-Var data assimilation. Left panel: mean ozone concentrations at ozonesonde locations for 3D-Var analyses and free model trajectories. Center panel: relative mean errors of predicted ozone concentrations with respect to ozonesonde measurements. Right panel: standard deviation of absolute values of errors with respect to ozonesonde measurements. The data is averaged over all ozonesonde launches. These plots were generated from 2 months simulation from 00:00 GMT July 1st to 23:00 GMT August, 2006 and compared against ozonesonde data available for the month of August.

In all our experiments, correlation lengths in latitudinal direction varied in proportion with correlation lengths in longitudinal direction. A value of 500 for ℓ_x implicitly in-

dicates ℓ_y is 400, and refers to correlation between two neighboring grid boxes both in East/West and North/South directions.

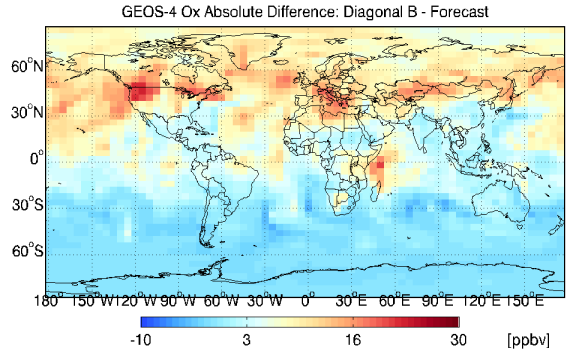
The results indicate that 3D-Var is sensitive to the correlation length used in the construction of the background error covariance matrix (a zero correlation length corresponds to a diagonal matrix). Note that the assimilation results using a non-diagonal $\mathbb{B}$ with higher correlation length are superior to those using a diagonal $\mathbb{B}$ in the lower and mid troposphere. Above 180 hPa, however, the errors in the assimilated ozone fields are larger for the non-diagonal case. This could be attributed to the fact that a uniform correlation length across all vertical levels is only a very coarse approximation of the real error correlations. Higher correlation lengths might be smearing off the ozone in the upper troposphere leading to an overestimate.

To further understand the effect of using non-diagonal background error covariance matrices in 3D-Var we consider the corrections obtained with different correlation lengths (i.e., the differences between the assimilated ozone fields and forecast, or the non-assimilated ozone fields). Panels (a)-(d) of Figure 4.4 show the global spatial distribution plots of these differences. The assimilation with non-diagonal covariance matrices generate much smoother analyses; note that the point-wise values of the increments is smaller, and that the corrections are distributed over larger areas. Panels (e)-(f) of Figure 4.4 compare directly the 3D-Var analyses obtained using a diagonal $\mathbb{B}$ and a non-diagonal $\mathbb{B}$ with a correlation length of 1000 Km. The corrections in the non-diagonal case are spread are less aggressive and smoother.

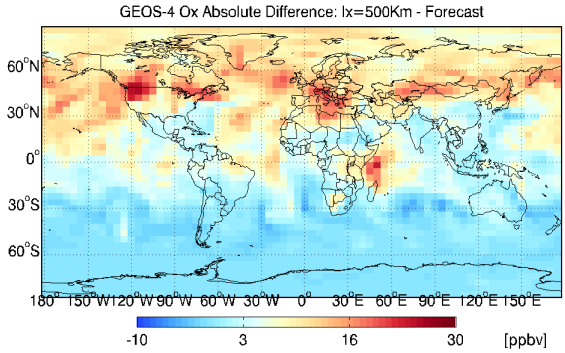
4.3.3 Impact of non-diagonal background error covariance in 4D-Var assimilation

We now study the effects of using non-diagonal background error covariance matrix in 4D-Var data assimilation. We compare the analyses ozone concentrations generated by 4D-Var with different background error correlation lengths against the ozonesonde observations. The left panel of Figure 4.5 shows the forecast, analysis and ozonesonde measured ozone concentrations averaged over the two months assimilation window. The model ozone fields are interpolated to the space-time location of each ozonesonde launch for comparison. The center panel shows the relative errors of model predictions with respect to ozonesonde data, averaged over all ozonesonde launches. The right panel provides the standard deviations of these errors.

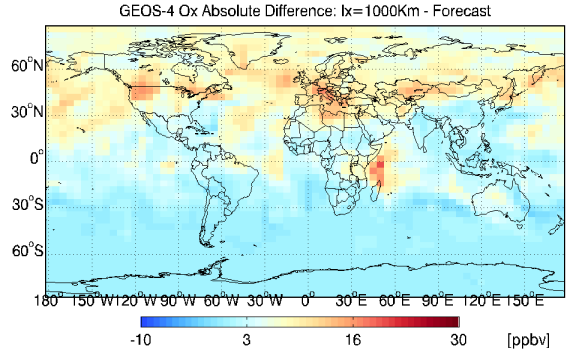
The results indicate that 4D-Var is also sensitive to the structure of the background error covariance matrix. The use of non-diagonal correlations leads to improved analyses. The best analysis is obtained with a correlation length of 500 Km (about one grid cell near the equator). Note that 4D-Var accounts for all the data available within the assimilation window.



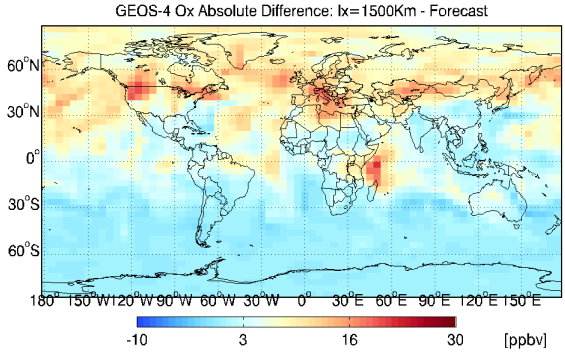
(a) Absolute difference between the 3D-Var analysis using diagonal $\mathbb{B}$ ($\ell_x = 0\ Km$) and the free model run



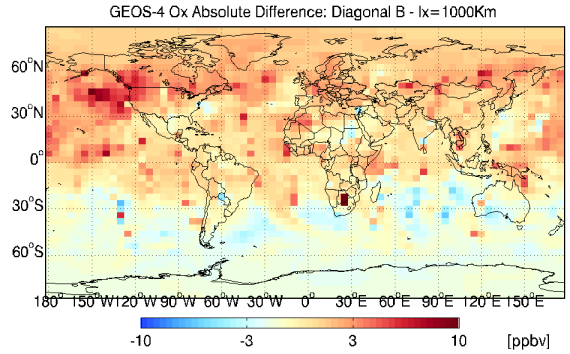
(b) Absolute difference between the 3D-Var analysis using non-diagonal $\mathbb{B}$ ($\ell_x = 500\ Km$) and the free model run



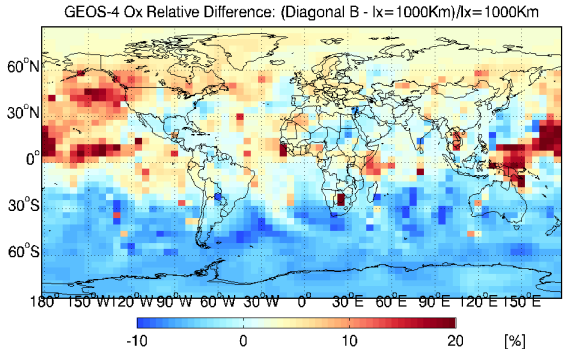
(c) Absolute difference between the 3D-Var analysis using non-diagonal $\mathbb{B}$ ($\ell_x = 1000\ Km$) and the free model run



(d) Absolute difference between the 3D-Var analysis using non-diagonal $\mathbb{B}$ ($\ell_x = 1500\ Km$) and the free model run



(e) Absolute difference between the 3D-Var analyses using diagonal $\mathbb{B}$ and non-diagonal $\mathbb{B}$ ($\ell_x = 1000\ Km$)



(f) Relative difference between the 3D-Var analyses using diagonal $\mathbb{B}$ and non-diagonal $\mathbb{B}$ ($\ell_x = 1000\ Km$)

Figure 4.4: Differences in global ozone concentrations at 23:00 GMT on August 31, 2006 averaged over the first 10 GEOS-Chem vertical levels. Panels (a)-(d): differences between the 3D-Var analysis fields and the model forecast (solution without data assimilation); the analyses use different correlation lengths between 0 Km and $1,500\ Km$. Panels (e)-(f): absolute and relative differences between 3D-Var analyses using diagonal and non-diagonal background covariance matrices.

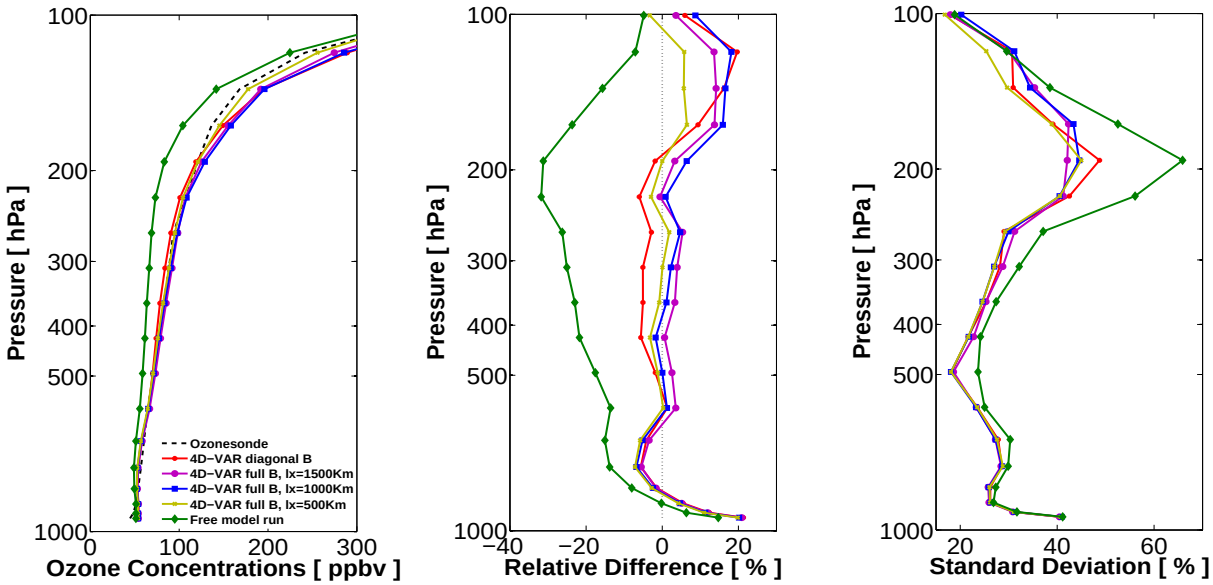
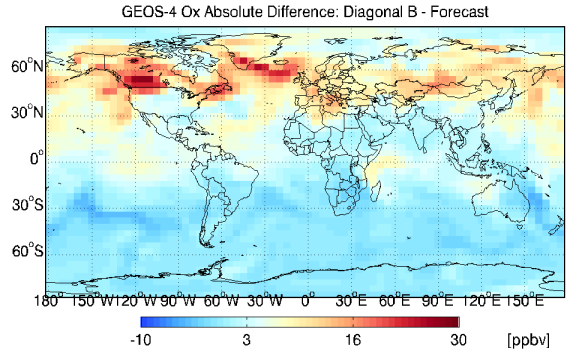


Figure 4.5: The impact of non-diagonal background error covariances on 4D-Var data assimilation. The results shown are for a single 5-day assimilation window from 00:00 GMT August 1st to 00:00 GMT August 6th, 2006. Left panel: mean ozone concentrations at ozonesonde locations for 4D-Var analyses and free model trajectories. Center panel: relative mean errors of predicted ozone concentrations with respect to ozonesonde measurements. Right panel: standard deviation of absolute values of errors with respect to ozonesonde measurements.

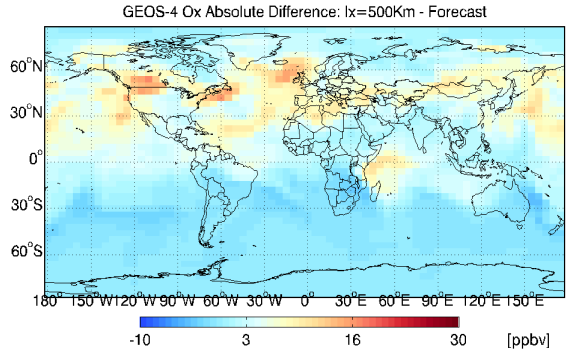
To better understand the impact of different background error correlation lengths on 4D-Var assimilation, we present in Figure 4.6 the differences in ozone concentrations generated by the free model run and by the 4D-Var assimilation using diagonal and non-diagonal $\mathbb{B}$. The use of a non-diagonal $\mathbb{B}$ with a properly-chosen correlation length not only provides a better estimate but also helps generate a smoother analysis. The panels (e)-(f) of Figure 4.6 compare directly the 4D-Var analyses obtained using a diagonal $\mathbb{B}$ and a non-diagonal $\mathbb{B}$ with a correlation length of 500 *Km*; the large localized corrections over North America provided by the diagonal $\mathbb{B}$ are smoothened out when the non-diagonal $\mathbb{B}$ is employed.

4.3.4 Determining the correlation length through experiments

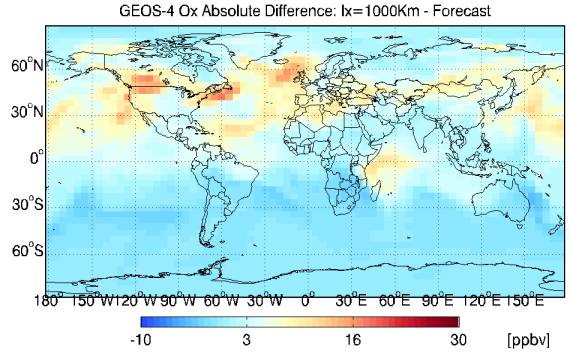
The correlation length is a very important parameter that impacts the quality of the assimilation when using non-diagonal error covariance matrix. The value of the correlation length depends on various factors such as the lifetime of the tracer under consideration,



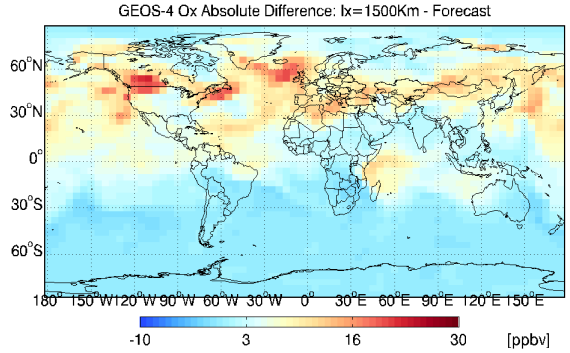
(a) Absolute difference between the 4D-Var analysis using diagonal $\mathbb{B}$ ($\ell_x = 0$ Km) and the free model run



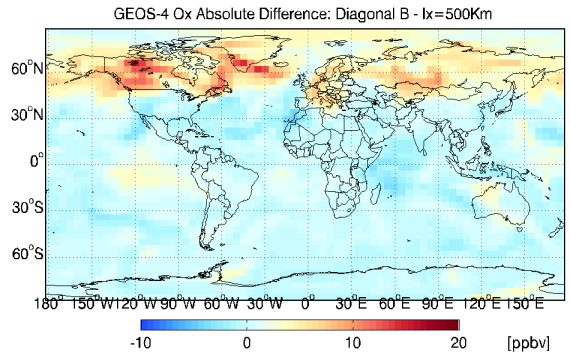
(b) Absolute difference between the 4D-Var analysis using non-diagonal $\mathbb{B}$ ($\ell_x = 500$ Km) and the free model run



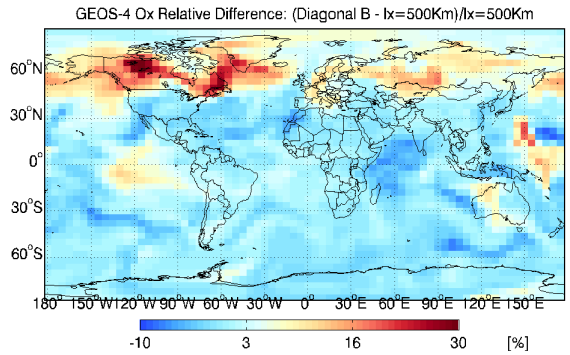
(c) Absolute difference between the 4D-Var analysis using non-diagonal $\mathbb{B}$ ($\ell_x = 1000$ Km) and the free model run



(d) Absolute difference between the 4D-Var analysis using non-diagonal $\mathbb{B}$ ($\ell_x = 1500$ Km) and the free model run

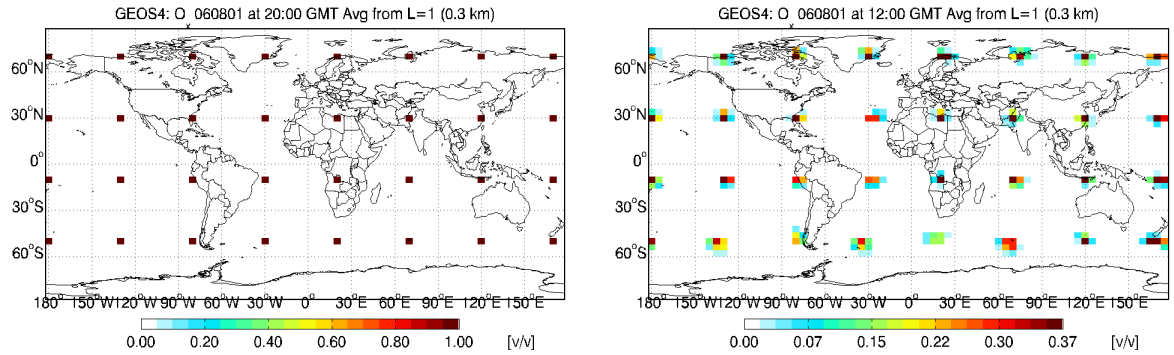


(e) Absolute difference between the 4D-Var analyses using diagonal $\mathbb{B}$ and non-diagonal $\mathbb{B}$ ($\ell_x = 500$ Km)



(f) Relative difference between the 4D-Var analyses using diagonal $\mathbb{B}$ and non-diagonal $\mathbb{B}$ ($\ell_x = 500$ Km)

Figure 4.6: Differences in global ozone concentrations at 00:00 GMT on August 06, 2006 (end of assimilation window) averaged over the first 10 GEOS-Chem vertical levels. Panels (a)-(d): differences between the 4D-Var analysis fields and the model forecast (solution without data assimilation); the analyses use different correlation lengths between 0 Km and $1,500$ Km. Panels (e)-(f): absolute and relative differences between 4D-Var analyses using diagonal and non-diagonal background covariance matrices.



(a) Adjoint ozone variables are initialized to one in selected grid cells

(b) Adjoint ozone variables spread after 8 hours of backward sensitivity run

Figure 4.7: Ground level ozone adjoint variable values are initialized to one on July 1st 2006, 20:00 GMT, every tenth grid point in longitudinal and latitudinal directions. An 8 hour backward adjoint integration spreads the adjoint fields, and helps identify grid cells where ozone errors are correlated.

the grid resolution, the pressure level, and the wind velocity. We propose a method to determine experimentally a value of the correlation length that is appropriate for the model and data at hand.

Recall the construction of one dimensional correlation matrices in equations (4.2) and (4.3). Our aim is to determine the number of grid cells (in each direction) where the errors are correlated. For this we use adjoint sensitivity analysis. Specifically, we initialize the adjoint variable to 1 in a specific cell at the end of a given window (and to zero everywhere else), perform a backwards adjoint simulation, and analyze the adjoint fields at the beginning of the window. The error in the specific cell is correlated with errors in those grid cells where the adjoint values are above $1/e$. The length of the time window depends on the time scale of the model under consideration.

Here we consider a time window of 8 hours. We run the forward GEOS-Chem model starting at July 1st, 00:00 GMT for 20 hours. The adjoint variable for ozone at 20:00 GMT are initialized to 1 in a subset of the grid points ((i, j) chosen such that $i \mod 10 = 1$ and $j \mod 10 = 1$). Adjoint variables for all other grid points and species are initialized to zero. The gap in the initialization helps avoid the interactions between adjoint "plumes" initialized at different locations. The ozone adjoint variable field is analyzed at 12:00 GMT to find out the number of grid cells where the value is greater than or equal to $1/e$. In our current setup, we use the same correlation length for all pressure levels and thus consider the spread only at ground level.

The procedure can be easily extended to considering different correlation lengths for different vertical levels and for different geographic areas.

Figure 4.7 shows that the ozone adjoint variables have spread (on average) over one to two grid cells in both longitudinal and latitudinal direction. For the $4^\circ \times 5^\circ$ model resolution each grid box is of size $400Km \times 500Km$ near the equator. The adjoint sensitivity analysis indicates that the correlation lengths should be chosen in the ranges of $\ell_x \in [500Km, 1000Km]$ and $\ell_y \in [400Km, 800Km]$ respectively. This confirms the best correlation lengths empirically observed in the data assimilation results reported in Figures 4.3 and 4.5.

4.4 Conclusion

This chapter presents an efficient methodology to construct non-diagonal background error covariance matrices for data assimilation. The two- or three- dimensional covariance matrices are not formed explicitly. Rather, multi-dimensional correlations are represented by tensor products of one dimensional correlation matrices along longitudinal and latitudinal directions. The technique can be easily extended to include correlations in the vertical direction as well. Highly efficient linear algebra operations are obtained by performing successive matrix-vector products, Cholesky decompositions, etc. with one-dimensional correlation matrices. The correlation lengths are important parameters that need to be specified for each directional correlation. We propose an adjoint sensitivity analysis approach to guide the choice of proper correlation lengths; the approach implicitly accounts for factors such as chemical activity, grid resolution, etc.

The approach to construct non-diagonal covariance matrices has been tested using the 3D-Var and 4D-Var data assimilation frameworks developed for GEOS-Chem. The experiments assimilate observations from TES satellite ozone profile retrievals, and validate the results against an independent data set provided by IONS ozonesondes. The change of the covariance matrix formulation in data assimilation from diagonal to non-diagonal adds only a negligible computational overhead. In the same time, the inclusion of spatial correlations leads consistently to improved analyses in both the 3D-Var and the 4D-Var settings.

Chapter 5

A Practical Method to Estimate Information Content in the Context of 4D-VAR Data Assimilation

Abstract

Data assimilation obtains improved estimates of the state of a physical system by combining imperfect model results with sparse and noisy observations of reality. Not all observations used in data assimilation are equally valuable. The ability to characterize the usefulness of different data points is important for analyzing the effectiveness of the assimilation system, for data pruning, and for the design of future sensor systems.

This chapter focuses on the four dimensional variational (4D-Var) data assimilation framework. Metrics from information theory are used to quantify the contribution of observations to decreasing the uncertainty with which the system state is known. We establish an interesting relationship between different information-theoretic metrics and the variational cost function/gradient under Gaussian linear assumptions. Based on this insight we derive an ensemble-based computational procedure to estimate the information content of various observations in the context of 4D-Var. The approach is illustrated on a global chemical data assimilation problem using satellite observations and the GEOS-Chem chemical transport model.

5.1 Introduction

The ability to characterize the usefulness of different data points in data assimilation is important for analyzing the effectiveness of the assimilation system, for data prun-

ing/data selection, for the design of future sensor systems, and for defining strategies for targeting observations. In order to quantify the contribution of observations to the improvements in state estimate obtained through data assimilation we employ metrics from information theory. Broadly speaking, the information content of a message in information theory describes the amount of novelty brought in by the message. Information theory has started in electrical engineering and has been applied to diverse areas as complexity theory, networking analysis, financial mathematics and mathematical statistics.

In the context of data assimilation the information content of observations is loosely defined by their contribution to decreasing the uncertainty in the state estimate [Fisher, 1922]. Several of the information theoretic metrics employed here measure the decrease in the (co-)variance of the error (the trace of the Fisher information matrix, the Shannon information, and the degrees of freedom for signal). Others measure the benefit of data assimilation in terms of adjusting the mean of the distribution (the signal information). Relative entropy offers a combination of both mean and variance effects.

Information theory has been used in atmospheric sciences for uncertainty studies, instrument development, and data selection. [Abramov, 2004; Majda, 2006] propose the use the relative entropy to quantify the lack of information in climate systems; their approach is applicable to non-Gaussian distributions and non-linear models. They demonstrate the methodology with two "toy" models, Burgers-Hopf [Lorenz, 1996]; the approach becomes computationally intractable for real large scale models. Information theoretic metrics like the entropy reduction and the degrees of freedom for signal are being used in the development of remote-sounding instruments [Rodgers, 1996, 1998, 2000; Rabier et al., 2002; Worden et al., 2004]. Data selection strategies were defined using information theory [Rabier et al., 2002].

The information theory has recently been used in data assimilation to characterize the information content of various observations (i.e., the usefulness of these observations). [Fisher, 2003] proposes methods to estimate the entropy reduction and degrees of freedom for signal with large variational analysis systems. [Cardinali et al., 2004] study the influence-matrix diagnostic of data assimilation systems.[Xu, 2006] analyses the relative entropy versus Shannon entropy difference to measure information content from observations for data assimilation. [Zupanski, 2009] discusses the use of information measures in ensemble data assimilation.

Here we discuss a characterization of the information content of observations in the context of four dimensional variational (4D-Var) data assimilation framework. The analysis carried out in this chapter assumes that errors are normally distributed and that the model dynamics is linear. It is shown that, under these assumptions, the posterior statistics of the variational cost function and its gradient can be used to quantify the information content of observations. This results leads to the following computational procedure. After data assimilation is complete, an ensemble of simulations is run with

the initial conditions drawn from (an approximation of) the analysis probability distribution. Mean values of the cost function and of adjoint norms are used to estimate the information content of various observations in the context of 4D-Var. Note that all information metrics obtained here are with respect to the beginning of each assimilation window (as 4D-Var provides the analysis in form of the model initial conditions). The impact of the assumptions on the accuracy of information estimates when the technique is applied to general systems is not discussed in this chapter, and will be addressed in future work.

The chapter is organized as follows. Section 5.2 reviews the variational approach to data assimilation from a Bayesian perspective. Various metrics for information content are discussed in Section 5.3. Section 5.4 develops computationally feasible estimation techniques for the information content of observations in the context of 4D-Var data assimilation; this is the main contribution of this work. The numerical results are presented and discussed in Section 5.5. Section 5.6 summarizes the findings of this work and points to future research directions.

5.2 Variational Data Assimilation

Variational methods solve the data assimilation problem in an optimal control framework [Courtier and Talagrand , 1987; LeDimet and Talagrand, 1986; Lions, 1971]. A detailed discussion on data assimilation methodologies is provided in Chapter 3. We reiterate it here from a statistical perspective.

Consider that the true state of the system $\mathbf{x}^t \in \mathbb{R}^n$ is unknown and needs to be estimated form the available information. In order to obtain an estimate of $\mathbf{x}^t$ *data assimilation combines three different sources of information*, as follows.

1. The background (prior) probability density $\mathcal{P}^{\mathrm{B}}(\mathbf{x})$ encapsulates our current knowledge of the true state of the system. Specifically, $\mathcal{P}^{\mathrm{B}}(\mathbf{x})$ describes the uncertainty with which one knows $\mathbf{x}^t$ at a given moment, before any (new) measurements are taken. The mean taken with respect to this probability density is denoted by

$$\mathbb{E}^{\mathrm{B}}[f] = \int f(\mathbf{x})\, \mathcal{P}^{\mathrm{B}}(\mathbf{x})\, d\mathbf{x} \; .$$

   The mean of the background distribution $\mathbf{x}^B = \mathbb{E}^{\mathrm{B}}[\mathbf{x}]$ is called the apriori, or the *background state*, and represents the current best estimate of the true state. The background estimation errors $\varepsilon^B = \mathbf{x}^B - \mathbf{x}^t$ are characterized by the *background error covariance matrix* $\mathbb{B} = \mathbb{E}^{\mathrm{B}}\left[\varepsilon^B \left(\varepsilon^B\right)^T\right] \in \mathbb{R}^{n \times n}$.

   A typical assumption is that the random background errors have a normal proba-

bility density, i.e.,

$$\begin{aligned} \varepsilon^B &= \mathbf{x}^B - \mathbf{x}^t \in \mathcal{N}(0, \mathbb{B}) \qquad (5.1) \\ \Leftrightarrow \quad \mathcal{P}^{\mathrm{B}}(\mathbf{x}) &= \mathcal{N}\left(\mathbf{x}^B, \mathbb{B}\right) \\ &= \frac{1}{\left((2\pi)^{n/2}\sqrt{\det \mathbb{B}}\right)} \exp\left(-\frac{1}{2}(\mathbf{x}-\mathbf{x}^B)^T \mathbb{B}^{-1}(\mathbf{x}-\mathbf{x}^B)\right) . \end{aligned}$$

With many nonlinear models this normality assumption is difficult to justify, but is nevertheless widely used because of its convenience.

2. The model $\mathcal{M}$ encapsulates our knowledge about physical and chemical laws that govern the evolution of the system. The model evolves an initial state $\mathbf{x}_0 \in \mathbb{R}^n$ at the initial time t_0 to future state values $\mathbf{x}_i \in \mathbb{R}^n$ at future times t_i,

$$\mathbf{x}_i = \mathcal{M}_{t_0 \to t_i}\, \mathbf{x}_0. \qquad (5.2)$$

The size of the state space in realistic chemical transport models is very large. For example, a GEOS-Chem simulation at the $2^o \times 2.5^o$ horizontal resolution has $n \in \mathcal{O}\left(10^8\right)$ variables.

3. Observations represent snapshots of reality available at several discrete time moments. Specifically, measurements $\mathbf{y}_i \in \mathbb{R}^m$ of the true state are taken at times t_i, $i = 1, \cdots, N$

$$\mathbf{y}_i = \mathcal{H}\left(\mathbf{x}_i^t\right) - \eta_i^{\mathrm{obs}}. \qquad (5.3)$$

The observation operator $\mathcal{H}$ maps the state space onto the observation space. In many practical situations $\mathcal{H}$ is a highly nonlinear mapping (as is the case, e.g., with satellite observation operators). Usually the observations are sparsely distributed, and the number of observations is small compared to the dimension of the state space, $m \ll n$. The measurement (instrument) errors are denoted by η_i^{obs}.

Equation (6.2) relates the true state with the observations. In order to relate the model state to observations we also consider the relation

$$\mathbf{y}_i = \mathcal{H}\left(\mathbf{x}_i\right) - \varepsilon_i^{\mathrm{obs}}. \qquad (5.4)$$

where the observation operator now acts on the model predicted state. The *observation error* term $\varepsilon_i^{\mathrm{obs}}$ accounts for both the measurement (instrument) errors, as well as representativeness errors (i.e., errors in the accuracy with which the model can reproduce reality). Typically observation errors are assumed to be unbiased and normally distributed

$$\varepsilon_i^{\mathrm{obs}} \in \mathcal{N}(0, \mathbb{R}_i) \ , \quad i = 1, \cdots, N \ . \qquad (5.5)$$

Moreover, observation errors at different times ($\varepsilon_i^{\mathrm{obs}}$ and $\varepsilon_j^{\mathrm{obs}}$ for $i \neq j$) are assumed to be independent.

Based on these three sources of information data assimilation computes the analysis (posterior) probability density $\mathcal{P}^{\mathrm{A}}(\mathbf{x})$. Specifically, $\mathcal{P}^{\mathrm{A}}(\mathbf{x})$ describes the uncertainty with which one knows $\mathbf{x}^t$ after all the information available from measurements has been accounted for. The mean taken with respect to this probability density is denoted by

$$\mathbb{E}^{\mathrm{A}}[f] = \int f(\mathbf{x})\, \mathcal{P}^{\mathrm{A}}(\mathbf{x})\, d\mathbf{x} \,.$$

The mean of the analysis distribution $\mathbf{x}^A = \mathbb{E}^{\mathrm{A}}[\mathbf{x}]$ is called the aposteriori, or the *analysis state*, (In the maximum likelihood approach the refined estimate of the true state is obtained from the analysis distribution mode $\mathbf{x}^A = \arg\max \mathcal{P}^{\mathrm{A}}(\mathbf{x})$). The analysis estimation errors $\varepsilon^A = \mathbf{x}^A - \mathbf{x}^t$ are characterized by the *analysis error covariance matrix* $\mathbb{A} = \mathbb{E}^{\mathrm{A}}\left[\varepsilon^A \left(\varepsilon^A\right)^T\right] \in \mathbb{R}^{n\times n}$.

If both the the background and the observation errors are Gaussian, and the error propagation through the model (6.6) is linear, then he probability density of the analysis (estimation) errors ε^A is also Gaussian,

$$\varepsilon^A = \mathbf{x}^A - \mathbf{x}^t \in \mathcal{N}(0, \mathbb{A}) \quad \Leftrightarrow \quad \mathcal{P}^{\mathrm{A}}(\mathbf{x}) = \mathcal{N}\left(\mathbf{x}^A, \mathbb{A}\right) \,. \tag{5.6}$$

5.2.1 The Bayesian point of view to data assimilation

The estimation problem is posed in a Bayesian framework. The analysis probability density is the probability density of the state *conditioned by all the available observations* $\mathbf{y} = [\mathbf{y}_1, \cdots, \mathbf{y}_N]$. Bayes theorem allows to express the analysis probability density as follows:

$$\mathcal{P}^{\mathrm{A}}(\mathbf{x}) = \mathbb{P}(\mathbf{x}|\mathbf{y}) = \frac{\mathbb{P}(\mathbf{y}|\mathbf{x}) \cdot \mathcal{P}^{\mathrm{B}}(\mathbf{x})}{\mathbb{P}(\mathbf{y})}, \tag{5.7}$$

The denominator $\mathbb{P}(\mathbf{y})$ is the marginal probability density of the observations and plays the role of a scaling factor. The probability of the observations conditioned by the states $\mathbb{P}(\mathbf{y}|\mathbf{x})$ is the probability that the observation errors in (6.3) assume certain values. If the observation errors at different times are independent we have that:

$$\begin{aligned}
\mathbb{P}(\mathbf{y}|\mathbf{x}) &= \mathbb{P}\left(\varepsilon_1^{\mathrm{obs}}, \cdots, \varepsilon_N^{\mathrm{obs}}\right) \\
&= \prod_{i=1}^{N} \mathbb{P}\left(\varepsilon_i^{\mathrm{obs}}\right) \\
&= \prod_{i=1}^{N} \mathbb{P}\left(\mathcal{H}(\mathbf{x}_i) - \mathbf{y}_i\right)
\end{aligned}$$

If, in addition, the observation errors are Gaussian (6.4) we have that

$$\mathbb{P}\left(\mathbf{y}|\mathbf{x}\right) = \frac{1}{\left((2\pi)^{mN/2}\sqrt{\prod_{i=1}^{N}\det \mathbb{R}_i}\right)} \exp\left(-\frac{1}{2}\sum_{i=1}^{N}\left(\mathcal{H}\left(\mathbf{x}_i\right)-\mathbf{y}_i\right)^T \mathbb{R}_i^{-1}\left(\mathcal{H}\left(\mathbf{x}_i\right)-\mathbf{y}_i\right)\right) \tag{5.8}$$

In the maximum likelihood approach one looks for the argument that maximizes the posterior distribution, or, equivalently, minimizes its negative logarithm:

$$\mathbf{x}^A = \arg\max_{\mathbf{x}} \mathcal{P}^{\mathrm{A}}(\mathbf{x}) = \arg\min_{\mathbf{x}} \mathcal{J}(\mathbf{x}) \,, \quad \mathcal{J}(\mathbf{x}) = -\ln \mathcal{P}^{\mathrm{A}}(\mathbf{x}) \,. \tag{5.9}$$

In this context the data assimilation problem is formulated as an optimization problem. Using (5.7) the minimization cost function can be written as

$$\underbrace{-\ln \mathcal{P}^{\mathrm{A}}(\mathbf{x})}_{\mathcal{J}(\mathbf{x})} = \underbrace{-\ln \mathcal{P}^{\mathrm{B}}\left(\mathbf{x}\right)}_{\left(\mathcal{J}^{\mathrm{B}}(\mathbf{x})+\mathrm{const}\right)} + \underbrace{-\ln \mathbb{P}\left(\mathbf{y}|\mathbf{x}\right)}_{\left(\mathcal{J}^{\mathrm{obs}}(\mathbf{x})+\mathrm{const}\right)} - \underbrace{-\ln \mathbb{P}\left(\mathbf{y}\right)}_{(\mathrm{const})} \tag{5.10}$$

The minimization function has two terms: the first one ($\mathcal{J}^{\mathrm{B}}$) comes from the negative logarithm of the background probability density, while the second one ($\mathcal{J}^{\mathrm{obs}}$) comes from the negative logarithm of the observation error probability density. Some scaling factors of the probability densities are usually left out as they give a constant component of the cost function and do not affect the minimization. The third term ($-\ln \mathbb{P}(\mathbf{y})$) does not depend on $\mathbf{x}$ and can also be left out of the minimization function. Under the assumption that the background errors are normally distributed (5.1), and after leaving out constant terms, we have that

$$\mathcal{J}^{\mathrm{B}}(\mathbf{x}) = \frac{1}{2}\left(\mathbf{x}-\mathbf{x}^B\right)^T \mathbb{B}^{-1}\left(\mathbf{x}-\mathbf{x}^B\right) \,. \tag{5.11}$$

Similarly, under the assumption that observation errors are normally distributed and independent (5.8), and after leaving out the constant terms,

$$\begin{aligned} \mathcal{J}^{\mathrm{obs}}(\mathbf{x}) &= \sum_{i=1}^{N} \mathcal{J}_i^{\mathrm{obs}}(\mathbf{x}) = \frac{1}{2}\sum_{i=1}^{N}\left(\mathcal{H}\left(\mathbf{x}_i\right)-\mathbf{y}_i\right)^T \mathbb{R}_i \left(\mathcal{H}\left(\mathbf{x}_i\right)-\mathbf{y}_i\right) \\ \mathcal{J}_i^{\mathrm{obs}}(\mathbf{x}) &= \frac{1}{2}\left(\mathcal{H}\left(\mathbf{x}_i\right)-\mathbf{y}_i\right)^T \mathbb{R}_i \left(\mathcal{H}\left(\mathbf{x}_i\right)-\mathbf{y}_i\right) \end{aligned} \tag{5.12}$$

Because observation errors are independent each set of observations $\mathbf{y}_i$ at time t_i brings its own contribution $\mathcal{J}_i^{\mathrm{obs}}$ to the total cost function.

5.2.2 Four dimensional variational (4D-Var) data assimilation

In strongly-constrained 4D-Var data assimilation all observations (6.3) at all times $t_1, \cdots, t_N$ are simultaneously considered. The control parameters are the initial conditions $\mathbf{x}_0$; they

uniquely determine the state of the system at all future times via the model equation (6.6). The background state is the prior value of the initial conditions $\mathbf{x}_0^{\mathrm{B}}$.

Given the background value of the initial state $\mathbf{x}_0^{\mathrm{B}}$, the covariance of the initial background errors $\mathbb{B}_0$, the observations $\mathbf{y}_i$ and the corresponding observation error covariances $\mathbb{R}_i$, $i = 1, \cdots, N$, the 4D-Var problem looks for the maximum likelihood estimate $\mathbf{x}_0^A$ of the true initial conditions by solving the optimization problem (5.9). Combining (5.10), (5.11), and (5.12) leads to the 4D-var cost function:

$$\mathcal{J}(\mathbf{x}_0) = \frac{1}{2}\left(\mathbf{x}_0 - \mathbf{x}_0^B\right)^T \mathbb{B}_0^{-1}\left(\mathbf{x}_0 - \mathbf{x}_0^B\right) + \frac{1}{2}\sum_{i=1}^{N}\left(\mathcal{H}(\mathbf{x}_i) - \mathbf{y}_i\right)^T \mathbb{R}_i^{-1}\left(\mathcal{H}(\mathbf{x}_i) - \mathbf{y}_i\right) \quad (5.13)$$

Note that the departure of the initial conditions from the background is weighted by the inverse background covariance matrix, while the differences between the model predictions $\mathcal{H}(\mathbf{x}_i)$ and observations $\mathbf{y}_i$ are weighted by the inverse observation error covariances. The 4D-Var analysis is computed as the initial condition which minimizes (6.10) subject to the model equation constraints (6.6)

$$\mathbf{x}_0^A = \arg\min \mathcal{J}(\mathbf{x}_0) \quad \text{subject to} (6.6). \quad (5.14)$$

The model (6.6) propagates the optimal initial condition (6.10) forward in time to provide the analysis at future times, $\mathbf{x}_i^A = \mathcal{M}_{t_0 \to t_i}\mathbf{x}_0^A$.

The optimization problem (6.11) is solved numerically using a gradient-based technique. The gradient of (6.10) reads

$$\nabla\mathcal{J}(\mathbf{x}_0) = \mathbb{B}_0^{-1}\left(\mathbf{x}_0 - \mathbf{x}_0^B\right) + \sum_{i=1}^{N}\left(\frac{\partial \mathbf{x}_i}{\partial \mathbf{x}_0}\right)^T \left(\mathcal{H}'(\mathbf{x}_i)\right)^T \mathbb{R}_i^{-1}\left(\mathcal{H}(\mathbf{x}_i) - \mathbf{y}_i\right) \quad (5.15)$$

The 4D-Var gradient requires not only the linearized observation operator $\mathcal{H}'$, but also the transposed derivative of future states with respect to the initial conditions. The 4D-Var gradient can be obtained effectively by forcing the adjoint model with observation increments, and running it backwards in time. The construction of an adjoint model requires considerable effort.

5.3 Information Metrics and Gaussian Probabilities

The 4D-Var data assimilation of the observations $\mathbf{y}$ changes the distribution of errors (uncertainty) in the initial conditions from the background probability density $\mathcal{P}^{\mathrm{B}}(\mathbf{x})$ to the analysis probability density $\mathcal{P}^{\mathrm{A}}(\mathbf{x})$. If the data assimilation is beneficial the uncertainty associated with the new distribution $\mathcal{P}^{\mathrm{A}}$ is smaller than the uncertainty associated with the original distribution $\mathcal{P}^{\mathrm{B}}$.

Roughly speaking, the *information content* of the observations $\mathbf{y}$ is measured by the decrease in uncertainty from before data assimilation ($\mathcal{P}^{\mathrm{B}}$) to after data assimilation ($\mathcal{P}^{\mathrm{A}}$). The information content depends not only on the data ($\mathbf{y}_i$) but also on the data accuracy ($\mathbb{R}_i^{-1}$), on the background uncertainty ($\mathbb{B}_0^{-1}$), and on the model dynamics $\mathcal{M}$.

We are interested to rigorously quantify the information content of observations in 4D-Var. For this we use several information theoretic metrics, which are reviewed below.

5.3.1 Fisher information matrix

The Fisher information matrix (FIM) [Fisher, 1922] associated with the probability density function $\mathbb{P}(\mathbf{x})$ is defined as

$$\mathcal{F}(\mathbb{P}) = \int_{\mathbb{R}^n} \left[\frac{\partial\ (-\ln \mathbb{P}(\mathbf{x}))}{\partial \mathbf{x}}\right] \left[\frac{\partial\ (-\ln \mathbb{P}(\mathbf{x}))}{\partial \mathbf{x}})\right]^T \mathbb{P}(\mathbf{x})\, d\mathbf{x} \quad \in \mathbb{R}^{n\times n}\,. \tag{5.16}$$

The trace of the FIM offers a measure of the total level of uncertainty associated with the distribution.

Under the assumption that the background errors are normally distributed (5.1) the Fisher information matrix of the background error probability density $\mathcal{P}^{\mathrm{B}}(\mathbf{x}) = \mathcal{N}(\mathbf{x}_0^B, \mathbb{B}_0)$ is just the inverse of the background error covariance:

$$\begin{aligned}
\mathcal{F}\left(\mathcal{P}^{\mathrm{B}}\right) &= \int_{\mathbb{R}^n} \left[\nabla\ \mathcal{J}^{\mathrm{B}}(\mathbf{x}_0)\right] \left[\nabla\ \mathcal{J}^{\mathrm{B}}(\mathbf{x}_0)\right]^T \mathcal{P}^{\mathrm{B}}(\mathbf{x}_0)\, d\mathbf{x}_0 \\
&= \int_{\mathbb{R}^n} \mathbb{B}_0^{-1} \left(\mathbf{x}_0 - \mathbf{x}_0^B\right)^T \left(\mathbf{x}_0 - \mathbf{x}_0^B\right) \mathbb{B}_0^{-1}\, \mathcal{P}^{\mathrm{B}}(\mathbf{x}_0)\, d\mathbf{x}_0 \\
&= \mathbb{B}_0^{-1}\, \mathbb{B}_0\, \mathbb{B}_0^{-1} \\
&= \mathbb{B}_0^{-1}\,.
\end{aligned} \tag{5.17}$$

Here we have used the relation (5.10) to link the background error probability densities with the background part of the 4D-Var cost function.

Similarly, the FIM associated with the analysis probability density is

$$\mathcal{F}\left(\mathcal{P}^{\mathrm{A}}\right) = \int_{\mathbb{R}^n} \left[\nabla\ \mathcal{J}(\mathbf{x}_0)\right] \left[\nabla\ \mathcal{J}(\mathbf{x}_0)\right]^T \mathcal{P}^{\mathrm{A}}(\mathbf{x}_0)\, d\mathbf{x}_0\,. \tag{5.18}$$

Assuming that the analysis error probability density is Gaussian (6.5) the analysis Fisher information matrix is the inverse of the analysis error covariance.

$$\mathcal{F}\left(\mathcal{P}^{\mathrm{A}}\right) = \mathbb{A}_0^{-1}\,.$$

The information content of the observations used in data assimilation can be measured as the trace of the background FIM (total uncertainty in the background) minus the trace of the analysis FIM (total uncertainty in the analysis) [Rodgers, 1998, 2000]:

$$\mathcal{I}^{\text{FIM}} = \text{trace}\left(\mathcal{F}\left(\mathcal{P}^{\text{A}}\right)\right) - \text{trace}\left(\mathcal{F}\left(\mathcal{P}^{\text{B}}\right)\right) \tag{5.19}$$

In the Gaussian case this reduces to the trace of difference between the analysis and background error covariance matrices.

$$\mathcal{I}^{\text{FIM}} = \text{trace}\left(\mathbb{A}_0^{-1} - \mathbb{B}_0^{-1}\right) \tag{5.20}$$

5.3.2 Shannon information

The entropy associated with a probability density is defined as [Shannon and Weaver, 1949; Bartlett, 1962]

$$\mathcal{H}\left(\mathbb{P}\right) = \int_{\mathbb{R}^n} \mathbb{P}(\mathbf{x}) \, \ln\left(\mathbb{P}(\mathbf{x})\right) \, d\mathbf{x}$$

and offers a measure of the *average uncertainty* with which one knows the state $\mathbf{x}$, if the estimation error has a probability density $\mathbb{P}$.

For example, assume that the background error distribution is Gaussian (5.1). The entropy of the background probability density is given by the relation [Rodgers, 2000]

$$\mathcal{P}^{\text{B}}(\mathbf{x}) = \mathcal{N}\left(\mathbf{x}_0^B, \mathbb{B}_0\right) \quad \Rightarrow \quad \mathcal{H}\left(\mathcal{P}^{\text{B}}\right) = n \, \ln\left(\sqrt{2\pi e}\right) + \frac{1}{2} \ln\det\left(\mathbb{B}_0\right) \, .$$

In this case, the entropy may be interpreted as a measure of the volume in phase space enclosed by a surface of constant probability.

Using the Bayes rule (5.7) the entropy of the analysis error probability distribution can be written as

$$\begin{aligned} \mathcal{H}\left(\mathcal{P}^{\text{A}}\right) &= \int \ln \frac{\mathcal{P}^{\text{B}}(\mathbf{x}) \cdot \mathbb{P}(\mathbf{y}|\mathbf{x})}{\mathbb{P}(\mathbf{y})} \, \mathcal{P}^{\text{A}}(\mathbf{x}) \, d\mathbf{x} \\ &= \int \left[\ln \mathcal{P}^{\text{B}}(\mathbf{x}) + \ln \mathbb{P}(\mathbf{y}|\mathbf{x}) - \ln \mathbb{P}(\mathbf{y})\right] \, \mathcal{P}^{\text{A}}(\mathbf{x}) \, d\mathbf{x} \end{aligned}$$

The Shannon information content of observations $\mathbf{y}$ used in 4D-Var data assimilation is defined as the decrease in the average uncertainty with which the initial state is known. Specifically, the Shannon information content is given by the difference between the background entropy and the analysis entropy,

$$\mathcal{I}^{\text{Shannon}} = \mathcal{H}\left(\mathcal{P}^{\text{B}}\right) - \mathcal{H}\left(\mathcal{P}^{\text{A}}\right) \, . \tag{5.21}$$

For example, under the assumption that both the background (5.1) and the analysis error probability densities are Gaussian (6.5), the Shannon information content of the observations used in data assimilation is

$$\begin{aligned} \mathcal{I}^{\text{Shannon}} &= \frac{1}{2} \ln \det (\mathbb{B}_0) - \frac{1}{2} \ln \det (\mathbb{A}_0) \\ &= \frac{1}{2} \ln \det \left(\mathbb{B}_0 \, \mathbb{A}_0^{-1} \right) \\ &= \frac{1}{2} \ln \det \left(\mathbb{A}_0^{-1/2} \, \mathbb{B}_0 \, \mathbb{A}_0^{-1/2} \right) \; . \end{aligned} \tag{5.22}$$

5.3.3 Degrees of freedom for signal

The Degrees of freedom for signal (DFS) metric for the information content has been previously employed in meteorological data assimilation [Rodgers, 1996; Fisher, 2003; Cardinali et al., 2004; Stewart et al., 2008; Zupanski, 2009].

Consider the symmetric matrix square root $\mathbb{B}_0^{1/2}$ of the background covariance; we have that

$$\mathbb{B}_0 = \mathbb{B}_0^{1/2} \, \mathbb{B}_0^{1/2} \; , \quad \mathbb{B}_0^{-1} = \mathbb{B}_0^{-1/2} \, \mathbb{B}_0^{-1/2} \; .$$

Consider also the orthogonal matrix $\mathbb{Q}$ whose columns are the eigenvectors of the symmetric matrix $\mathbb{B}_0^{-1/2} \, \mathbb{A}_0 \, \mathbb{B}_0^{-1/2}$

$$\mathbb{Q}^T \left(\mathbb{B}_0^{-1/2} \, \mathbb{A}_0 \, \mathbb{B}_0^{-1/2} \right) \mathbb{Q} = \Sigma \; .$$

with Σ a diagonal matrix.

The matrix $\mathbb{L} = \mathbb{B}_0^{-1/2} \, \mathbb{Q}$ has the property that it transforms simultaneously the background and the analysis covariances to diagonal forms [Fisher, 2003] when it is symmetrically applied:

$$\mathbb{L}^T \, \mathbb{B}_0 \, \mathbb{L} = \mathbb{I} \; , \quad \mathbb{L}^T \, \mathbb{A}_0 \, \mathbb{L} = \Sigma \; .$$

The diagonal elements of the transformed background error covariance matrix are equal to unity and each corresponds to an individual degree of freedom. The eigenvalues of the transformed matrix Σ, on the other hand, can be interpreted as the relative reduction in variance in each of the n statistically independent directions corresponding to the n components of error in the state vector. The degrees of freedom for signal measures the total reduction in variance and is defined as

$$\mathcal{I}^{\text{DFS}} = \text{trace} \, (\mathbb{I} - \Sigma) = n - \text{trace} \left(\mathbb{B}_0^{-1/2} \, \mathbb{A}_0 \, \mathbb{B}_0^{-1/2} \right) = n - \text{trace} \left(\mathbb{B}_0^{-1} \, \mathbb{A}_0 \right) \; . \tag{5.23}$$

The relative reduction in variance $\mathbb{B}_0^{-1} \, \mathbb{A}_0$ could also be interpreted as the gradient of the analysis in observation space with respect to the observations.

5.3.4 Relative entropy

The information content of the observations used in data assimilation can also be measured by the relative entropy (RE) of the analysis probability density with respect to the background probability density:

$$\mathcal{I}^{\mathrm{RE}} \;=\; \int_{\mathbb{R}^n} \mathcal{P}^{\mathrm{A}}(\mathbf{x})\, \ln \frac{\mathcal{P}^{\mathrm{A}}(\mathbf{x})}{\mathcal{P}^{\mathrm{B}}(\mathbf{x})}\, d\mathbf{x}\,.$$

Under the assumption that both the background (5.1) and the analysis error probability densities are Gaussian (6.5), the relative entropy of the analysis over the background is [**?**]:

$$\begin{aligned}
\mathcal{I}^{\mathrm{RE}} \;=\; & \frac{1}{2}\left(\mathbf{x}_0^A - \mathbf{x}_0^B\right)^T \mathbb{B}_0^{-1}\left(\mathbf{x}_0^A - \mathbf{x}_0^B\right) && \text{(5.24a)}\\
& + \frac{1}{2}\,\mathrm{trace}\left(\mathbb{B}_0^{-1/2}\,\mathbb{A}_0\,\mathbb{B}_0^{-1/2}\right) && \text{(5.24b)}\\
& - \frac{n}{2} && \text{(5.24c)}\\
& + \frac{1}{2}\ln\,\det\left(\mathbb{B}_0^{1/2}\,\mathbb{A}_0^{-1}\,\mathbb{B}_0^{1/2}\right)\,. && \text{(5.24d)}
\end{aligned}$$

The *signal part* of the relative entropy

$$\mathcal{I}^{\mathrm{Signal}} = \frac{1}{2}\left(\mathbf{x}_0^A - \mathbf{x}_0^B\right)^T \mathbb{B}_0^{-1}\left(\mathbf{x}_0^A - \mathbf{x}_0^B\right) \qquad (5.25)$$

measures the reduction of uncertainty due to the change in the best estimate from the background state to the analysis state. The terms (5.24b), (5.24c), and (5.24d) together form the *dispersion part* of the relative entropy.

Comparing (5.24a)–(5.24b)–(5.24c)–(5.24d) and (5.22), (5.23), (5.25) reveals that

$$\mathcal{I}^{\mathrm{RE}} = \underbrace{\mathcal{I}^{\mathrm{Signal}}}_{(5.24a)} + \underbrace{\mathcal{I}^{\mathrm{Shannon}}}_{(5.24d)} - \underbrace{\frac{1}{2}\,\mathcal{I}^{\mathrm{DFS}}}_{(5.24b)-(5.24c)}\,.$$

Let us have a closer look at the relative entropy between two Gaussian distributions:

$$\begin{aligned}
\mathcal{I}^{\mathrm{RE}} \;=\; & \int_{\mathbb{R}^n} \mathcal{P}^A(\mathbf{x}) \cdot \left(\ln \mathcal{P}^A(\mathbf{x}) - \ln \mathcal{P}^B(\mathbf{x})\right) d\mathbf{x}\\
=\; & \int_{\mathbb{R}^n} \mathcal{P}^A(\mathbf{x}) \cdot \left(-\frac{1}{2}\ln\det\,\mathbb{A}_0 - \frac{1}{2}\left(\mathbf{x}-\mathbf{x}_0^A\right)^T \mathbb{A}_0^{-1}\left(\mathbf{x}-\mathbf{x}_0^A\right)\right.\\
& \qquad \left. + \frac{1}{2}\ln\det\,\mathbb{B}_0 + \frac{1}{2}\left(\mathbf{x}-\mathbf{x}_0^B\right)^T \mathbb{B}_0^{-1}\left(\mathbf{x}-\mathbf{x}_0^B\right)\right) d\mathbf{x}\\
=\; & \frac{1}{2}\ln\det\,\mathbb{A}_0^{-1}\,\mathbb{B}_0 - \frac{n}{2} + \frac{1}{2}\int_{\mathbb{R}^n} \mathcal{P}^A(\mathbf{x}) \cdot \left(\mathbf{x}-\mathbf{x}_0^B\right)^T \mathbb{B}_0^{-1}\left(\mathbf{x}-\mathbf{x}_0^B\right) dx
\end{aligned}$$

We see that the Shannon part (5.24d) of the relative entropy comes from the scaling factors of the Gaussian distributions (the difference between the logarithms of the factors $((2\pi)^{n/2}\sqrt{\det \mathbb{B}_0})^{-1}$ and $((2\pi)^{n/2}\sqrt{\det \mathbb{A}_0})^{-1}$). Since 4D-Var cost functions do not account for this scaling we cannot hope to accurately recover the Shannon part of the dispersion just by analyzing the cost function.

The constant term (5.24c) comes from the integration (averaging) of the exponent of the analysis distribution; this is shown in Appendix A.2 in relation (A.4). The signal part (5.24a) and the DFS part (5.24b) come from the integration (averaging) of the exponent of the background distribution; this is shown in Appendix A.2 in relation (A.5).

The three terms (5.24c), (5.24a), and t(5.24b) are represented in the 4D-Var cost function, and we should be able to estimate them by studying different statistics of different parts of the 4D-Var cost function.

5.4 Estimation of the Data Information Content in the Context of 4D-Var Data Assimilation

We seek to derive a computationally-easy way to estimate the information content of various observations in the context of 4D-Var. The proposed approach is based on an approximate sampling from the posterior error distribution in 4D-Var. Thus, our approach is a hybrid one: ensembles are used to infer the information content of observations used in variational data assimilation.

Sampling from the posterior probability density at t_0 is challenging since this probability density is not explicitly computed by 4D-Var. Approximate sampling can be performed using second order adjoints, and computing a few eigenvectors corresponding to the dominant eigenvalues of the inverse Hessian.

One approach uses the fact that the analysis covariance matrix is approximated by the inverse Hessian of the cost function, evaluated at the optimum [Thacker, 1989; Gejadze et al., 2008]. A few eigenvectors corresponding to the dominant eigenvalues of the inverse Hessian are computed; they approximate the principal components of the posterior error and can be used for approximate sampling from the posterior distribution. The computation of the dominant eigenpairs of the inverse Hessian can be done using only Hessian vector products, for example obtained via a second order adjoint. Alternatively, if a quasi-Newton method is used in optimization (e.g., L-BFGS) the low rank quasi-Newton approximation of the inverse Hessian is constructed by the method and available for use in approximate sampling.

An alternative approach is based on a subspace analysis of 4D-Var and is explained in Appendix A.1. The latter is the particular approximate sampling used in this chapter.

Therefore, we assume that we have the ability to obtain the following sample of initial conditions from the posterior distribution:

$$\mathbf{x}_0^r \in \mathcal{P}^{\mathrm{A}}(\mathbf{x}_0) \ , \quad r = 1, \cdots, \mathrm{N}_{\mathrm{ens}} \ . \tag{5.26}$$

Based on it we can approximate expected values with respect to the posterior density by posterior ensemble averages as follows:

$$\mathbb{E}^{\mathrm{A}}\left[f\left(\mathbf{x}_0\right)\right] \approx \left\langle f\left(\mathbf{x}_0\right)\right\rangle^A = \frac{1}{\mathrm{N}_{\mathrm{ens}}} \sum_{r=1}^{\mathrm{N}_{\mathrm{ens}}} f\left(\mathbf{x}_0^r\right) \ . \tag{5.27}$$

5.4.1 Estimation of the FIM information content

In the 4D-Var setting a gradient based optimization method is typically employed to minimize the cost function $\mathcal{J}(\mathbf{x})$. The gradients are evaluated by the adjoint model; specifically, the value of the adjoint variable at the initial time equals the gradient of the cost function with respect to the initial state

$$\lambda_0(\mathbf{x}_0) = \nabla_{\mathbf{x}_0} \mathcal{J}(\mathbf{x}_0) \ .$$

The adjoint variable depends on the forward model trajectory about which the linearization is performed. This is indicated explicitly by making the adjoint variable a function of the forward initial condition.

The trace of the analysis FIM (5.18) can be expressed as:

$$\begin{aligned}
\mathrm{trace}\left(\mathcal{F}\left(\mathcal{P}^{\mathrm{A}}\right)\right) &= \int_{\mathbb{R}^n} \mathrm{trace}\left(\lambda_0(\mathbf{x}_0)\,\lambda_0^T(\mathbf{x}_0)\right)\,\mathcal{P}^{\mathrm{A}}(\mathbf{x}_0)\,d\mathbf{x}_0 \\
&= \int_{\mathbb{R}^n} \left(\sum_{\ell=1}^{n} \left(\lambda_0(\mathbf{x}_0)\right)_\ell^2\right)\,\mathcal{P}^{\mathrm{A}}(\mathbf{x}_0)\,d\mathbf{x}_0 \\
&= \int_{\mathbb{R}^n} \left\|\lambda_0(\mathbf{x}_0)\right\|^2\,\mathcal{P}^{\mathrm{A}}(\mathbf{x}_0)\,d\mathbf{x}_0 \\
&= \mathbb{E}^{\mathrm{A}}\left[\left\|\lambda_0(\mathbf{x}_0)\right\|^2\right] \\
&\approx \left\langle \left\|\lambda_0(\mathbf{x}_0)\right\|^2 \right\rangle^A \ .
\end{aligned}$$

The trace of the analysis FIM is the average value of the adjoint variable norm with respect to the analysis distribution. Using the sample of initial conditions (5.26) the statistical average can be approximated by the ensemble average.

Under the typical assumption that the background probability is Gaussian and using (5.17) and (5.19) we obtain the following estimate for the FIM information content of all observations:

$$\mathcal{I}^{\mathrm{FIM}} \approx \left\langle \left\|\lambda_0(\mathbf{x}_0)\right\|^2 \right\rangle^A - \mathrm{trace}\left(\mathbb{B}_0^{-1}\right) \ . \tag{5.28}$$

Computational procedure for estimating the FIM information

After the data assimilation has been performed, one runs the forward *and the adjoint* models N_{ens} times starting with forward initial conditions sampled from the analysis probability density (5.26). Each run produces an adjoint gradient, whose norm is computed. The ensemble average of these gradient norms estimates the trace of the analysis FIM.

5.4.2 Estimation of the DFS information content

In this section we consider the idealized situation detailed in Appendix A.3. Specifically, we assume that the model is linear (A.6), the observation operator is also linear (A.7), and both the background errors and the observation errors are normally distributed. The analysis relies on the properties of random quadratic functionals presented in Appendix A.2.

Consider running the model with random initial conditions taken from the distribution $\widehat{\mathbf{x}}_0 \in \mathcal{N}(\mu, \mathbb{C})$. Each run results in different values of the 4D-Var cost function; we are interested to understand the information provided by the statistics of the (ensemble of) cost function values.

Note that $\widehat{\mathbf{x}}_0 - \mathbf{x}_0^B \in \mathcal{N}\left(\mu - \mathbf{x}_0^B, \mathbb{C}\right)$. A direct application of (A.3a) reveals that the background component of the cost function has the following mean:

$$\begin{aligned} \mathcal{J}^{\mathrm{B}}(\widehat{\mathbf{x}}_0) &= \frac{1}{2}\left(\widehat{\mathbf{x}}_0 - \mathbf{x}_0^B\right)^T \mathbb{B}_0^{-1}\left(\widehat{\mathbf{x}}_0 - \mathbf{x}_0^B\right) \\ \mathbb{E}\left[\mathcal{J}^{\mathrm{B}}(\widehat{\mathbf{x}}_0)\right] &= \frac{1}{2}\left(\mu - \mathbf{x}_0^B\right)^T \mathbb{B}_0^{-1}\left(\mu - \mathbf{x}_0^B\right) + \frac{1}{2}\,\mathrm{trace}\left(\mathbb{B}_0^{-1}\,\mathbb{C}\right) \\ &= \mathcal{J}^B(\mu) + \frac{1}{2}\,\mathrm{trace}\left(\mathbb{C}^{1/2}\,\mathbb{B}_0^{-1}\,\mathbb{C}^{1/2}\right) . \end{aligned}$$

Since the dynamics is linear, for a given observation data vector $\mathbf{y}_i$ we have that

$$H_i\,M_i\,\widehat{\mathbf{x}}_0 - \mathbf{y}_i \;\in\; \mathcal{N}\left(H_i\,M_i\,\mu - \mathbf{y}_i\;,\; H_i\,M_i\,\mathbb{C}\,M_i^T\,H_i^T\right) .$$

Note that the above relation characterizes only the uncertainty in the initial conditions. The data is given; the same data values $\mathbf{y}_i$ are used for each initial condition $\widehat{\mathbf{x}}_0$.

The observation component of the cost function:

$$\mathcal{J}^{\mathrm{obs}}(\widehat{\mathbf{x}}_0) = \frac{1}{2}\sum_{i=0}^{N}\left(H_i\,M_i\,\widehat{\mathbf{x}}_0 - \mathbf{y}_i\right)^T \mathbb{R}_i^{-1}\left(H_i\,M_i\,\widehat{\mathbf{x}}_0 - \mathbf{y}_i\right)$$

has the following mean:

$$\begin{aligned}
\mathbb{E}\left[\mathcal{J}^{\text{obs}}\left(\widehat{\mathbf{x}}_0\right)\right] &= \frac{1}{2}\sum_{i=0}^{N}\mathbb{E}\left[\left(H_i\,M_i\,\widehat{\mathbf{x}}_0-\mathbf{y}_i\right)^T\,\mathbb{R}_i^{-1}\,\left(H_i\,M_i\,\widehat{\mathbf{x}}_0-\mathbf{y}_i\right)\right] \\
&= \frac{1}{2}\sum_{i=0}^{N}\left(H_i\,M_i\,\mu-\mathbf{y}_i\right)^T\,\mathbb{R}_i^{-1}\,\left(H_i\,M_i\,\mu-\mathbf{y}_i\right) \\
&\quad +\frac{1}{2}\sum_{i=0}^{N}\text{trace}\left(\mathbb{R}_i^{-1}\,H_i\,M_i\,\mathbb{C}\,M_i^T\,H_i^T\right) \\
&= \mathcal{J}^{\text{obs}}\left(\mu\right)+\frac{1}{2}\sum_{i=0}^{N}\text{trace}\left(\mathbb{C}^{1/2}\,M_i^T\,H_i^T\,\mathbb{R}_i^{-1}\,H_i\,M_i\,\mathbb{C}^{1/2}\right) \\
&= \mathcal{J}^{\text{obs}}\left(\mu\right)+\frac{1}{2}\text{trace}\left(\mathbb{C}^{1/2}\,\left(\sum_{i=0}^{N}M_i^T\,H_i^T\,\mathbb{R}_i^{-1}\,H_i\,M_i\right)\,\mathbb{C}^{1/2}\right) \\
&= \mathcal{J}^{\text{obs}}\left(\mu\right)+\frac{1}{2}\text{trace}\left(\mathbb{C}^{1/2}\,\left(\mathbb{A}_0^{-1}-\mathbb{B}_0^{-1}\right)\,\mathbb{C}^{1/2}\right)
\end{aligned}$$

Putting the two formulas together results in

$$\mathbb{E}\left[\mathcal{J}\left(\widehat{\mathbf{x}}_0\right)\right]-\mathcal{J}(\mu)=\frac{1}{2}\,\text{trace}\left(\mathbb{C}^{1/2}\,\mathbb{A}_0^{-1}\,\mathbb{C}^{1/2}\right)\,. \tag{5.29}$$

Sampling from a diagonal distribution

Recall that in the Gaussian case the Fisher information matrix (FIM) is just the inverse of the covariance. Let $\mathbb{C}=\sigma^2\,\mathbb{I}$. Then the total reduction in uncertainty is given by the trace of the difference between the analysis and the background FIMs:

$$\mathbb{E}\left[\mathcal{J}^{\text{obs}}\left(\widehat{\mathbf{x}}_0\right)\right]-\mathcal{J}^{\text{obs}}\left(\mu\right) \;=\; \frac{\sigma^2}{2}\,\text{trace}\left(\mathbb{A}_0^{-1}-\mathbb{B}_0^{-1}\right)\,.$$

Consequently the FIM information content of all observations $\mathbf{y}_1\cdots\mathbf{y}_N$ is

$$\mathcal{I}^{\text{FIM}}_{\mathbf{y}_1\cdots\mathbf{y}_N} \;=\; \frac{2}{\sigma^2}\,\left(\mathbb{E}\left[\mathcal{J}^{\text{obs}}\left(\widehat{\mathbf{x}}_0\right)\right]-\mathcal{J}^{\text{obs}}\left(\mu\right)\right)\,.$$

The contribution of the observations $\mathbf{y}_i$ taken at time t_i to the decrease of the trace of FIM, i.e., the FIM information content of $\mathbf{y}_i$ is:

$$\begin{aligned}
\mathcal{I}^{\text{FIM}}_{\mathbf{y}_i} &= \frac{2}{\sigma^2}\,\left(\mathbb{E}\left[\mathcal{J}_i^{\text{obs}}\left(\widehat{\mathbf{x}}_0\right)\right]-\mathcal{J}_i^{\text{obs}}\left(\mu\right)\right) \\
&= \frac{1}{\sigma^2}\,\mathbb{E}\left[\left(H_i\,M_i\widehat{\mathbf{x}}_0-\mathbf{y}_i\right)^T\,\mathbb{R}_i^{-1}\,\left(H_i\,M_i\,\widehat{\mathbf{x}}_0-\mathbf{y}_i\right)\right] \\
&\quad -\frac{1}{\sigma^2}\left(H_i\,M_i\,\mu-\mathbf{y}_i\right)^T\,\mathbb{R}_i^{-1}\,\left(H_i\,M_i\,\mu-\mathbf{y}_i\right)\,.
\end{aligned}$$

While in the linear case this expression does not depend on μ, in the nonlinear case we can take $\mu = \mathbf{x}_0^A$ (after the analysis to assess the impact the observation *had* on the FIM) and $\mu = \mathbf{x}_0^B$ (before the analysis to assess the impact the observation *will have* on the FIM).

Sampling from the analysis distribution

A sample $\widehat{\mathbf{x}}_0 \in \mathcal{N}\left(\mathbf{x}_0^A, \mathbb{A}_0\right)$ from the posterior distribution leads to

$$\begin{aligned}
\mathbb{E}^{\mathrm{A}}\left[\mathcal{J}^B\left(\widehat{\mathbf{x}}_0\right)\right] &= \mathcal{J}^B(\mathbf{x}_0^A) + \frac{1}{2}\,\mathrm{trace}\left(\mathbb{A}_0^{1/2}\,\mathbb{B}_0^{-1}\,\mathbb{A}_0^{1/2}\right) \\
\mathbb{E}^{\mathrm{A}}\left[\mathcal{J}^{\mathrm{obs}}\left(\widehat{\mathbf{x}}_0\right)\right] &= \mathcal{J}^{\mathrm{obs}}\left(\mathbf{x}_0^A\right) + \frac{1}{2}\mathrm{trace}\left(\mathbb{A}_0^{1/2}\,\left(\mathbb{A}_0^{-1} - \mathbb{B}_0^{-1}\right)\,\mathbb{A}_0^{1/2}\right) \\
&= \mathcal{J}^{\mathrm{obs}}\left(\mathbf{x}_0^A\right) + \frac{n}{2} - \frac{1}{2}\mathrm{trace}\left(\mathbb{A}_0^{1/2}\,\mathbb{B}_0^{-1}\,\mathbb{A}_0^{1/2}\right) \\
\mathbb{E}^{\mathrm{A}}\left[\mathcal{J}\left(\widehat{\mathbf{x}}_0\right)\right] &= \mathcal{J}(\mathbf{x}_0^A) + \frac{n}{2}\,.
\end{aligned}$$

The signal part of the relative entropy (5.24a) is given by $\mathcal{J}^B(x_0^A)$. Attributing the contribution of each observation to the signal part of the entropy is more involved.

We have the following estimate of the DFS information content (5.24b) of all observations $\mathbf{y}_1 \cdots \mathbf{y}_N$:

$$\mathcal{I}^{\mathrm{DFS}}_{\mathbf{y}_1 \cdots \mathbf{y}_N} = n - \mathrm{trace}\left(\mathbb{A}_0^{1/2}\,\mathbb{B}^{-1}\,\mathbb{A}_0^{1/2}\right) = 2\,\mathbb{E}^{\mathrm{A}}\left[\mathcal{J}^{\mathrm{obs}}\left(\widehat{\mathbf{x}}_0\right)\right] - 2\,\mathcal{J}^{\mathrm{obs}}\left(\mathbf{x}_0^A\right)\,. \tag{5.30}$$

This method allows to account for the contribution of each observation $\mathbf{y}_i$ to the DFS information as follows:

$$\begin{aligned}
\mathcal{I}^{\mathrm{DFS}}_{\mathbf{y}_i} &= 2\,\mathbb{E}^{\mathrm{A}}\left[\mathcal{J}_i^{\mathrm{obs}}\left(\widehat{\mathbf{x}}_0\right)\right] - 2\,\mathcal{J}_i^{\mathrm{obs}}\left(\mathbf{x}_0^A\right) \\
&= \mathbb{E}^{\mathrm{A}}\left[\left(\mathcal{H}_i\left(\widehat{\mathbf{x}}_i\right) - \mathbf{y}_i\right)^T\,\mathbb{R}_i^{-1}\,\left(\mathcal{H}_i\left(\widehat{\mathbf{x}}_i\right) - \mathbf{y}_i\right)\right] \\
&\quad - \left(\mathcal{H}_i\left(\mathbf{x}_0^A\right) - \mathbf{y}_i\right)^T\,\mathbb{R}_i^{-1}\,\left(\mathcal{H}_i\left(\mathbf{x}_0^A\right) - \mathbf{y}_i\right)
\end{aligned}$$

For nonlinear models this relation holds within some approximation margin.

In practice the posterior expected value is replaced by the ensemble expected value

$$\mathcal{I}^{\mathrm{DFS}}_{\mathbf{y}_i} \;\approx\; 2\left\langle \mathcal{J}_i^{\mathrm{obs}}\left(\widehat{\mathbf{x}}_0\right)\right\rangle^{\mathrm{A}} - 2\,\mathcal{J}_i^{\mathrm{obs}}\left(\mathbf{x}_0^A\right)\,. \tag{5.31}$$

Sampling from the background distribution

A sample $\widehat{\mathbf{x}}_0 \in \mathcal{N}\left(\mathbf{x}_0^B, \mathbb{B}_0\right)$ from the background distribution leads to

$$\begin{aligned}
\mathbb{E}^{\mathrm{B}}\left[\mathcal{J}^B\left(\widehat{\mathbf{x}}_0\right)\right] &= \mathcal{J}^B(\mathbf{x}_0^B) + \frac{n}{2} \\
\mathbb{E}^{\mathrm{B}}\left[\mathcal{J}^{\mathrm{obs}}\left(\widehat{\mathbf{x}}_0\right)\right] &= \mathcal{J}^{\mathrm{obs}}\left(\mathbf{x}_0^B\right) + \frac{1}{2}\,\mathrm{trace}\left(\mathbb{B}_0^{1/2}\,\left(\mathbb{A}_0^{-1} - \mathbb{B}_0^{-1}\right)\,\mathbb{B}_0^{1/2}\right) \\
\mathbb{E}^{\mathrm{B}}\left[\mathcal{J}\left(\widehat{\mathbf{x}}_0\right)\right] &= \mathcal{J}\left(\mathbf{x}_0^B\right) + \frac{1}{2}\,\mathrm{trace}\left(\mathbb{B}_0^{1/2}\,\mathbb{A}_0^{-1}\,\mathbb{B}_0^{1/2}\right)\,.
\end{aligned}$$

Computational procedure for estimating the DFS information

After the data assimilation has been performed, one runs the forward model $\mathrm{N_{ens}}$ times. The initial conditions are sampled from the analysis distribution (5.26) (or from another distribution, e.g., diagonal, to obtain different statistics). An additional run is performed starting from the analysis initial conditions. During each run one records all individual contributions $\mathcal{J}_i^{\mathrm{obs}}$ of all observations $\mathbf{y}_i$ to the cost function. This data is post-processed according to (5.31). The ensemble average of the contributions $\mathcal{J}_i^{\mathrm{obs}}$, minus the contribution obtained from the analysis run, estimates (half of) the DFS information content of the data $\mathbf{y}_i$.

5.4.3 Estimation of the RE information content

The relative entropy (RE) information content of all observations $\mathbf{y}_1 \cdots \mathbf{y}_N$ is measured by the relative entropy of the posterior probability density $\mathcal{P}^{\mathrm{A}}$ over the background probability density $\mathcal{P}^{\mathrm{B}}$

$$\begin{aligned}
\mathcal{I}^{\mathrm{RE}}_{\mathbf{y}_1\cdots\mathbf{y}_N} &= \int_{\mathbb{R}^n} \mathcal{P}^{\mathrm{A}}\left(\mathbf{x}\right) \cdot \ln \frac{\mathcal{P}^{\mathrm{A}}\left(\mathbf{x}\right)}{\mathcal{P}^{\mathrm{B}}\left(\mathbf{x}\right)}\, d\mathbf{x} \\
&= \int_{\mathbb{R}^n} \mathcal{P}^{\mathrm{A}}\left(\mathbf{x}\right) \cdot \ln \frac{\mathbb{P}\left(\mathbf{y}|\mathbf{x}\right)}{\mathbb{P}\left(\mathbf{y}\right)}\, d\mathbf{x} \\
&= \int_{\mathbb{R}^n} \mathcal{P}^{\mathrm{A}}\left(\mathbf{x}\right) \cdot \Big(\ln \mathbb{P}\left(\mathbf{y}|\mathbf{x}\right) - \ln \mathbb{P}(\mathbf{y})\Big)\, d\mathbf{x} \\
&= \mathbb{E}^{\mathrm{A}}\left[\ln \mathcal{P}\left(\mathbf{y}|\mathbf{x}\right)\right] - \ln \mathcal{P}(\mathbf{y}) \\
&= \mathrm{const} - \mathbb{E}^{\mathrm{A}}\left[\mathcal{J}^{\mathrm{obs}}(\mathbf{x})\right]
\end{aligned}$$

where we have made use of Bayes rule (5.7) to derive the second relation, and of (5.10) to derive the last equation. The marginal distribution of observations $\mathbf{y}$ does not depend on $\mathbf{x}$ and its expected value is a constant.

Assuming we can sample the posterior distribution this expected value can be approximated by the ensemble mean. The RE information content of all observations is estimated as

$$\mathcal{I}^{\mathrm{RE}}_{\mathbf{y}_1 \cdots \mathbf{y}_N} \approx \text{const} - \left\langle \mathcal{J}^{\mathrm{obs}}(\mathbf{x}) \right\rangle^{\mathrm{A}} . \tag{5.32}$$

The relative entropy information content is larger when the 4D-Var process decreases more the observation part of the cost function. In other words, the lower the mismatch between model predictions and observations after assimilation the higher the relative entropy information content of observations is.

The RE information content of the particular observation $\mathbf{y}_i$ can be quantified as follows. Data assimilation using all observations $\mathbf{y}_1 \cdots \mathbf{y}_N$ results in a posterior probability density $\mathcal{P}^{\mathrm{A}}(\mathbf{x})$. Data assimilation using all observations except $\mathbf{y}_i$ results in another posterior probability density $\mathcal{P}^{\mathrm{A}}_{-i}(\mathbf{x})$. The RE information contribution of data $\mathbf{y}_i$ is measured by the relative entropy of the full-data posterior probability density $\mathcal{P}^{\mathrm{A}}$ over the partial-data posterior density $\mathcal{P}^{\mathrm{A}}_{-i}$. If the observation errors at different times are independent it can be shown that

$$\begin{aligned} \mathcal{I}^{\mathrm{RE}}_{\mathbf{y}_i} &= \int_{\mathbb{R}^n} \mathcal{P}^{\mathrm{A}}(\mathbf{x}) \cdot \ln \frac{\mathcal{P}^{\mathrm{A}}(\mathbf{x})}{\mathcal{P}^{\mathrm{A}}_{-i}(\mathbf{x})} \, d\mathbf{x} \\ &= \text{const}_i - \mathbb{E}^{\mathrm{A}} \left[\mathcal{J}^{\mathrm{obs}}_i(\mathbf{x}) \right] \approx \text{const}_i - \left\langle \mathcal{J}^{\mathrm{obs}}_i(\mathbf{x}) \right\rangle^{\mathrm{A}} . \end{aligned} \tag{5.33}$$

The constant comes from the marginal probability of the observation $\mathbf{y}_i$ and is different for each data point. Therefore it is difficult to apportion the information gain to individual observations using this metric.

An alternative, more computationally involved approach would be to repeat the data assimilation without the data point $\mathbf{y}_i$, and to build another ensemble drawn from $\mathcal{P}^{\mathrm{A}}_{-i}(\mathbf{x})$. For each data assimilation experiment one computes the total RE information content (5.32). The information gain due to the data $\mathbf{y}_i$ is the measured by

$$\mathcal{I}^{\mathrm{RE}}_{\mathbf{y}_i} = \mathcal{I}^{\mathrm{RE}}_{\mathbf{y}_1 \cdots \mathbf{y}_N} - \mathcal{I}^{\mathrm{RE}}_{\mathbf{y}_1 \cdots y_{i-1}, y_{i+1} \cdots \mathbf{y}_N} . \tag{5.34}$$

Computational procedure for estimating the RE information

The computational procedure is similar to the one for the DFS information presented in Section 5.4.2. An ensemble of models is run with the initial conditions sampled from the analysis distribution (5.26). The ensemble average of the observation part $\mathcal{J}^{\mathrm{obs}}$ of the cost function estimates the RE information content of all observations (5.32), modulo a constant. This procedure can be repeated for different data assimilation scenarios, where individual data points are being withheld; the difference between the resulting metrics estimates the RE information content of the withheld data.

5.4.4 Estimation of the Shannon information content

We have seen that the Shannon information is related to the scaling of the Gaussian probability densities. This information is ignored by the 4D-Var cost function. Therefore, we cannot expect to obtain accurate estimates of the Shannon information content by mining the cost function information.

A (very) rough approximation can be obtained using the eigenvalues of the ensemble covariance matrices, as follows. Consider a set of perturbations drawn from the background ensemble, and a set of perturbations drawn from the analysis ensemble; in matrix notation

$$\Delta \mathbf{x}_0^{\mathrm{B}} \in \mathbb{R}^{n \times \mathrm{N_{ens}}} \; ; \quad \Delta \mathbf{x}_0^{\mathrm{A}} \in \mathbb{R}^{n \times \mathrm{N_{ens}}} \; ; \quad \mathrm{N_{ens}} \ll n \; .$$

The error covariance matrices are approximated by the ensemble covariance

$$\mathbb{B}_0 \approx \frac{1}{(\mathrm{N_{ens}}-1)} \cdot \left(\Delta \mathbf{x}_0^B\right)^T \cdot \Delta \mathbf{x}_0^B \; ; \quad \mathbb{A}_0 \approx \frac{1}{(\mathrm{N_{ens}}-1)} \cdot \left(\Delta \mathbf{x}_0^A\right)^T \cdot \Delta \mathbf{x}_0^A \; . \tag{5.35}$$

Denote the nonzero eigenvalues of the two ensemble covariance matrices by λ_i^B and λ_i^A respectively, $i = 1, 2, \ldots, \mathrm{N_{ens}}$. The nonzero eigenvalues can be efficiently computed by solving small $\mathrm{N_{ens}} \times \mathrm{N_{ens}}$ eigenvalue problems since

$$\Lambda = \mathrm{eig} \underbrace{\left(\Delta \mathbf{x} \cdot \Delta \mathbf{x}^T\right)}_{n \times n} \in \mathbb{R}^n \; , \quad \lambda = \mathrm{eig} \underbrace{\left(\Delta \mathbf{x}^T \cdot \Delta \mathbf{x}\right)}_{\mathrm{N_{ens}} \times \mathrm{N_{ens}}} \in \mathbb{R}^{\mathrm{N_{ens}}} \quad \Rightarrow \quad \Lambda_i = \lambda_i \; , \; i = 1, \cdots, \mathrm{N_{ens}} \; . \tag{5.36}$$

An estimate of the Shannon information content (5.24d) can be given in terms of eigenvalues as follows:

$$\frac{1}{2} \ln \det \mathbb{B}_0 \, \mathbb{A}_0^{-1} = \frac{1}{2} \ln \prod_{i=1}^{\mathrm{N_{ens}}} \left(\frac{\lambda_i^B}{\lambda_i^A}\right) = \frac{1}{2} \sum_{i=1}^{\mathrm{N_{ens}}} \ln \left(\frac{\lambda_i^B}{\lambda_i^A}\right) \; . \tag{5.37}$$

Similarly, the part (5.24b) of the DFS metric can be estimated by

$$\frac{1}{2} \mathrm{trace} \left(\mathbb{B}_0^{-1/2} \, \mathbb{A}_0 \, \mathbb{B}_0^{-1/2}\right) = \frac{1}{2} \sum_{i=1}^{\mathrm{N_{ens}}} \left(\frac{\lambda_i^A}{\lambda_i^B}\right) \; . \tag{5.38}$$

Computational procedure for estimating the Shannon information

One constructs two ensembles of initial conditions, one from the background distribution, and one from the analysis distribution. The nonzero eigenvalues of the corresponding ensemble covariances are computed using (5.36). These eigenvalues are used to estimate the Shannon information content via (5.37) and the DFS information content

via (5.38). The computational procedure is direct - no additional model runs are necessary. However, for a small number of ensemble members, the ensemble covariance eigenvalues may poorly represent the eigenvalues of the true covariances. In this case the resulting estimates of the Shannon or DFS information content are expected to be inaccurate.

5.4.5 Estimation of the Signal information content

In this section we assume a linear system with linear observation operators and Gaussian uncertainties as discussed in Appendix A.3. The analysis state obtained using all the available information is $\mathbf{x}_0^A$, Consider one particular observation $\mathbf{y}_\ell$, remove it from the set of data, and repeat the data assimilation. Let $\mathbf{x}_0^C$ be the analysis state when the data assimilation is carried out *without the observation* $\mathbf{y}_\ell$.

We use the notation of Appendix A.3. Furthermore, denote the contribution of observation ℓ to the right hand side and to the 4D-Var system matrix (A.9) by

$$b_\ell = M_\ell^T \, H_\ell^T \, \mathbb{R}_\ell^{-1} \, \left(\mathbf{y}_\ell - H_\ell \, M_\ell \, \mathbf{x}_0^B \right) \ , \quad D_\ell = M_\ell^T \, H_\ell^T \, \mathbb{R}_\ell^{-1} \, H_\ell \, M_\ell \ .$$

Following equation (A.9) the two 4D-Var problems have the following solutions:

$$\mathbb{A}_0^{-1} \cdot \left(\mathbf{x}_0^A - \mathbf{x}_0^B \right) = b \ , \quad \left(\mathbb{A}_0^{-1} - D_\ell \right) \cdot \left(\mathbf{x}_0^C - \mathbf{x}_0^B \right) = b - b_\ell \ .$$

We assume a case where there are many observations such that the contribution of b_ℓ to the total right hand side vector is relatively small, $b - b_\ell \approx b$, and the contribution of D_ℓ to the total inverse covariance is relatively small, $\mathbb{A}_0^{-1} - D_\ell \approx \mathbb{A}_0^{-1}$. The following approximations are obtained:

$$\mathbb{A}_0^{-1} \cdot \left(\mathbf{x}_0^C - \mathbf{x}_0^B \right) \approx b - b_\ell \ , \quad \mathbb{A}_0^{-1} \cdot \left(\mathbf{x}_0^A - \mathbf{x}_0^C \right) \approx b_\ell \ .$$

The difference in the signal part due to the assimilation of observation $\mathbf{y}_\ell$ is

$$\begin{aligned}
\mathcal{I}_{\mathbf{y}_\ell}^{\text{Signal}} &= \frac{1}{2}\left(\mathbf{x}_0^A-\mathbf{x}_0^B\right)^T \mathbb{B}_0^{-1}\left(\mathbf{x}_0^A-\mathbf{x}_0^B\right)-\frac{1}{2}\left(\mathbf{x}_0^C-\mathbf{x}_0^B\right)^T \mathbb{B}_0^{-1}\left(\mathbf{x}_0^C-\mathbf{x}_0^B\right)\\
&= \frac{1}{2}\left(\mathbf{x}_0^A-\mathbf{x}_0^B\right)^T \mathbb{B}_0^{-1}\left(\mathbf{x}_0^A-\mathbf{x}_0^B\right)-\frac{1}{2}\left(\mathbf{x}_0^C-\mathbf{x}_0^B\right)^T \mathbb{B}_0^{-1}\left(\mathbf{x}_0^A-\mathbf{x}_0^B\right)\\
&\quad +\frac{1}{2}\left(\mathbf{x}_0^A-\mathbf{x}_0^B\right)^T \mathbb{B}_0^{-1}\left(\mathbf{x}_0^C-\mathbf{x}_0^B\right)-\frac{1}{2}\left(\mathbf{x}_0^C-\mathbf{x}_0^B\right)^T \mathbb{B}_0^{-1}\left(\mathbf{x}_0^C-\mathbf{x}_0^B\right)\\
&= \frac{1}{2}\left(\mathbf{x}_0^A-\mathbf{x}_0^C\right)^T \mathbb{B}_0^{-1}\left(\mathbf{x}_0^A-\mathbf{x}_0^B\right)+\frac{1}{2}\left(\mathbf{x}_0^A-\mathbf{x}_0^C\right)^T \mathbb{B}_0^{-1}\left(\mathbf{x}_0^C-\mathbf{x}_0^B\right)\\
&= \frac{1}{2}\left(\mathbb{A}_0^{-1}(\mathbf{x}_0^A-\mathbf{x}_0^C)\right)^T \mathbb{A}_0\,\mathbb{B}_0^{-1}\,\mathbb{A}_0\left(\mathbb{A}_0^{-1}\left(\mathbf{x}_0^A-\mathbf{x}_0^B\right)+\mathbb{A}_0^{-1}(\mathbf{x}_0^C-\mathbf{x}_0^B)\right)\\
&\approx \frac{1}{2}\,(b_\ell)^T\,\mathbb{A}_0\,\mathbb{B}_0^{-1}\,\mathbb{A}_0\,(2b-b_\ell)\\
&\approx b_\ell^T\,\mathbb{A}_0\,\mathbb{B}_0^{-1}\,\mathbb{A}_0\,b\\
&= \left(\mathbf{y}_\ell-H_\ell M_\ell \mathbf{x}_0^B\right)^T \mathbb{R}_\ell^{-1}\,H_\ell\,M_\ell\,\mathbb{A}_0\,\mathbb{B}_0^{-1}\left(\mathbf{x}_0^A-\mathbf{x}_0^B\right)\ .
\end{aligned}$$

Let

$$\begin{aligned}
&\widetilde{\mathbf{x}}_0^A=\mathbb{A}_0\,\mathbb{B}_0^{-1}\,\mathbf{x}_0^A\ ,\quad \widetilde{\mathbf{x}}_0^B=\mathbb{A}_0\,\mathbb{B}_0^{-1}\,\mathbf{x}_0^B\\
&H_\ell\,M_\ell\,\mathbb{A}_0\,\mathbb{B}_0^{-1}\left(\mathbf{x}_0^A-\mathbf{x}_0^B\right)\approx H_\ell\,\widetilde{\mathbf{x}}_\ell^A-H_\ell\,\widetilde{\mathbf{x}}_\ell^B\ .
\end{aligned}\tag{5.39}$$

The contribution of measurement $\mathbf{y}_\ell$ to the signal information can therefore be approximated as:

$$\mathcal{I}_{\mathbf{y}_\ell}^{\text{Signal}} \approx \left(\mathbf{y}_\ell-\mathcal{H}_\ell\left(\mathbf{x}_\ell^B\right)\right)^T \mathbb{R}_\ell^{-1}\left(\mathcal{H}_\ell\left(\widetilde{\mathbf{x}}_\ell^A\right)-\mathcal{H}_\ell\left(\widetilde{\mathbf{x}}_\ell^B\right)\right)\tag{5.40a}$$

$$\mathcal{I}_{\mathbf{y}_\ell}^{\text{Signal}} \approx \left(\mathbf{y}_\ell-\mathcal{H}_\ell\left(\mathbf{x}_\ell^B\right)\right)^T \mathbb{R}_\ell^{-1}\left(\mathcal{H}_\ell\left(\mathbf{x}_\ell^A\right)-\mathcal{H}_\ell\left(\mathbf{x}_\ell^B\right)\right)\tag{5.40b}$$

where the last approximation is rather coarse.

Computational procedure for estimating the Signal information

Two modified initial conditions are computed by (5.39). (If this is not feasible, the background and the analysis initial conditions can be used, at the price of a larger approximation error). The model is run from the modified analysis and the "synthetic observations" $\mathcal{H}_\ell\left(\widetilde{\mathbf{x}}_\ell^A\right)$ are recorded. The model is run again from the modified background and the "synthetic observations" $\mathcal{H}_\ell\left(\widetilde{\mathbf{x}}_\ell^B\right)$ are also recorded (this run is not necessary if one uses (5.40b)). Finally, the model is run from the background state, and the estimates (5.40a) or (5.40b) are evaluated for each data point $\mathbf{y}_\ell$.

5.5 Numerical Experiments

We first illustrate the estimation methodology developed in Section 5.4 with a linear test case with Gaussian uncertainties. Next, we apply the estimation methodology to a 4D-Var data assimilation study with a global chemical transport model. The data assimilation experiment focuses on ozone. We estimate the information content of satellite observations taken at different times using different information theoretic metrics.

5.5.1 A linear test case

In order to illustrate the estimates of various information metrics described in section 5.4 we first consider a linear test case. The model is

$$\mathbf{x}_k = M \cdot \mathbf{x}_{k-1} \ , \quad k = 1, \cdots, 4 \ , \quad \mathbf{x}_k \in \mathbb{R}^{10} \ . \tag{5.41}$$

The model matrix M has eigenvalues log-equally distributed in the interval $[10^{-2}, 10^{2}]$. There are 5 eigenvalues greater than 1 (with the errors growing along the corresponding eigendirections) and 5 eigenvalues smaller than 1 (with the errors decreasing along the corresponding eigendirections). Observations of odd numbered states (1,3,5,7, and 9) are taken at each step. The background errors are normal and characterized by a diagonal background covariance matrix; the standard deviation of the error in each component is 10% of the background mean value. The observation errors are assumed normal and independent of each other; the standard deviation of each observation error is 1% of the reference observation value.

For this problem analytical solutions are available for the analysis state $\mathbf{x}_0^A$ and for the analysis covariance matrix $\mathbb{A}_0$. Based on them a direct evaluation of the different information metrics is possible. The results are summarized in Table 5.1 and show that the ensemble estimates of information metrics are accurate.

Table 5.1: Results with the linear test problem (5.41). The Fisher information is estimated using equation (5.28), DFS using (5.31), Shannon using (5.37), and the signal using (5.40a)

	Direct	$N_{ens} = 10$	$N_{ens} = 10^2$	$N_{ens} = 10^3$	$N_{ens} = 10^4$
Fisher	2.001e+05	1.923e+05	2.138e+05	1.977e+05	1.979e+05
DFS	4.999e+00	4.802e+00	5.336e+00	4.934e+00	4.942e+00
Shannon	2.234e+01	1.998e+01+1.571i	2.222e+01	2.245e+01	2.232e+01
Signal	3.347e+00	3.347e+00	3.347e+00	3.347e+00	3.347e+00

5.5.2 Experimental Setting

The GEOS-Chem simulations are carried out at a resolution of $4^\circ \times 5^\circ$. The dimension of the state space in our simulations is $n \approx 8$ million (72 longitude grid points, times 46 latitude grid points, times 55 vertical levels, times 43 chemical tracers).

The control variables are the initial concentrations of ozone throughout the simulation domain. While GEOS-Chem is capable of performing simulations up to 75 Km (55 vertical levels), the model error increases with height and the model bias is non-negligible in the upper troposphere and into the stratosphere. For this reason we perform data assimilation only up to the first 23 model vertical levels (21 Km) similar to assimilation experiments described in Chapter 3 and 4.

The assimilation time window has a length of 5 days, starting on August 1st, 2006 (00 GMT) and ending on August 6th, 2006 (00 GMT). The observation time window is 4 hours, i.e., the observation operator treats all retrievals available in a 4 hour window as a single data point. Specifically, the observation $\mathbf{y}_i$ at time t_i consists of all the data available for the time interval $[\, t_i - 2\,\text{hours}\,, t_i + 2\,\text{hours}\,]$.

We estimate the information content of ozone profile retrievals from TES when used to improve the ozone initial conditions in GEOS-Chem through 4D-Var data assimilation. The main computational costs come from: (1) the 4D-Var run, which requires 12 iterations of the optimization routine, with each iteration performing a forward and adjoint model run; and (2) an ensemble of 20 additional model runs, including adjoints, to gather the data needed for the estimation of different information content metrics. Concentrations and other time dependent variables are checkpointed during the forward runs, and are read during the adjoint runs. We consider a diagonal background error covariance matrix ($\mathbb{B}_0$), as defined in Chapter 3, in all our experiments for simplicity.

The following simple technique is employed to approximately sample the analysis distribution. We perform data assimilation and compare the background and the analysis fields against the INTEX ozonesonde validation data set. This provides a vertical distribution of mean errors and of their variance. We make the following assumptions: the analysis covariance matrix is diagonal (the correlation length is smaller than one grid size); the relative error reduction realized through data assimilation is similar in all gridpoints at the same vertical level; and the relative error reduction is similar throughout the assimilation window. Under these assumptions the error reduction measured against the INTEX ozonesonde data is representative of the reduction in error at the initial time throughout the entire computational grid. Consequently, the analysis error standard deviation at a given grid point is obtained by scaling the background standard deviation. The scaling factor is the ratio of the standard deviation of the analysis against INTEX data over the standard deviation of the background against INTEX data; the same scaling factor is applied to all grids at the same vertical level. In summary, the analysis mean is provided by the result of the data assimilation. The analysis covariance

matrix is diagonal, with the diagonal entries obtained by scaling the corresponding background variances. The scaling factors are obtained by comparing the background and the analysis against the validation data set. A more sophisticated method for sampling the posterior distribution is described in Appendix A.1.

5.5.3 Information content of TES ozone column retrievals

We exhibit four different sets of results that provide estimates of information content of aggregated and individual observation data sets in the context of 4D-Var data assimilation.

Aggregated information content of all available data

We first compute the aggregated information content of *all* the available data, i.e., of all the TES ozone profile retrievals available within the 5 days assimilation window. Since 4D-Var adjusts the initial conditions of ozone, the information content metrics describe the data impact on reducing the uncertainty at time t_0.

The estimate of the FIM information content (5.28) requires an ensemble of $\mathrm{N_{ens}}$ gradient values. Each gradient λ_0^r is calculated by running the forward and the adjoint models starting from one of the initial conditions $\mathbf{x}_0^r$ drawn from the posterior ensemble (5.27) . The ensemble average of the squared gradient entries is computed following (5.27)

$$\left\langle \lambda_0(i,j,\ell,s_{O3})^2 \right\rangle^A = \frac{1}{\mathrm{N_{ens}}} \sum_{r=1}^{\mathrm{N_{ens}}} \left(\lambda_0^r(i,j,\ell,s_{O3})\right)^2 \ .$$

Using the average squared gradient values and the background error covariance matrix (3.15)–(3.19), the numerical approximation to Fisher information is calculated as

$$\begin{aligned}
\mathcal{I}^{\mathrm{FIM}} &= \left\langle \|\lambda_0\|^2 \right\rangle^A - \mathrm{trace}\left(\mathbb{B}_0^{-1}\right) \\
&= \sum_{i=1}^{Nlon} \sum_{j=1}^{Nlat} \sum_{\ell=1}^{Nlev} \left(\left\langle \lambda_0(i,j,\ell,s_{O3})^2 \right\rangle^A - \frac{1}{\sigma_\ell^2} \right) \\
&= \sum_{\ell=1}^{Nlev} \mathcal{I}_\ell^{\mathrm{FIM}} \\
\mathcal{I}_\ell^{\mathrm{FIM}} &= \sum_{i=1}^{Nlon} \sum_{j=1}^{Nlat} \left(\left\langle \lambda_0(i,j,\ell,s_{O3})^2 \right\rangle^A - \frac{1}{\sigma_\ell^2} \right), \quad \ell = 1,2,\cdots,Nlev \ .
\end{aligned}$$

The first relation provides the scalar value for the FIM information content of all available observations. The last relation provides the Fisher information content relative to the

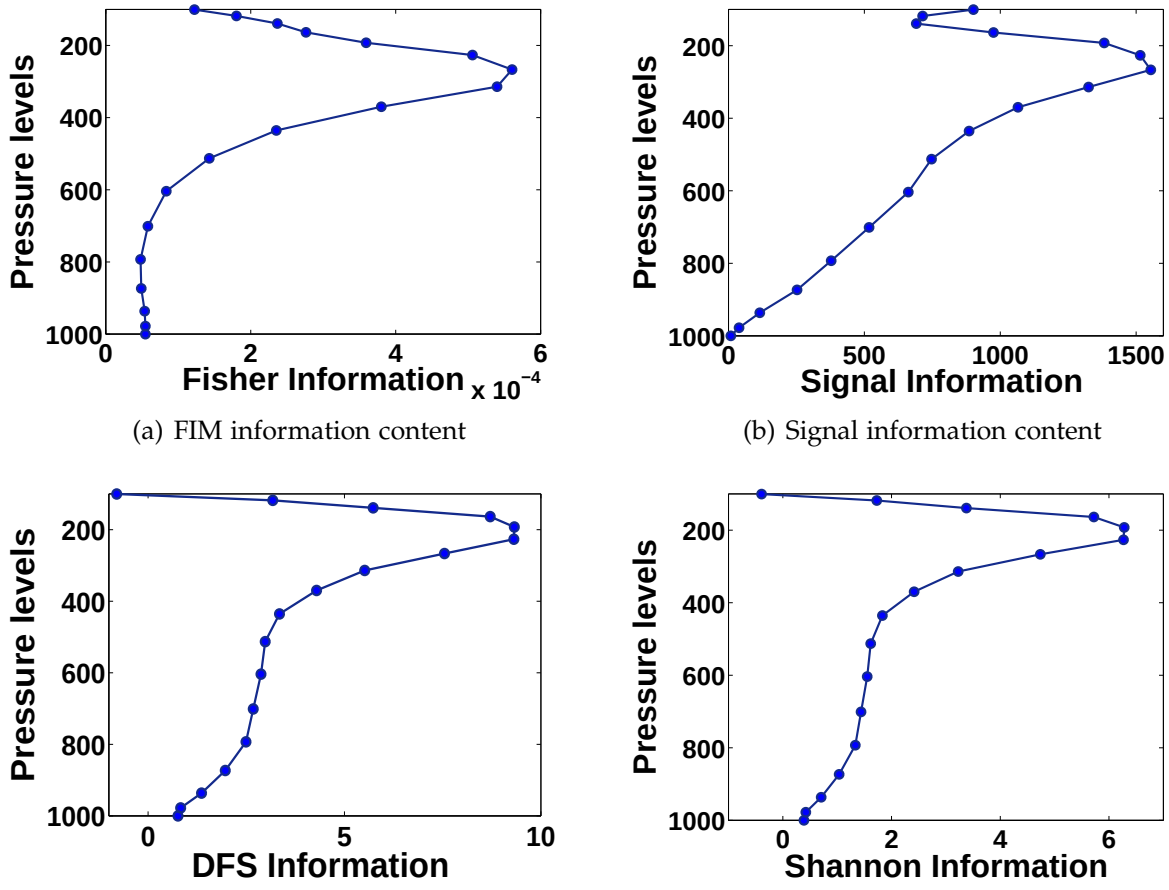
vertical level ℓ of the model; this is a metric of how level ℓ benefits from the assimilation of the data. It is important to note that the breakdown of the information by vertical levels is possible only under the assumption that there is no correlation among errors at different levels. While this is not the case in general, the breakdown provides insight into how the uncertainty is reduced in models with varying pressure levels. The results are shown in Figure 5.1(a). The FIM information content is large between 400 hPa and 200 hPa, and is small for all other pressure levels. The uncertainty in the initial ozone field is reduced the most in the higher tropospheric area, according to the FIM metric; the levels between 400 hPa and 200 hPa benefit the most from the assimilation of TES ozone column retrievals.

The signal information content of all the observations (5.25) is the background cost function evaluated at the optimal initial condition. Using the formula for background error covariance matrix(3.19), the level-wise signal contribution could be defined as

$$\begin{aligned}
\mathcal{I}^{\text{Signal}} &= \frac{1}{2}\left(\mathbf{x}_0^A - \mathbf{x}_0^B\right)^T \mathbb{B}_0^{-1}\left(\mathbf{x}_0^A - \mathbf{x}_0^B\right) \\
&= \frac{1}{2}\sum_{i=1}^{Nlon}\sum_{j=1}^{Nlat}\sum_{\ell=1}^{Nlev}\left(\frac{\mathbf{x}_0^A(i,j,\ell,s_{O3}) - \mathbf{x}_0^B(i,j,\ell,s_{O3})}{\sigma_\ell}\right)^2 \\
&= \sum_{\ell=1}^{Nlev}\mathcal{I}_\ell^{\text{Signal}} \\
\mathcal{I}_\ell^{\text{Signal}} &= \frac{1}{2}\sum_{i=1}^{Nlon}\sum_{j=1}^{Nlat}\left(\frac{\mathbf{x}_0^A(i,j,\ell,s_{O3}) - \mathbf{x}_0^B(i,j,\ell,s_{O3})}{\sigma_\ell}\right)^2, \quad \ell = 1,2,\cdots,Nlev\,.
\end{aligned}$$

The results for the Signal information content of all observations are shown in Figure 5.1(b). The Signal information content is the largest between 400 hPa and 200 hPa, which correlates well with the distribution of the FIM information. The Signal information content decreases (almost) linearly for higher pressure levels, and approaches zero near the ground level. This indicates that the assimilation of TES ozone column data does little to reduce the uncertainty in ozone concentrations near ground level.

The DFS information (5.24b) and the Shannon information content (5.24d) are estimated from ensemble covariance eigenvalues using the formulas (5.38) and (5.37, respectively. The results for DFS are shown in Figure 5.1(c); the results for Shannon information are shown in Figure 5.1(d). The two information metrics have highest values between 400 hPa and 200 hPa indicating larger uncertainty reduction in the upper troposphere and slight reduction in the mid and lower troposphere.



(a) FIM information content

(b) Signal information content

(c) DFS information content

(d) Shannon information content

Figure 5.1: The aggregated information content *of all observations*, as measured by different information theoretic metrics. The breakdown of information content by vertical layers is possible if the vertical error correlations are negligible.

The Signal information content

The signal information content of individual data points $\mathbf{y}_\ell$ is estimated using the formula (5.40b). No gradient calculations are necessary. The estimate depends only on the innovation vectors associated with the background trajectory $d_\ell^B = \mathbf{y}_\ell - H_\ell \mathbf{x}_\ell^B$, and with the analysis trajectory $d_\ell^A = \mathbf{y}_\ell - H_\ell \mathbf{x}_\ell^A$. Equation (5.40b) can be written as

$$\mathcal{I}_{\mathbf{y}_\ell}^{\text{Signal}} \approx \left(d_\ell^B \right)^T \mathbb{R}_\ell^{-1} \left(d_\ell^B - d_\ell^A \right) . \tag{5.42}$$

We first perform a forward model run starting from the optimal initial condition $\mathbf{x}_0^A$ and save the innovation vectors d_ℓ^A for each observation location and for all observation windows. We then perform a second run starting with the background initial condition

$\mathbf{x}_0^B$. During this run we compute the innovation vectors d_ℓ^B, and, using the saved d_ℓ^A values, we also compute the Signal information content (5.42).

The time series of the Signal information content per each observation window is shown in Figure 5.2. The difference between the contribution of observations taken earlier and taken later during the assimilation window is small. This difference is relatively large for the DFS information metric, as observed in Figure 5.5.

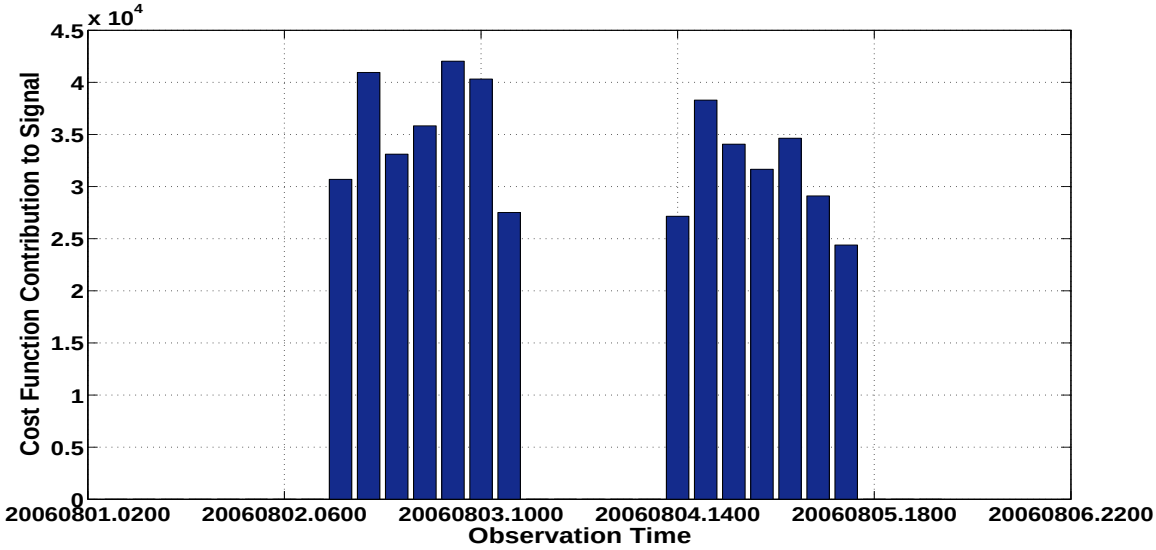
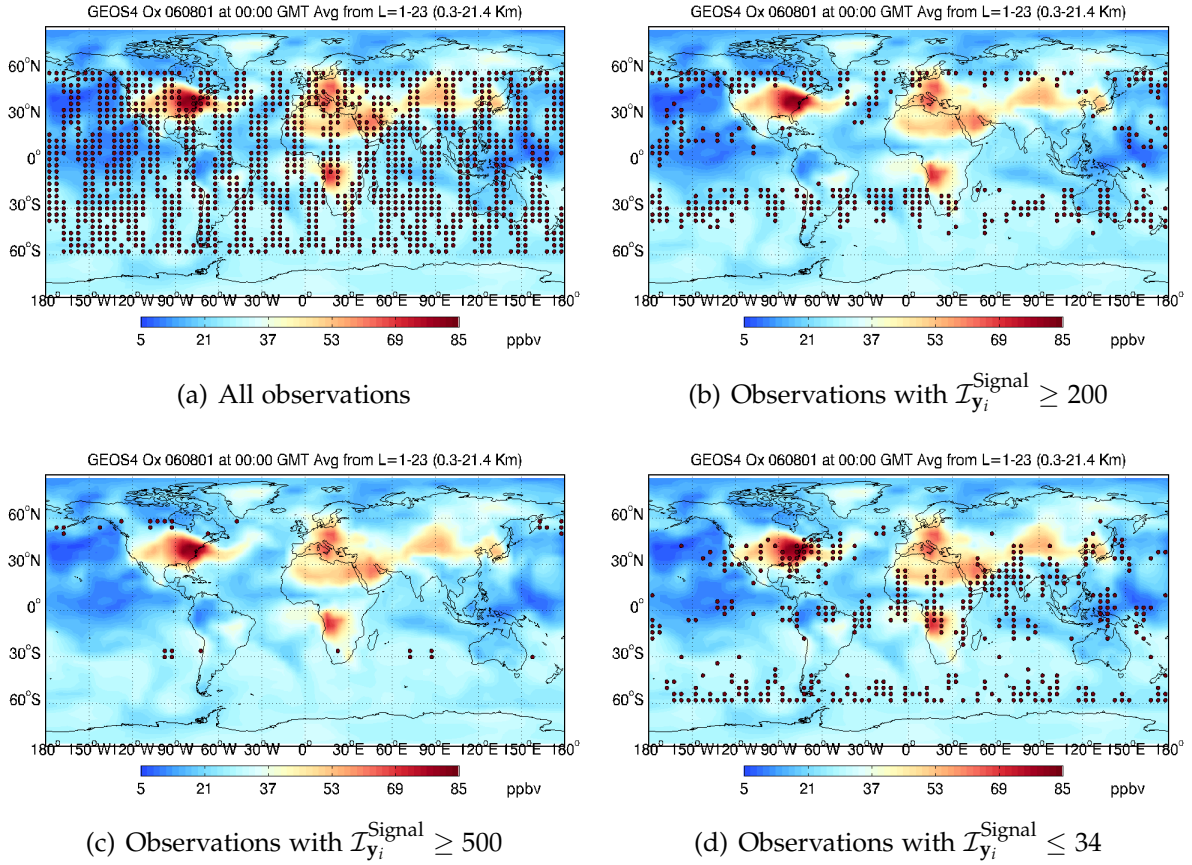


Figure 5.2: The Signal information content of observations taken at different times within the assimilation window.

We next relate the signal information content with the location of observations. What makes this approach significant is that it provides a direct way to locate observations with higher information contents. Figure 5.3 presents the locations of observations with a Signal information content greater than a given threshold. We use IDL visualization software (`http://www.ittvis.com`) in combination with Global Atmospheric Model Analysis Package to overlay the global ozone distribution on August 1st, 00:00 GMT, with the location of the (subset of) observations..

First three panels in Figure 5.3 use progressively larger thresholds, thus successively filtering out the "less important" data points. The top left panel represents all 1,342 observations available over the 5 days assimilation window. The bottom right panel reflects points with lesser contribution to the assimilation system. The plots reveal that measurements taken about 60° N and 30° S have the highest signal information content.

Assimilation of subsets of observations: We now investigate the relationship between the estimated information content and the benefit for the 4D-Var data assimilation. For this we repeat the 4D-Var data assimilation using only subsets of observations. The subsets are filtered based on signal information content. All data assimilation experiments



(a) All observations

(b) Observations with $\mathcal{I}_{\mathbf{y}_i}^{\text{Signal}} \geq 200$

(c) Observations with $\mathcal{I}_{\mathbf{y}_i}^{\text{Signal}} \geq 500$

(d) Observations with $\mathcal{I}_{\mathbf{y}_i}^{\text{Signal}} \leq 34$

Figure 5.3: The location of the most important observations, filtered by their signal information content.

use the same covariance matrices and the background field $\mathbf{x}_0^B$. Specifically, we use the following subsets of observations: $\mathcal{I}_{\mathbf{y}_i}^{\text{Signal}} \geq 500$ units (32 data points, 2.3% of all observations), $\mathcal{I}_{\mathbf{y}_i}^{\text{Signal}} \geq 200$ units (363 data points, 27% of all observations) and $\mathcal{I}_{\mathbf{y}_i}^{\text{Signal}} \leq 34$ units (367 grid points, 27% of all observations).

Figure 5.4 presents the results of the different data assimilation experiments. The errors are measured against the independent data set of INTEX Ozonesonde Network Study 2006 (IONS-6). The leftmost panel presents the mean ozone concentration vertical profiles. The central panel shows the mean errors, i.e., the relative difference between the mean model profiles and ozonesondes. The rightmost panel presents the corresponding error standard deviations. The results reveal that the observations with a higher signal information content contribute more to the 4D-Var analysis. The quality of the analysis using only the top 2.3% observations ($\mathcal{I}_{\mathbf{y}_i}^{\text{Signal}} \geq 500$ units) is similar to the quality of the analysis using the bottom 27% observations ($\mathcal{I}_{\mathbf{y}_i}^{\text{Signal}} \leq 34$ units). Detailed discussion and results of 4D-Var data assimilation with all the observations is provided in [Singh et al.,

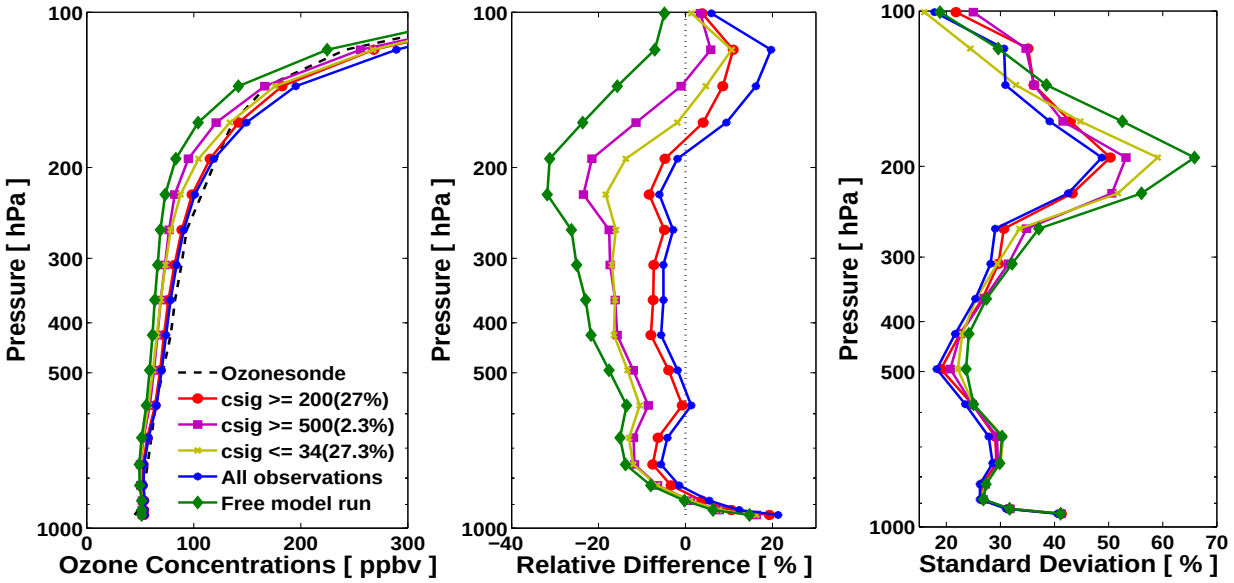


Figure 5.4: Plot of ozonesonde data, free model run, and 4D-VAR analysis trajectories obtained using subsets of observation points. The subsets are selected according to their signal information content.

2010].

The DFS information content

Loosely speaking, the DFS metric (as discussed in Section 5.3.3) indicates the number of states that benefit from the assimilation of observations. The closer the DFS is to the total number of model states, the more information the observations have brought into the system through data assimilation. While the information content measures the change in the mean field obtained through assimilation, the DFS measures the relative decrease in the error (co-)variance through assimilation. Thus the two metrics measure different aspects of the data assimilation benefits.

The DFS information content for individual data points $\mathbf{y}_i$ is estimated using equation (5.31). Recall that in our simulations one data point $\mathbf{y}_i$ consists of all the ozone column retrievals available in the 4 hours interval $[\, t_i - 2 \text{ hours}\,,\, t_i - 2 \text{ hours}\,]$. As the Aura satellite orbits the Earth the observations are taken over different locations and at different times of day. It is therefore expected that some data points will contain more information than other, i.e., are more useful in reducing uncertainty when assimilated. We utilize the data from the ensemble of $\mathrm{N}_{\mathrm{ens}} + 1$ model runs initialized with states drawn from the analysis distribution (this is the same set of runs used for the aggregated information

content calculations). During each of the runs the cost function contribution of each data point is saved. These ensemble of results is used to calculate the DFS information content according to (5.31).

Figure 5.5 presents the DFS information content of the data in each observation window. The data in the observation window 16:00 GMT - 20:00 GMT, August 3rd, 2006 has the highest DFS information content. The DFS information content decreases with time, and the impact of the observations taken later in the assimilation window is smaller. This trend is similar to the one of the signal information content.

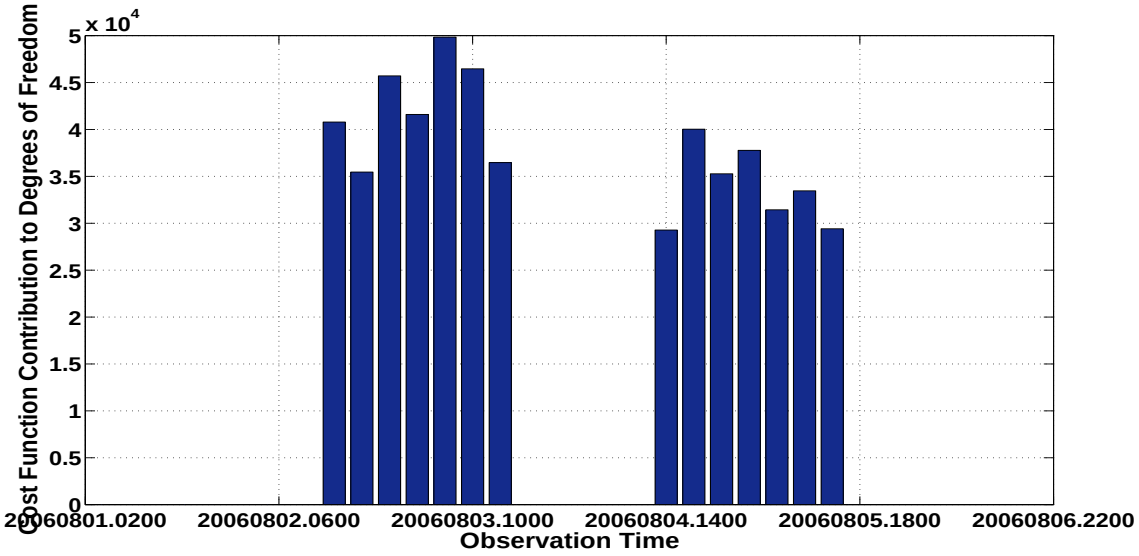
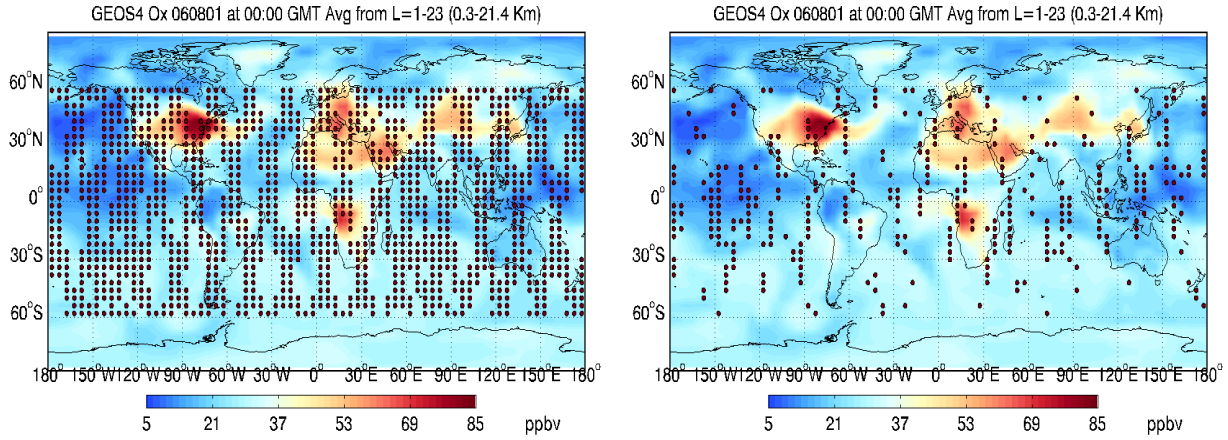


Figure 5.5: The DFS information content of observations taken at different times within the assimilation window.

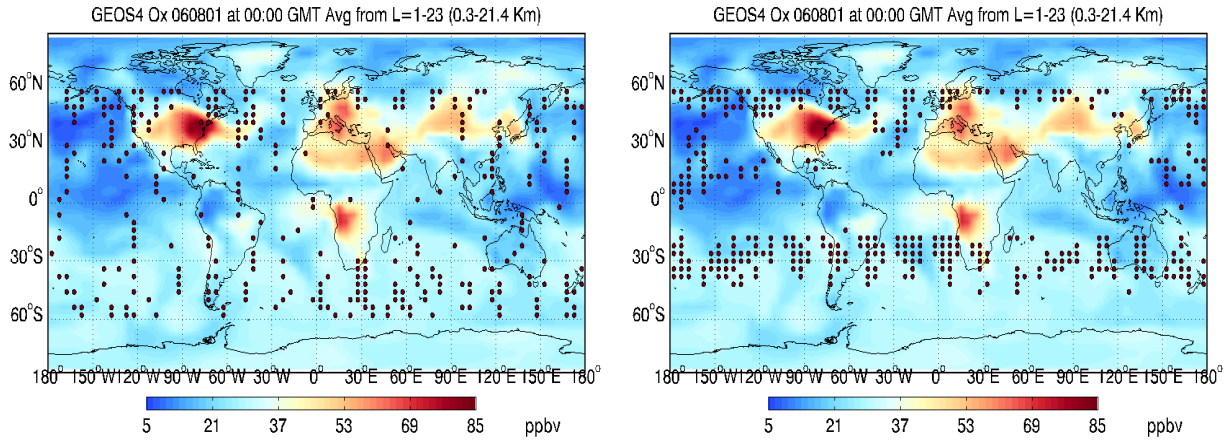
We next study the DFS contribution of each observation point to the assimilation results. Specifically, the data points are classified into subsets according to their estimated DFS information values. Figure 5.6 shows the location of different observation subsets plotted over the global ozone distribution (averaged over the first 23 levels on August 1st, 2006, 00 GMT). Panel (a) represents all 1,342 observation points available. Panel (b) depicts the location of observations with $\mathcal{I}_{\mathbf{y}_i}^{\mathrm{DFS}} \geq 300$ (27% of all the data points). There is no apparent spatial structure. Panel (c) presents the location of observations with $\mathcal{I}_{\mathbf{y}_i}^{\mathrm{DFS}} \leq 10$ (27% of all the data points). These points have higher densities between $30^\circ - 60^\circ$ both North and South. Panel (d) presents the location of observations with high signal content for reference; a simple visual comparison reveals that points with lesser DFS information are collocated with the points with higher signal information.

Assimilation of subsets of observations. We perform several data assimilation experiments using only subsets of observations, filtered by their estimated DFS information



(a) All observations

(b) Observations with $\mathcal{I}_{\mathbf{y}_i}^{\text{DFS}} \geq 300$



(c) Observations with $\mathcal{I}_{\mathbf{y}_i}^{\text{DFS}} \leq 10$

(d) Observations with $\mathcal{I}_{\mathbf{y}_i}^{\text{Signal}} \geq 200$

Figure 5.6: The location of the most important observations, filtered by their DFS information content.

content. The results are presented in Figure 5.7. The assimilation results using the top 27% data points (points with highest DFS) show a decent performance. However, the assimilation results using the bottom 27% data points (points with $\mathcal{I}_{\mathbf{y}_i}^{\text{DFS}} \leq 10$) are remarkable. The performance is considerably better than that of the top DFS subset, and is also better than that obtained with assimilating the top signal information contributors. This is not completely unexpected since the low DFS data points are located among the high signal data points.

We cannot fully explain these results. We note that for this experiment the data points that contribute most to changing the mean field (highest signal information content)

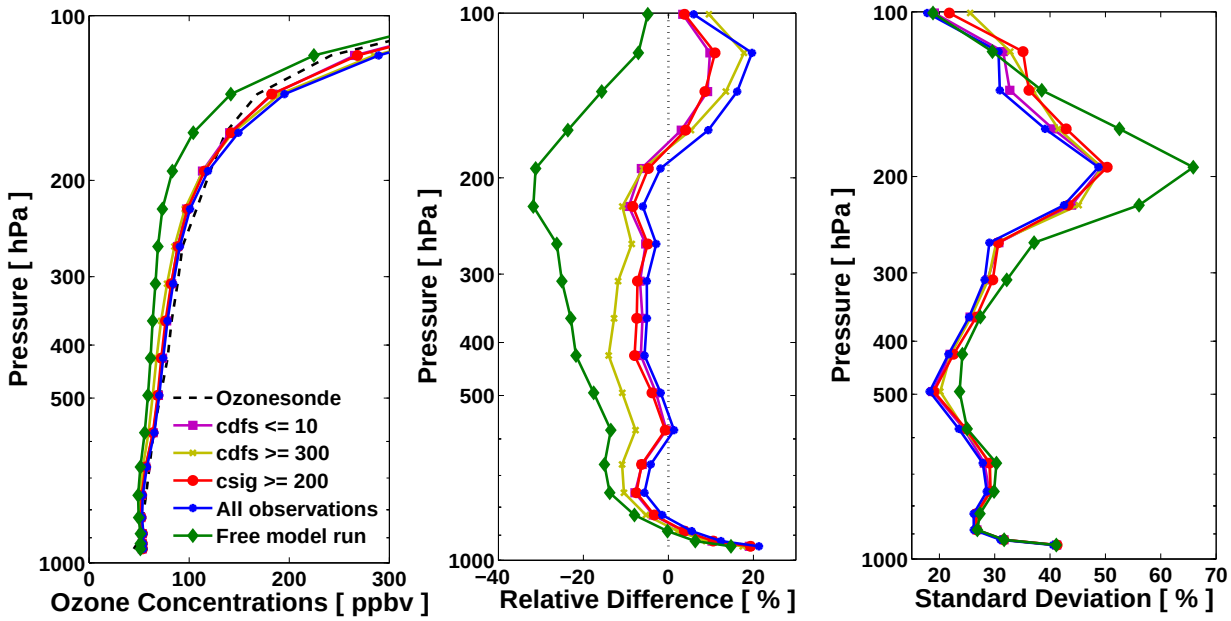


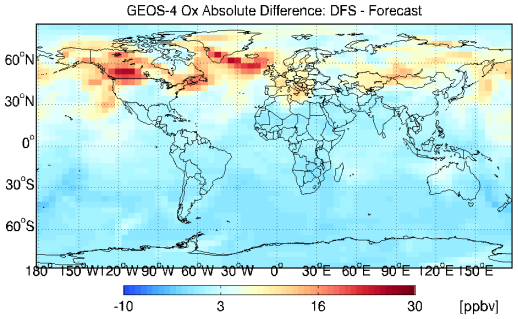
Figure 5.7: Plot of ozonesonde data, free model run, and 4D-VAR analysis trajectories obtained using subsets of observation points. The subsets are selected according to their DFS information content.

seem to contribute least to decreasing the error (co-)variance.

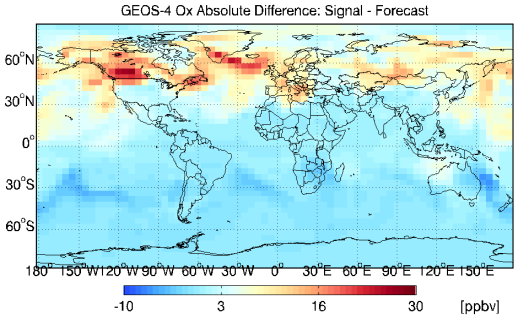
A direct comparison of different assimilation results is shown in Figure 5.8. Specifically, we plot the differences in global ozone concentrations at the beginning of the assimilation window (00:00 GMT on August 6, 2006) averaged over the first 10 GEOS-Chem vertical levels. Panels (a)-(b) show differences between the 4D-Var analysis fields and the model forecast (solution without data assimilation); the analyses use observation data points with $\mathcal{I}_{\mathbf{y}_i}^{\text{DFS}} \leq 10$ and with $\mathcal{I}_{\mathbf{y}_i}^{\text{Signal}} \geq 200$. Panels (c)-(d) present the differences between the 4D-Var analysis fields using all observation points and using observation data points with $\mathcal{I}_{\mathbf{y}_i}^{\text{DFS}} \leq 10$ and with $\mathcal{I}_{\mathbf{y}_i}^{\text{Signal}} \geq 200$. Panels (e)-(f) show absolute and relative differences between 4D-Var analyses using points with $\mathcal{I}_{\mathbf{y}_i}^{\text{DFS}} \leq 10$ and with $\mathcal{I}_{\mathbf{y}_i}^{\text{Signal}} \geq 200$.

Virtual ground-level observations

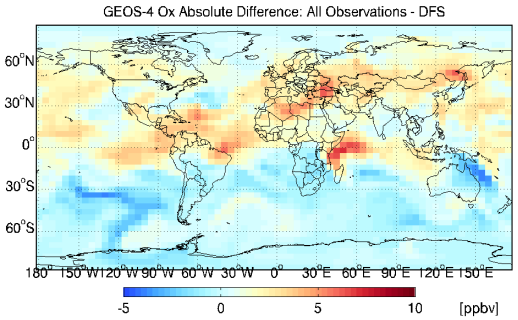
So far we have analyzed the information content of real data: the ozone profile retrievals from TES. We next illustrate the use of the proposed methodology to asses the potential impact of *virtual* observations. This is useful for planning new field campaigns, and for guiding the design of new observing networks.



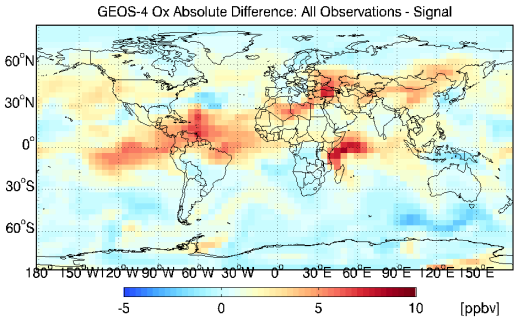
(a) Absolute difference between the 4D-Var analysis using data points with $\mathcal{I}_{\mathbf{y}_i}^{\text{DFS}} \leq 10$ and the free model run



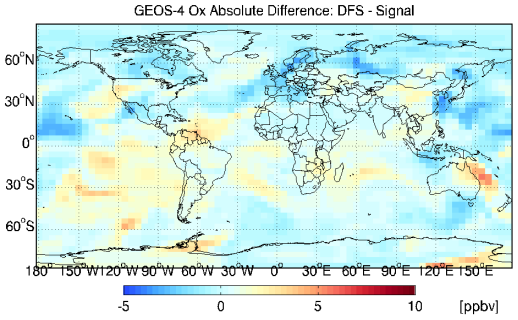
(b) Absolute difference between the 4D-Var analysis using data points with $\mathcal{I}_{\mathbf{y}_i}^{\text{Signal}} \geq 200$ and the free model run



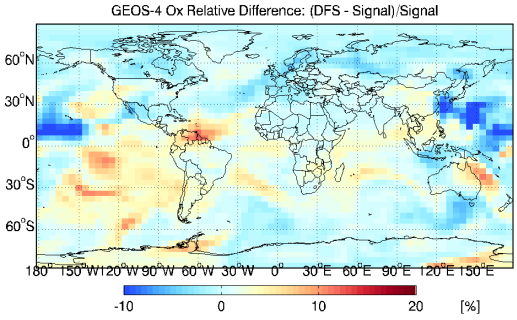
(c) Absolute difference between the 4D-Var analysis using all observations and data points with $\mathcal{I}_{\mathbf{y}_i}^{\text{DFS}} \leq 10$



(d) Absolute difference between the 4D-Var analysis using all observations and data points with $\mathcal{I}_{\mathbf{y}_i}^{\text{Signal}} \geq 200$



(e) Absolute difference between the 4D-Var analyses using data points with $\mathcal{I}_{\mathbf{y}_i}^{\text{DFS}} \leq 10$ and with $\mathcal{I}_{\mathbf{y}_i}^{\text{Signal}} \geq 200$



(f) Relative difference between the 4D-Var analyses using data points with $\mathcal{I}_{\mathbf{y}_i}^{\text{DFS}} \leq 10$ and with $\mathcal{I}_{\mathbf{y}_i}^{\text{Signal}} \geq 200$

Figure 5.8: Direct comparison of different assimilation results. Differences in global ozone concentrations are shown at 00:00 GMT on August 6, 2006 and averaged over the first 10 GEOS-Chem vertical levels.

This section focuses on virtual observations taken at ground level. The concentrations of the analysis field $\mathbf{x}^A$ provide the virtual observations. We perform a forward model run starting from $\mathbf{x}_0^B$ and compute the following approximation of the signal information content at hourly intervals

$$\mathcal{I}_{\text{ground}}^{\text{Signal}}\left(\mathbf{x}^B\right) = \frac{1}{2}\left(\mathbf{x}_{\text{ground}}^B - \mathbf{x}_{\text{ground}}^A\right)^T \mathbb{G}^{-1}\left(\mathbf{x}_{\text{ground}}^B - \mathbf{x}_{\text{ground}}^A\right) \tag{5.43}$$

Note that (5.43) is derived from equation (5.42) when the observation data is replaced by the analysis field, and when the observation operator selects the ground level ozone concentrations. The error covariance matrix $\mathbb{G}$ of the virtual observations is diagonal; the standard deviation of each virtual observation is 10% of the analysis field. Figure 5.9 presents the time series of the signal information content of the virtual ground observations. The information level increases initially, reaches a peak on August 2nd, 2006, 18:00 GMT, and then decreases. Note that the peak information time for virtual ground level observations is the same as the peak DFS information time for TES ozone column retrievals. This indicates that the ground level observations (possibly) taken on August 2nd at 18:00 GMT are most useful for the assimilation scenario under consideration.

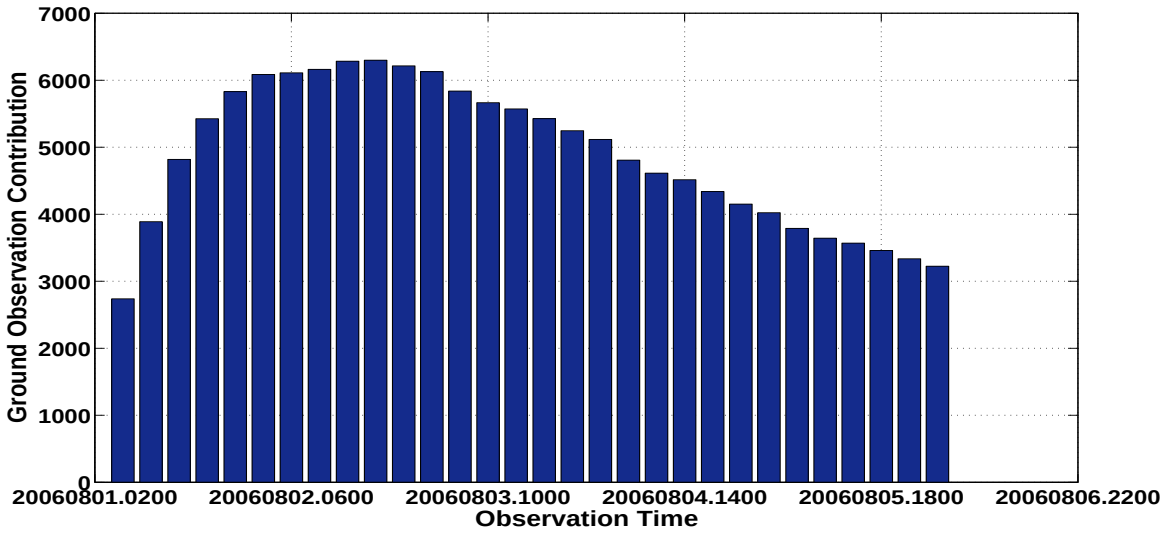
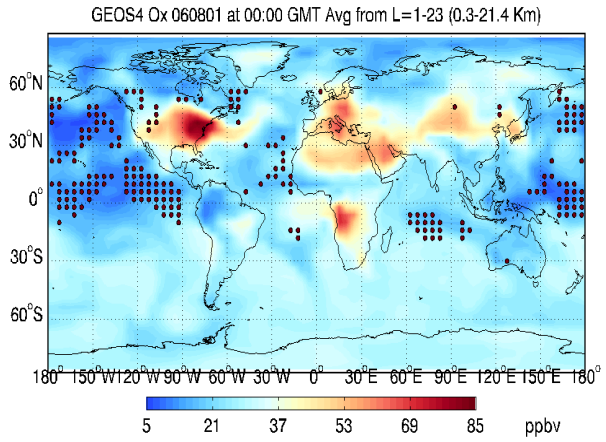
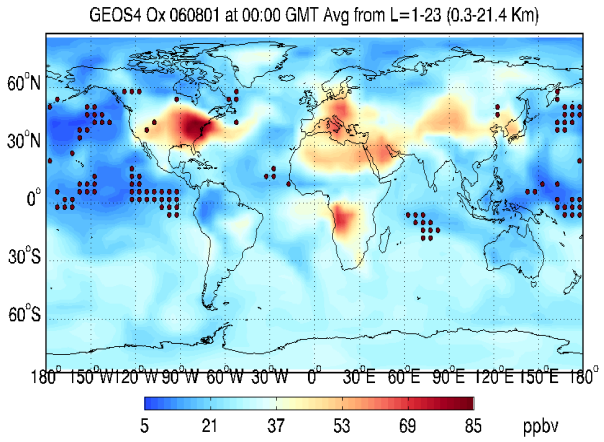


Figure 5.9: Signal information content of virtual ground level observations during the assimilation window.

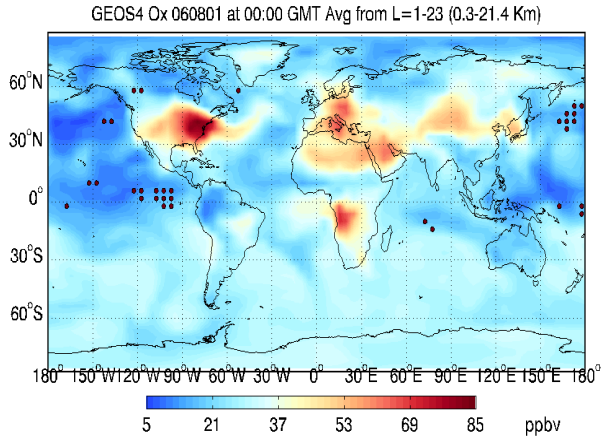
Figure 5.10 plots the locations of the most important virtual ground level observations, ranked based on their signal information content. These locations are overlaid on top of the global ozone distribution on August 1st, 2006, 00:00 GMT. The maximum signal information is associated with the region between 60° N and 30° S. The reason for this scattering in ground observation case could be attributed to the northern and southern hemisphere subtropical jet streams. The few locations with the largest signal information are located around the Equator and at about 45° N.



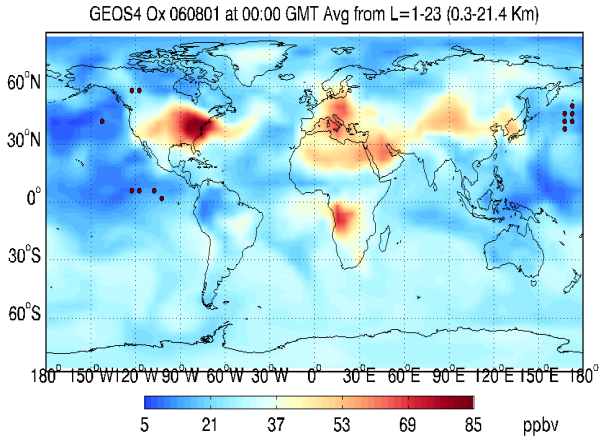
(a) Virtual ground observations with $\mathcal{I}_{\text{ground}}^{\text{Signal}} \geq 3$



(b) Virtual ground observations with $\mathcal{I}_{\text{ground}}^{\text{Signal}} \geq 5$



(c) Virtual ground observations with $\mathcal{I}_{\text{ground}}^{\text{Signal}} \geq 10$



(d) Virtual ground observations with $\mathcal{I}_{\text{ground}}^{\text{Signal}} \geq 15$

Figure 5.10: The location of virtual ground level observations with the largest signal information content.

5.6 Conclusions and Future Work

This chapter discusses a characterization of the information content of observations in the context of four dimensional variational (4D-Var) data assimilation framework. The ability to characterize the usefulness of different data points is important for analyzing the effectiveness of the assimilation system, for data pruning, and for the design of future sensor systems.

Several metrics from information theory are used to quantify the information content of data, including the trace of the Fisher information matrix, the Shannon information, the relative entropy, the signal information, and the degrees of freedom for signal. In the Gaussian case the signal information measures the benefit of data assimilation in terms of adjusting the mean of the distribution. Fisher, Shannon, and DFS all measure the benefit of data assimilation in terms of decreasing the (co-)variance of the error. Relative entropy offers a combination of metrics.

The analysis is carried out under the assumptions that errors have Gaussian distributions and that the model dynamics is linear. The analysis reveals that the information content of observations is intimately related to the statistics of the variational cost function and its gradient. These statistics are obtained with respect to the analysis probability distribution. The theoretical results lead to a new computational procedure to estimate the information content of various observations in the context of 4D-Var. After data assimilation is complete, and ensemble of simulations is run with the initial conditions drawn from the posterior probability distribution. Mean values of the adjoint norms are used to estimate the trace of the Fisher information matrix. The mean value of the observation part of the cost function, minus its value for the analysis, is used to estimate the DFS information content. Scaled dot products between the background innovation and the difference between the background and the analysis innovations provide estimates of the signal information content.

The estimates require a sampling from the posterior distribution, which is not readily available in 4D-Var data assimilation. Different approximate methods are possible to obtain analysis samples. Here we use a normal distribution with the mean given by the assimilation result, a diagonal covariance matrix, and the analysis variances obtain by properly scaling the background variances. The error ratios obtained by comparing the model results against an independent data set are used to determine the scaling factors. More sophisticated methods for sampling the posterior distribution are possible, e.g., see Appendix A.1. Another approach that we plan to explore in the future is based on the Hessian of the cost function (or its quasi-Newton approximation) which is related to the inverse of the posterior covariance [Thacker, 1989; Gejadze et al., 2008].

The information content estimation approach is illustrated on linear test problem. Next, the approach is applied to a global ozone data assimilation problem using TES satellite observations and the GEOS-Chem chemical transport model. The quality of the assim-

ilation is assessed by comparing the results against an independent data set (INTEX ozonesonde measurements). The observations with the highest signal information content are roughly along the latitudes 30° S and 60° N. The assimilation results using the top 27% data points (with the highest signal information) are similar to the results obtained using all the data points. The approach is extended to asses the potential benefit of ground level observations, and to select those ground level locations that are associated with the highest information content. The observations with the hight estimated DFS information are located throughout the globe, while the observations with low DFS information content are at similar locations as the ones with high signal content. Moreover the assimilation results are better when the data points with low DFS are used. It seems that the observations that contribute most to changing the mean are the ones that contribute least to reducing the error variance. We do not fully understand this result. Future work is needed to explain the relationship between the information content of observations measured by the signal and by the DFS metrics.

The assumptions and approximations made during the analysis and computations impact the accuracy of the information content estimates. While the analysis assumes normal error distributions and a linear dynamics, it is desirable to apply the methodology to nonlinear systems and arbitrary uncertainty distributions. The analysis distribution is not explicitly available, samples are taken from distributions that only approximate the analysis under certain assumptions. Finally, relatively small ensembles lead to relatively large sampling errors. Future effort will focus on quantifying the impact that each of these issues (nonlinearity, non-normality, approximate posterior distributions, and small samples) has on the accuracy of the information content estimates.

Chapter 6

Quasi 4D-VAR: An Approach Towards Building a Cost Efficient Global Assimilation System

Abstract

Data assimilation obtains improved estimates of the state of a physical system by combining imperfect model results with sparse and noisy observations of reality. In the four dimensional variational (4D-Var) framework data assimilation is formulated as an optimization problem. A numerical solution is obtained using gradient based optimization methods. The 4D-Var gradient requires the linearized observation operator, and transposed derivative of future states with respect to the initial conditions. The 4D-Var gradient can be obtained effectively by forcing the adjoint model with observation increments, and running it backwards in time. The construction of the adjoint model requires considerable development effort. Moreover, running the adjoint model requires considerable CPU time (typically, a small multiple of the time needed to run the forward model).

In this chapter we propose the *quasi 4D-Var* (Q4D-Var), a technique which uses approximations of the adjoint model in order to decrease the development effort, and to reduce the computational time associated with the computation of the 4D-Var gradient. The approach is illustrated on a global chemical data assimilation problem using satellite observations and the GEOS-Chem chemical transport model.

6.1 Introduction

The variational approach to data assimilation is rooted in control theory, and formulates data assimilation as a minimization problem of a cost functional that measures the model-observations mismatch. In the three dimensional variational (3D-Var) approach the observations available at a certain time are used to correct the state at that time. In the four dimensional variational (4D-Var) approach all observations available within an assimilation time window are considered simultaneously and used to adjust the initial state (state at the beginning of the window). The Kalman filter approach to data assimilation is rooted in statistical estimation theory and provides the analysis covariance together with the best state estimate. Suboptimal Kalman filters employ different approximations of the covariances in order to make the computations feasible with large models.

Among the variational techniques, the three dimensional version (3D-Var) is easier to implement as it does not involve the constructing model adjoints, and is computationally inexpensive to run. The four dimensional variational method (4D-Var) is more accurate but requires an adjoint construction, and more computational time and memory to run. Assimilation techniques that are more accurate than 3D-Var but less expensive than 4D-Var have the potential to provide important practical benefits. We propose an assimilation strategy that is based on 4D-Var, but uses inexact gradient information. We call this approach Quasi 4D-Var (Q4D-Var). Results with global chemical data assimilation indicate that Q4D-Var is slightly more expensive than 3D-Var, and that the quality of the Q4D-Var analyses is comparable to that of strongly-constrained 4D-Var analyses. The difference between the model adjoint and the inexact gradients determines the accuracy of the approach. The quality of the Q4D-Var analysis is expected to deteriorate with increase of the assimilation window length.

This chapter is organized as follows. Section 6.2 reviews the variational approach and discusses the differences between 3D-Var and 4D-Var. Section 6.3 introduces the Q4D-Var approach and Section 6.4 assesses it performance on a global chemical data assimilation study using real satellite data. Conclusions and future work are presented in Section 6.5.

6.2 Variational Data Assimilation

Let us recall the variational data assimilation with details on the requirements for 3D-Var and 4D-Var systems.

Consider that the true state of the system $\mathbf{x}^t \in \mathbb{R}^n$ is unknown and needs to be estimated form the available information. In order to obtain an estimate of $\mathbf{x}^t$ *data assimilation*

combines three different sources of information, as follows.

The background (prior) information describes the uncertainty with which one knows $\mathbf{x}^t$ at a given moment, before any (new) measurements are taken. The current best estimate of the true state $\mathbf{x}^B$ is called the apriori, or the *background state*. The background estimation errors $\varepsilon^B = \mathbf{x}^B - \mathbf{x}^t$ are characterized by the *background error covariance matrix* $\mathbb{B} \in \mathbb{R}^{n \times n}$. A typical assumption is that the random background errors have a normal probability density, $\varepsilon^B \in \mathcal{N}(0, \mathbb{B})$. With many nonlinear models this normality assumption is difficult to justify, but is nevertheless widely used because of its convenience.

The model $\mathcal{M}$ encapsulates our knowledge about physical and chemical laws that govern the evolution of the system. The model evolves an initial state $\mathbf{x}_0 \in \mathbb{R}^n$ at the initial time t_0 to future state values $\mathbf{x}_i \in \mathbb{R}^n$ at future times t_i,

$$\mathbf{x}_i = \mathcal{M}_{t_0 \to t_i} \mathbf{x}_0. \tag{6.1}$$

The size of the state space in realistic chemical transport models is very large. For example, a GEOS-Chem simulation at the $2^o \times 2.5^o$ horizontal resolution has $n \in \mathcal{O}\left(10^8\right)$ variables.

Observations represent snapshots of reality available at several discrete time moments. Specifically, measurements $\mathbf{y}_i \in \mathbb{R}^m$ of the true state are taken at times t_i, $i = 1, \cdots, N$

$$\mathbf{y}_i = \mathcal{H}\left(\mathbf{x}_i^t\right) - \eta_i^{\text{obs}}. \tag{6.2}$$

The observation operator $\mathcal{H}$ maps the state space onto the observation space. In many practical situations $\mathcal{H}$ is a highly nonlinear mapping (as is the case, e.g., with satellite observation operators). Usually the observations are sparsely distributed, and the number of observations is small compared to the dimension of the state space, $m \ll n$. The measurement (instrument) errors are denoted by η_i^{obs}.

Equation (6.2) relates the true state with the observations. In order to relate the model state to observations we also consider the relation

$$\mathbf{y}_i = \mathcal{H}\left(\mathbf{x}_i\right) - \varepsilon_i^{\text{obs}}. \tag{6.3}$$

where the observation operator now acts on the model predicted state. The *observation error* term $\varepsilon_i^{\text{obs}}$ accounts for both the measurement (instrument) errors, as well as representativeness errors (i.e., errors in the accuracy with which the model can reproduce reality). Typically observation errors are assumed to be unbiased and normally distributed

$$\varepsilon_i^{\text{obs}} \in \mathcal{N}\left(0, \mathbf{R}_i\right) \ , \quad i = 1, \cdots, N \ . \tag{6.4}$$

Moreover, observation errors at different times ($\varepsilon_i^{\text{obs}}$ and $\varepsilon_j^{\text{obs}}$ for $i \neq j$) are assumed to be independent.

Based on these three sources of information data assimilation computes the analysis *analysis state* $\mathbf{x}^A$. The analysis estimation errors $\varepsilon^A = \mathbf{x}^A - \mathbf{x}^t$ are characterized by the *analysis error covariance matrix* $\mathbb{A} \in \mathbb{R}^{n \times n}$.

If both the the background and the observation errors are Gaussian, and the error propagation through the model (6.6) is linear, then he probability density of the analysis (estimation) errors ε^A is also Gaussian,

$$\varepsilon^A = \mathbf{x}^A - \mathbf{x}^t \in \mathcal{N}(0, \mathbb{A}) \quad \Leftrightarrow \quad \mathcal{P}^{\mathrm{A}}(\mathbf{x}) = \mathcal{N}\left(\mathbf{x}^A, \mathbb{A}\right) . \tag{6.5}$$

6.2.1 Three dimensional variational (3D-Var) data assimilation

In 3D-Var data assimilation the observations (6.3) are sequentially accounted for. The model (6.6) advances the state to time t_i

$$\mathbf{x}_i^B = \mathcal{M}_{t_{i-1} \to t_i} \mathbf{x}_{i-1}^A . \tag{6.6}$$

The model forecast at t_i provides the background state. An analysis state is obtained by considering the observations $\mathbf{y}_i$, and by computing the maximum likelihood estimate

$$\mathbf{x}_i^A = \arg\min \mathcal{J}(\mathbf{x}_i) \tag{6.7}$$

where

$$\mathcal{J}(\mathbf{x}_i) = \frac{1}{2}\left(\mathbf{x}_i - \mathbf{x}_i^B\right)^T \mathbb{B}_i^{-1} \left(\mathbf{x}_i - \mathbf{x}_i^B\right) + \frac{1}{2} \left(\mathcal{H}(\mathbf{x}_i) - \mathbf{y}_i\right)^T \mathbb{R}^{-1} \left(\mathcal{H}(\mathbf{x}_i) - \mathbf{y}_i\right) \tag{6.8}$$

Typically the same covariance matrix is used for all steps, $\mathbb{B}_0 = \mathbb{B}_i \; \forall \, i$

The optimization problem (6.7) is solved numerically using a gradient-based technique. The gradient of (6.8) reads

$$\nabla_{\mathbf{x}_i} \mathcal{J}(\mathbf{x}_i) = \mathbb{B}_i^{-1} \left(\mathbf{x}_i - \mathbf{x}_i^B\right) + \left(\mathcal{H}'(\mathbf{x}_i)\right)^T \mathbb{R}_i^{-1} \left(\mathcal{H}(\mathbf{x}_i) - \mathbf{y}_i\right) \tag{6.9}$$

The 3D-Var gradient requires the transpose of the linearized observation operator $\mathcal{H}'$.

6.2.2 Four dimensional variational (4D-Var) data assimilation

In strongly-constrained 4D-Var data assimilation all observations (6.3) at all times $t_1, \cdots, t_N$ are simultaneously considered. The control parameters are the initial conditions $\mathbf{x}_0$; they uniquely determine the state of the system at all future times via the model equation (6.6). The background state is the prior value of the initial conditions $\mathbf{x}_0^{\mathrm{B}}$.

Given the background value of the initial state $\mathbf{x}_0^{\mathrm{B}}$, the covariance of the initial background errors $\mathbb{B}_0$, the observations $\mathbf{y}_i$ and the corresponding observation error covariances $\mathbf{R}_i$, $i = 1, \cdots, N$, the 4D-Var problem looks for the maximum likelihood estimate $\mathbf{x}_0^A$ of the true initial conditions by solving an optimization problem. The 4D-var cost function measures the departure of the initial conditions from the background (weighted by the inverse background covariance matrix) and the differences between the model predictions $\mathcal{H}(\mathbf{x}_i)$ and observations $\mathbf{y}_i$ (weighted by the inverse observation error covariances)

$$\mathcal{J}\left(\mathbf{x}_0\right) = \frac{1}{2}\left(\mathbf{x}_0 - \mathbf{x}_0^B\right)^T \mathbb{B}_0^{-1}\left(\mathbf{x}_0 - \mathbf{x}_0^B\right) + \frac{1}{2}\sum_{i=0}^{N}\left(\mathcal{H}(\mathbf{x}_i) - \mathbf{y}_i\right)^T \mathbb{R}^{-1}\left(\mathcal{H}(\mathbf{x}_i) - \mathbf{y}_i\right) \quad (6.10)$$

The 4D-Var analysis is computed as the initial condition which minimizes (6.10) subject to the model equation constraints (6.6)

$$\mathbf{x}_0^A = \arg\min \mathcal{J}\left(\mathbf{x}_0\right) \quad \text{subject to} \ \ (6.6). \quad (6.11)$$

The model (6.6) propagates the optimal initial condition (6.10) forward in time to provide the analysis at future times, $\mathbf{x}_i^A = \mathcal{M}_{t_0 \to t_i}\, \mathbf{x}_0^A$.

The optimization problem (6.11) is solved numerically using a gradient-based technique. The gradient of (6.10) reads

$$\nabla_{\mathbf{x}_0}\mathcal{J}\left(\mathbf{x}_0\right) = \mathbb{B}_0^{-1}\left(\mathbf{x}_0 - \mathbf{x}_0^B\right) + \sum_{i=0}^{N}\left(\frac{\partial \mathbf{x}_i}{\partial \mathbf{x}_0}\right)^T \left(\mathcal{H}'(\mathbf{x}_i)\right)^T \mathbb{R}_i^{-1}\left(\mathcal{H}(\mathbf{x}_i) - \mathbf{y}_i\right) \quad (6.12)$$

The 4D-Var gradient requires not only the linearized observation operator $\mathcal{H}'$, but also the transposed derivative of future states with respect to the initial conditions

$$M_{t_i \to t_0}^T = \left(\frac{\partial \mathbf{x}_i}{\partial \mathbf{x}_0}\right)^T \quad (6.13)$$

The 4D-Var gradient can be obtained effectively by forcing the adjoint model with observation increments, and running it backwards in time. The construction of the adjoint model requires considerable development effort. Moreover, running the adjoint model requires considerable CPU time (typically, a small multiple of the time needed to run the forward model).

6.3 Quasi 4D-Var

We consider approximations of the adjoint model in order to decrease the development effort needed to build a full adjoint model, and to reduce the computational time associated with the computation of the gradient (6.12) in numerical optimization. We call the resulting technique *quasi 4D-Var*.

Consider the following quasi-adjoint model, which approximates the model adjoint derivative matrix (6.13)

$$N^T_{t_i \to t_0} \approx M^T_{t_i \to t_0} = \left(\frac{\partial \mathbf{x}_i}{\partial \mathbf{x}_0} \right)^T . \tag{6.14}$$

Using (6.14) in (6.12) leads to the following approximation of the adjoint gradient:

$$g(\mathbf{x}_0) = \mathbb{B}_0^{-1} \left(\mathbf{x}_0 - \mathbf{x}_0^B \right) + \sum_{i=0}^{N} N^T_{t_i \to t_0} \left(\mathcal{H}'(\mathbf{x}_i) \right)^T \mathbb{R}_i^{-1} \left(\mathcal{H}(\mathbf{x}_i) - \mathbf{y}_i \right) \tag{6.15}$$

Assume that the quasi-adjoint approximation (6.14) is close enough to the exact adjoint in the sense that

$$\left\| N^T_{t_i \to t_0} - M^T_{t_i \to t_0} \right\| \leq C , \quad \forall\, i = 0, \cdots, N , \quad \forall \mathbf{x}_0 \,:\, \left\| \mathbf{x}_0 - \mathbf{x}_0^A \right\| \leq R . \tag{6.16}$$

Here C is a uniform bound for all observation times and for all initial conditions sufficiently close to the optimal solution. The approximation error in the quasi-adjoint gradient (6.15) is:

$$\begin{aligned} \Delta g(\mathbf{x}_0) &= g(\mathbf{x}_0) - \nabla_{\mathbf{x}_0} \mathcal{J}(\mathbf{x}_0) \\ \| \Delta g(\mathbf{x}_0) \| &\leq C \sum_{i=0}^{N} \left\| \left(\mathcal{H}'(\mathbf{x}_i) \right)^T \mathbb{R}_i^{-1} \left(\mathcal{H}(\mathbf{x}_i) - \mathbf{y}_i \right) \right\| . \end{aligned} \tag{6.17}$$

The error bound (6.20) is reduced when the magnitude of the model-observation residuals $\| \mathcal{H}(\mathbf{x}_i) - \mathbf{y}_i \|$ decreases. In particular, if the model is accurate and the noise in the observations is small, the residuals become small close to the analysis trajectory. Thus the approximation (6.20) becomes more accurate as the iterations converge.

6.3.1 Gradient calculation without model adjoint

The simplest approximation is to completely ignore model dynamics and set the quasi-adjoint to the identity matrix

$$N^T_{t_i \to t_0} = \mathbb{I}_{n \times n} \quad \forall\, i = 0, \cdots, N .$$

Assuming that the adjoint model is smooth in time, the quasi-adjoint error is bounded by a constant times the length of the simulation interval

$$\left\| N^T_{t_i \to t_0} - M^T_{t_i \to t_0} \right\| \leq C \, |t_N - t_0| \quad \forall\, i = 0, \cdots, N . \tag{6.18}$$

The quasi-gradient reads

$$g(\mathbf{x}_0) = \mathbb{B}_0^{-1} \left(\mathbf{x}_0 - \mathbf{x}_0^B \right) + \sum_{i=0}^{N} \left(\mathcal{H}'(\mathbf{x}_i) \right)^T \mathbb{R}_i^{-1} \left(\mathcal{H}(\mathbf{x}_i) - \mathbf{y}_i \right) \tag{6.19}$$

and has the following error bound

$$\|\Delta\, g\,(\mathbf{x}_0)\| \leq C\ |t_N - t_0|\ \sum_{i=0}^{N} \left\| \left(\mathcal{H}'(\mathbf{x}_i)\right)^T\ \mathbb{R}_i^{-1}\ \left(\mathcal{H}(\mathbf{x}_i) - \mathbf{y}_i\right) \right\| \ . \tag{6.20}$$

The quasi-gradient (6.19) is a direct extension of the 3D-Var gradient (6.15): the observations at $t_1, \cdots, t_N$ are all considered to have taken place at t_0.

The error bound (6.20) can be reduced by reducing the length of the assimilation window. For $N = 0$ the gradient is exact, and equal to the 3D-Var gradient (6.15). In order to reduce the error bound an extension of this approach can be obtained as follows. Assimilate several observations at once at a time chosen in the middle of the observation interval. Assuming the number of observation times N is even we have the following quasi-gradient:

$$g\,(\mathbf{x}_0) = \mathbb{B}_0^{-1}\ \left(\mathbf{x}_0 - \mathbf{x}_0^B\right) + \sum_{i=-N/2}^{N/2} \left(\mathcal{H}'(\mathbf{x}_i)\right)^T\ \mathbb{R}_i^{-1}\ \left(\mathcal{H}(\mathbf{x}_i) - \mathbf{y}_i\right) \tag{6.21}$$

6.3.2 Gradient calculation with a coarse resolution adjoint model

An alternative approach to reduce the gradient CPU time is to run the adjoint model with a coarser resolution. In this case the quasi-adjoint approximation error is

$$\left\| N_{t_i \to t_0}^T - M_{t_i \to t_0}^T \right\| \leq C\, \tau^p \quad \forall\ i = 0, \cdots, N\ ,$$

Here τ is a discretization or model parameter that controls the accuracy of the adjoint model. For example, the adjoint model can be run with larger timesteps or with coarser grids; the coarse adjoint solution is then interpolated onto the grid of the forward model.

6.3.3 Gradient calculation with simplified physics adjoint

In case of multi-physics systems the quasi adjoint can leave out some of he physical processes which have a smaller impact on the system output. For example, in chemical transport models, the quasi adjoint can include the transport but leave out the chemistry, particle processes, etc. This seems to be a reasonable approximation when the measurements of a certain chemical species (e.g., ozone columns) are used to adjust the initial concentration of the same species (ozone in our example). By leaving out the chemistry the data assimilation system ignores the correlations between multiple chemical species.

We assume that the error in quasi-adjoint approximation (6.16) is similar to (6.18) and increases linearly with time

$$\left\| N_{t_i \to t_0}^T - M_{t_i \to t_0}^T \right\| \leq C\ |t_N - t_0| \quad \forall\ i = 0, \cdots, N\ . \tag{6.22}$$

6.3.4 Impact of inexact gradients on the optimization process

The use of inexact gradients in optimization have been extensively studied in the literature. In this section we review several results which are relevant for quasi 4D-Var.

[Kelley and Sachs, 1999, Section 2.3.1] considers the behavior of errors in Newton's method when the Hessian and the gradient are inexact. Let $e^{(k)} = \mathbf{x}_0^{(k)} - \mathbf{x}_0^A$ be the error in the numerical solution at iteration k. Let $\Delta H(\mathbf{x}_0)$ be the error in the Hessian used by Newton method. Then,

$$\left\| e^{(k+1)} \right\| \leq C \left(\left\| e^{(k)} \right\|^2 + \left\| \Delta H \left(\mathbf{x}_0^{(k)} \right) \right\| \cdot \left\| e^{(k)} \right\| + \left\| \Delta g \left(\mathbf{x}_0^{(k)} \right) \right\| \right) .$$

This implies that the accuracy of the optimal solution cannot exceed the accuracy of the approximate gradient.

[Carter, 1991] has studied the convergence of trust region algorithms using inexact gradients. He established that if the gradient approximation error is such that

$$\left\| \nabla_{\mathbf{x}_0} \mathcal{J} \left(\mathbf{x}_0 \right) - g \left(\mathbf{x}_0 \right) \right\| \leq \zeta \cdot \left\| \nabla_{\mathbf{x}_0} \mathcal{J} \left(\mathbf{x}_0 \right) \right\|$$

for a constant ζ then the trust region algorithm converges strongly in the sense that for the iterates $\mathbf{x}_0^{(k)}$ we have that

$$\lim_{k \to \infty} \left\| \nabla_{\mathbf{x}_0} \mathcal{J} \left(\mathbf{x}_0^{(k)} \right) \right\| = 0 .$$

Carter indicates that a typical value for the constant is $\zeta \sim 0.8$. The condition on the approximation error implies that the approximate gradient goes to zero near the exact solution (when the exact gradient goes to zero). This is possible in an idealized (twin experiment) setting where the model-observations mismatch is zero for the optimal solution.

6.4 Numerical Experiments

In order to compare the computational cost and quality of analyses for Quasi 4D-Var systems, we conducted data assimilation experiments similar to as described in section 3.6 of Chapter 3. To summarize, we assimilated Tropospheric Emission Spectrometer (TES) satellite ozone profile retrievals into the GEOS-Chem v7 for two different assimilation windows and validated the generated analyses against direct ozone profile measurements from ozonesondes. The numerical optimization method used in all variational experiments is the limited memory bound-constrained BFGS. All experiments were carried out at $4^\circ \times 5^\circ$ resolution.

6.4.1 Computational Costs

Presented in Table 6.1 are the timing results for performing a 24 hour simulation using GEOS-Chem free model run using different chemistry solvers, and subotimal Kalman filter, 3D-Var, Q4D-Var, and 4D-Var data assimilation systems. Diagonal background error covariance matrices are used for all the experiments. 3D-Var is the least expensive of all variational data assimilation systems as it requires no model adjoint. Next is the Quasi 4D-Var with no model adjoint. A single iteration of this technique costs less than 3D-Var since no sequential corrections were involved, however, in order to generate an optimal stable analysis, we performed 12 iterations and overall it is about 10 times costlier than 3D-Var. The Quasi 4D-Var with advection adjoint is a bit expensive than no adjoint which was expected. Compared to 3D-Var it is about 13 times costlier. 4D-Var is the most expensive of all due to full model adjoint calculations and checkpointing of forward variables. Compared to 3D-Var, it requires computational time which is 51 times that of the 3D-Var for the same assimilation window length.

Table 6.1: Timing results for GEOS-Chem free model run, suboptimal KF, 3D-Var, Q4D-Var, and 4D-Var data assimilations for a 24 hour simulation starting at 00:00 GMT on July 1st, 2006.

Experiment Description	CPU Time	Scaled time
Free model run, SMVGEAR chemistry	2 min 50 sec	1.00
Free model run, KPP chemistry	3 min 18 sec	1.16
Suboptimal Kalman filter with diagonal $\mathbb{P}^f$	3 min 08 sec	1.11
3D-Var with diagonal $\mathbb{B}$	3 min 57 sec	1.39
Quasi 4D-Var, no adjoint (per iteration)	3 min 11 sec	1.12
Quasi 4D-Var, advection adjoint (per iteration)	4 min 23 sec	1.55
4D-Var with diagonal $\mathbb{B}$ (per iteration)	16 min 51 sec	5.95

We next illustrate in Figure 6.1, the disk memory usage for Quasi 4D-Var methods as compared to 4D-Var. The plots reflect that Quasi 4D-Var systems have a lot less memory requirements as compared to the 4D-Var. All the three assimilation systems use 60 MB constant over any assimilation window length to save parameters related to l-BFGS optimization. Q4D-Var with no adjoint requires no additional disk memories. Q4D-Var with advection adjoint although does not depend on any forward variables, state vectors from the forward mode at the end of every observation window are required for forcing calculations in the backward mode. Precisely it requires an additional 62 MB every observation window compared to the no adjoint case. In the plot, we have broken it down to a requirement of 16 MB per hour. 4D-Var is the most costly of all as it writes out 570 MB worth of data every for each simulation hour. Hence, Quasi 4D-Var with advection adjoint is 36 times efficient that 4D-Var in terms of memory.

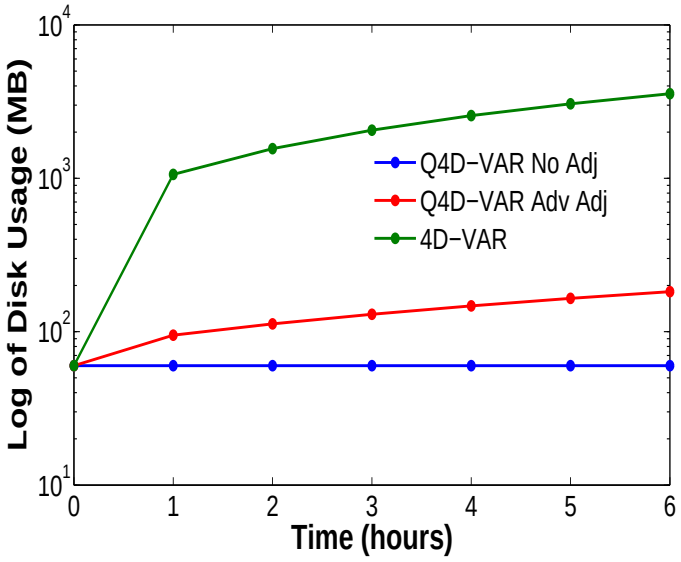


Figure 6.1: Plot of computational time against log of disk memory usage for 4D-VAR and Q4D-VAR assimilations over a period of 6 hours from 00:00 GMT to 00:06 GMT August 1st, 2006.

6.4.2 Global ozone estimates through Quasi 4D-VAR

We first consider an assimilation window length of 5 days. Presented in Figure 6.2 is a comparison of analysis profiles obtained from different assimilation systems, and free GEOS-Chem model run against ozonesonde measurement data. The plots provide an assessment of the quality of tropospheric ozone as estimated by suboptimal Kalman filter, 3D-Var, Q4D-Var and 4D-Var assimilation systems. The left panel is the plot of pressure level against ozone estimates averaged over all ozonesonde locations for all ozonesonde measurements available in the 5-day assimilation window. The center panel provides the relative difference of mean ozone estimates from assimilation systems and free model run against mean ozonesonde measurements. The rightmost panel provides the standard deviation of absolute values of the differences between ozone predictions and ozonesonde measurements.

It is quite interesting to see how using the same gradients from 3D-Var, the overall analysis is improved. Both the Quasi 4D-Var systems improve the tropospheric ozone estimates. Q4D-Var with no adjoints causes a decrease in the relative difference between the mean ozone analysis field and the ozonesonde measurements to less than 6% as compared to 5-20% in cases of suboptimal KF and 3D-Var. It overstimates mid tropospheric ozone by a slight margin as compared to the 4D-Var, however, it greatly improves ozone in the upper troposphere. Quasi 4D-Var with advection adjoint surprisingly performed better than 4D-Var throughout the tropospher. It brought down the analysis-ozonesonde

relative difference by 2-3% further as compared to 4D-Var. It would be worth finding out if this phenomenon persists if a trace gas is optimized using observations of its precursors. In our case, we are assimilating ozone using ozone profile retrievals. Another interesting idea would be to use non-diganal background error covariance matrices for better distribution of the forcing information.

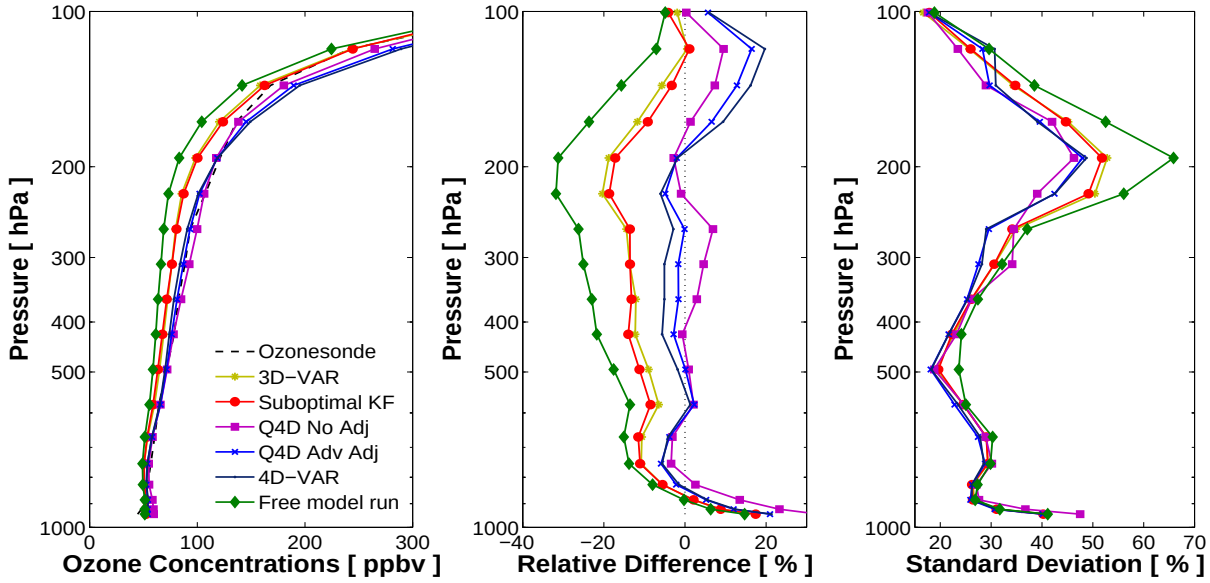
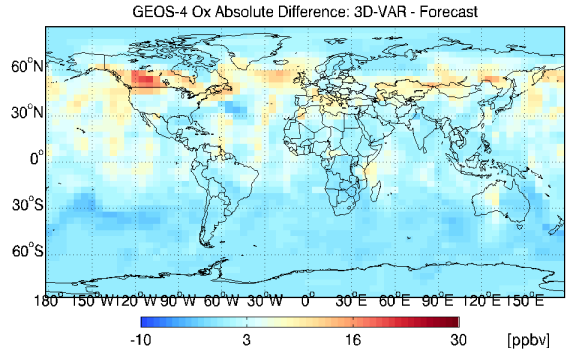


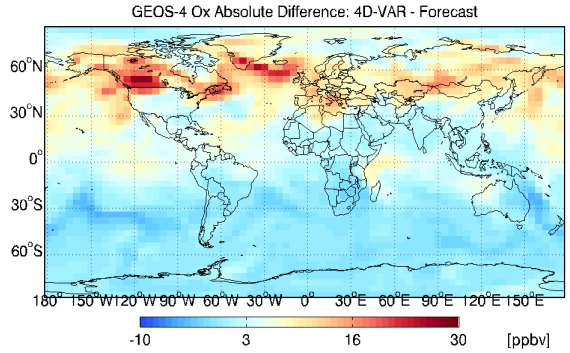
Figure 6.2: The results shown are for a 5-day simulation from 00:00 GMT August 1st, 2006 to 00:00 GMT August 6th, 2006. Left panel: mean ozone concentrations at ozonesonde locations for 3D-Var, 4D-Var and Q4D-Var analyses and free model trajectories. Center panel: relative mean errors of predicted ozone concentrations with respect to ozonesonde measurements. Right panel: standard deviation of absolute values of errors with respect to ozonesonde measurements.

To view the global distribution of the corrections in tropospheric ozone brought in by variational data assimilation systems, we provide in Figure 6.3 panels (a)-(d), the difference between ozone analyses from different assimilation systems and forecast state as obtained by running GEOS-Chem model. The correction profile for Q4D-Var with advection adjoint is quite similar to 4D-Var, while no adjoint case has few localized errors over North Atlantic and north Australian region. For a more precise comparison, we provide in panels (e)-(f), the difference of the analysis profiles obtained from Q4D-Var and 4D-Var systems.

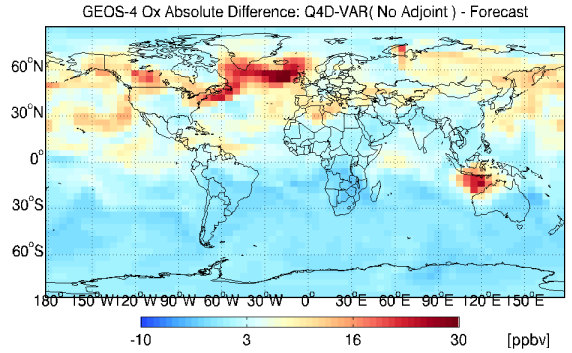
Since the difference between the Quasi 4D-Var and 4D-Var implementations are in their gradient calculations, we provide in Figure 6.4, the difference in their gradients accumulated over the 5-day assimilation window. This would provide an insight into how difference in gradient calculations affect the quality of an analysis. The plots reflect that



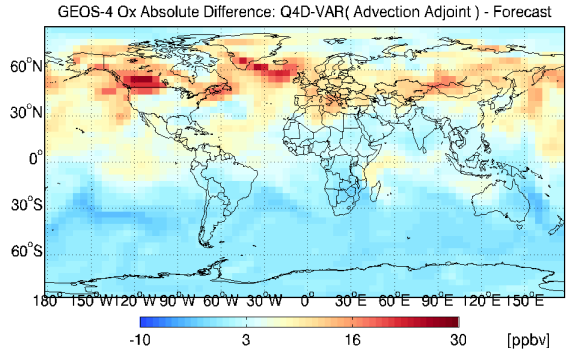
(a) Absolute difference between the 3D-Var analysis and the free model run



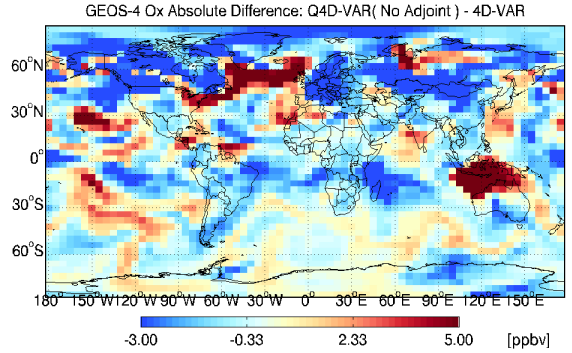
(b) Absolute difference between the 4D-Var analysis and the free model run



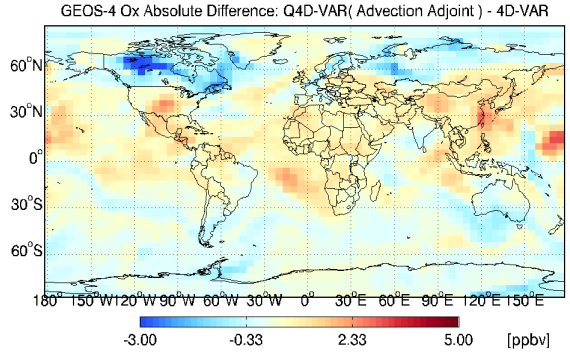
(c) Absolute difference between the Quasi 4D-Var analysis using no model adjoint and the free model run



(d) Absolute difference between the Quasi 4D-Var analysis using advection adjoint and the free model run

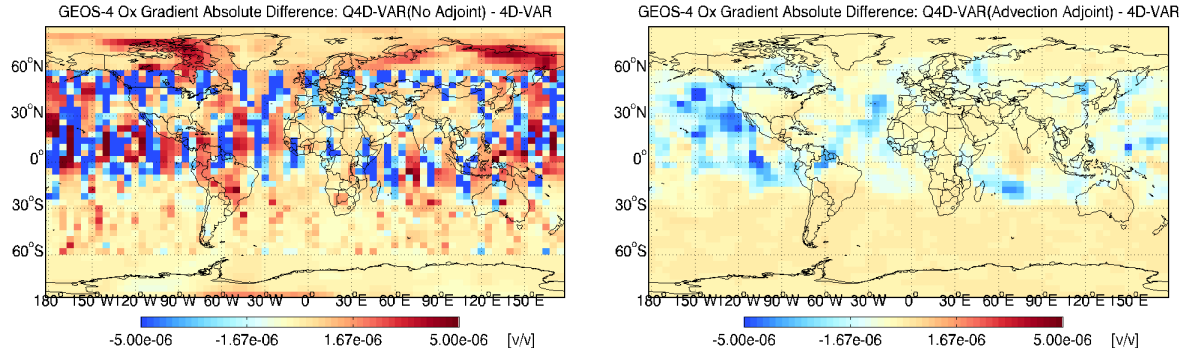


(e) Absolute difference between the Quasi 4D-Var analysis using no model adjoint and 4D-Var analysis



(f) Relative difference between the Quasi 4D-Var analysis using advection adjoint and 4D-Var analysis

Figure 6.3: Differences in global ozone concentrations at 00:00 GMT on August 06, 2006 (end of assimilation window) averaged over the first 10 GEOS-Chem vertical levels. Panels (a)-(d): differences between the 3D-Var, 4D-Var, Q4D-Var with no model adjoint and with advection adjoint analysis fields against the model forecast (solution without data assimilation). Panels (e)-(f): absolute differences between Quasi 4D-Var analyses using no adjiont and using advection adjoint against 4D-Var analysis.



(a) Absolute difference between Quasi 4D-Var gradient using no model adjoint and 4D-Var gradient

(b) Absolute difference between Quasi 4D-Var gradient using advection adjoint and 4D-Var gradient

Figure 6.4: Differences in global ozone gradient fields at 00:00 GMT on August 01, 2006 (start of assimilation window) after the first iteration of Q4D-Var and 4D-Var assimilations with a 5-day assimilation window averaged over the first 10 GEOS-Chem vertical levels.

for shorter assimilation window lengths, the gradients calculated using only advection adjoint approximate well the full model adjoint, while gradients without any adjoints differ mostly along the satellite trajectories. It is also highly interesting to compare these plots with panels (e)-(f) of Figure 6.4. In case of Q4D-Var with advection adjoints, the difference in the analysis profiles as compared to 4D-Var have almost exactly the same profile as the difference in the gradients. This result could be highly useful in designing future Quasi 4D-Var systems that could resemble 4D-Var without full model adjoint construction.

We next consider simulations with assimilation window length of 2 weeks. A longer assimilation window provides an insight into how ozone estimates obtained through different assimilation systems vary with time and if the corrections maintain structures similar to 5-day case.

Similar to Figure 6.2, we present in Figure 6.5, a comparison of analysis profiles obtained from different assimilation systems against ozonesonde measurement data. The plots reflect that the accuracy of analysis obtained from Q4D-Var with no model adjoint is compromised significantly, while analysis obtained from Q4D-Var with advection adjoint is still comparable to 4D-Var. However, if we look at the difference plots of ozone estimates as presented in Figure 6.6, the overcorrection as seen in Figure 6.5 for Quasi 4D-Var with no adjoints, is not visible here.

Also provided in Figure 6.7, are the global difference plots of the gradients obtained through Q4D-Var and 4D-Var assimilation systems. The difference between Q4D-Var and 4D-Var gradients is still smoothly structured, however, the resemblance between

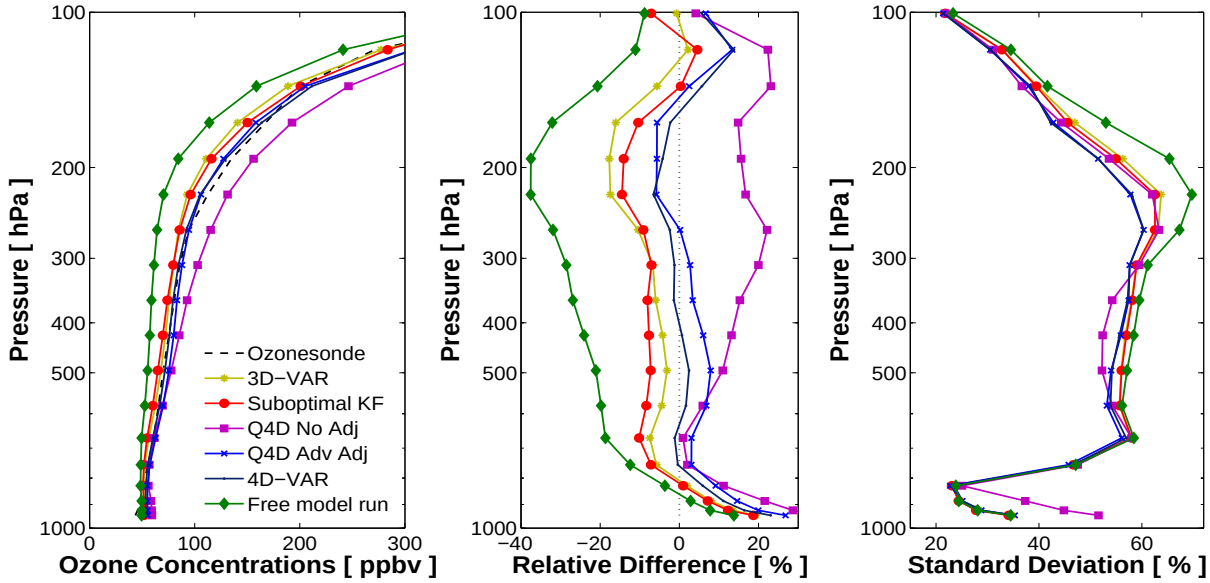
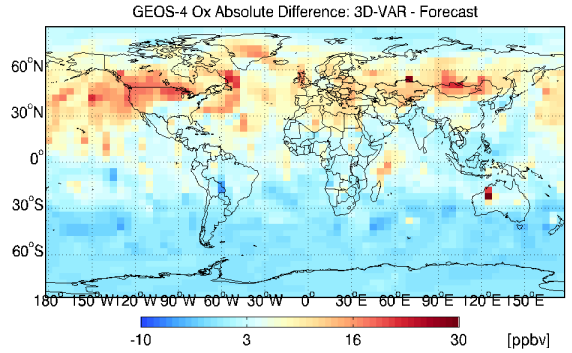


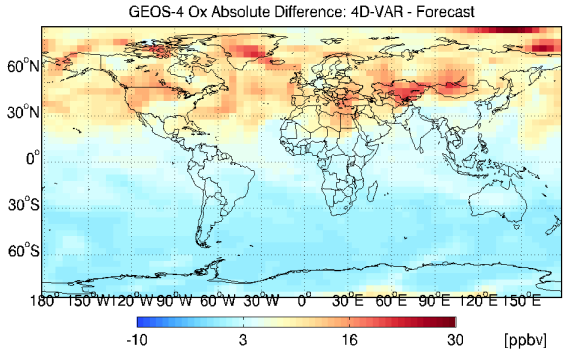
Figure 6.5: The results shown are for a 2-week simulation from 00:00 GMT August 1st, 2006 to 00:00 GMT August 15th, 2006. Left panel: mean ozone concentrations at ozonesonde locations for 3D-Var, 4D-Var and Q4D-Var analyses and free model trajectories. Center panel: relative mean errors of predicted ozone concentrations with respect to ozonesonde measurements. Right panel: standard deviation of absolute values of errors with respect to ozonesonde measurements.

gradient difference and analysis difference is lost.

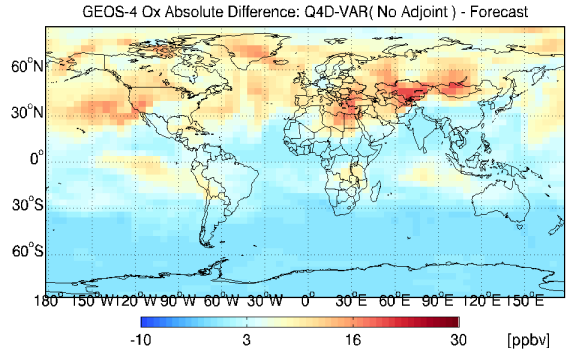
We next illustrate through Figure 6.8, the decrease in L1-norm of the relative differences between mean ozonesonde measurements and mean estimates provided by GEOS-Chem free model run and various data assimilation systems studied in this dissertation. It reflects clearly that development of Quasi 4D-Var algorithms is a worthy effort and provides a suite of assimilation systems that could improve the accuracy of model estimates with huge reduction in the effort involved in building full model adjoints and the computational resources.



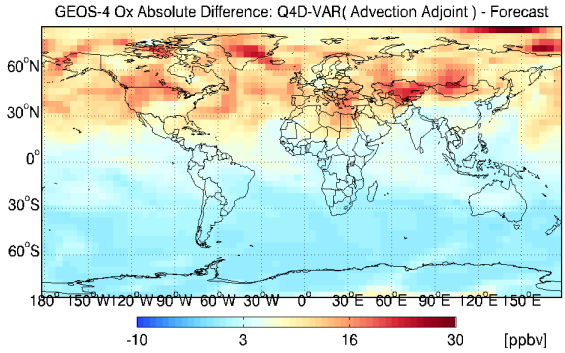
(a) Absolute difference between the 3D-Var analysis and the free model run



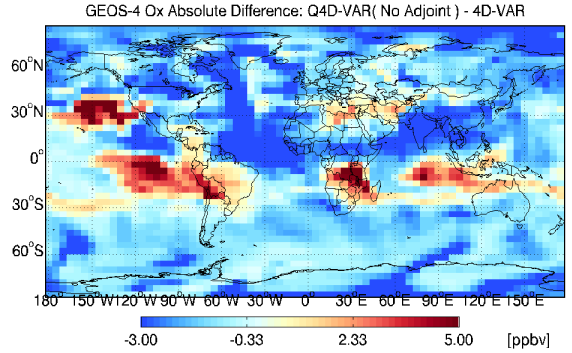
(b) Absolute difference between the 4D-Var analysis and the free model run



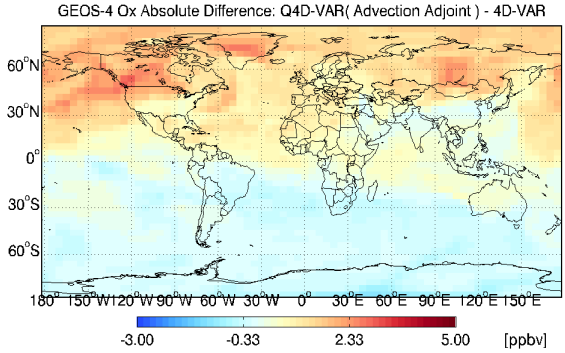
(c) Absolute difference between the Quasi 4D-Var analysis using no model adjoint and the free model run



(d) Absolute difference between the Quasi 4D-Var analysis using advection adjoint and the free model run

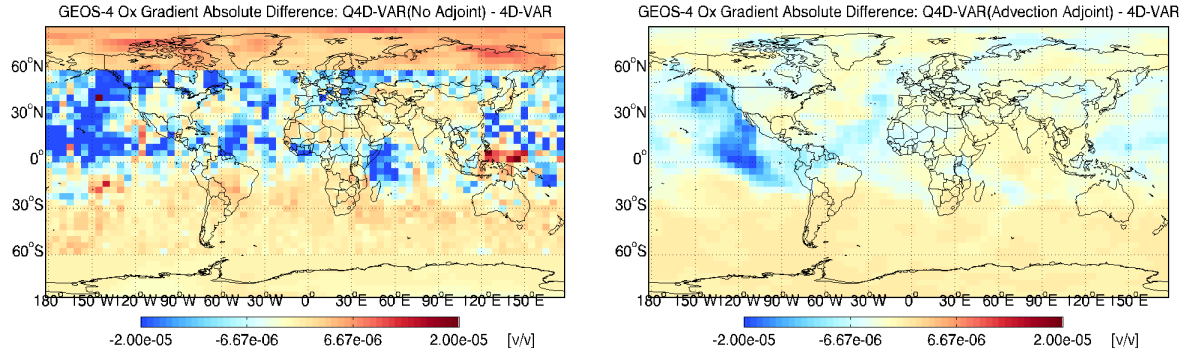


(e) Absolute difference between the Quasi 4D-Var analysis using no model adjoint and 4D-Var analysis



(f) Relative difference between the Quasi 4D-Var analysis using advection adjoint and 4D-Var analysis

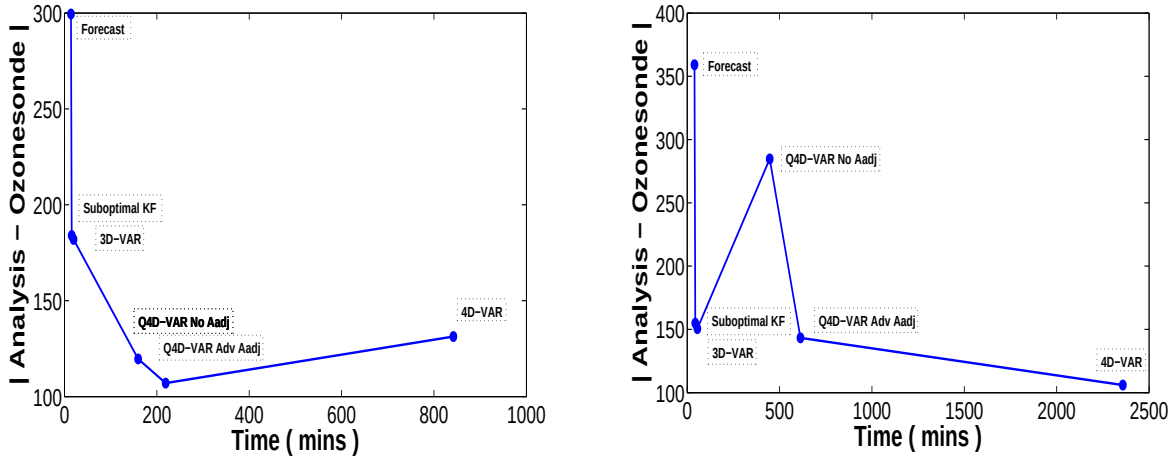
Figure 6.6: Differences in global ozone concentrations at 00:00 GMT on August 15, 2006 (end of assimilation window) averaged over the first 10 GEOS-Chem vertical levels. Panels (a)-(d): differences between the 3D-Var, 4D-Var, Q4D-Var with no model adjoint and with advection adjoint analysis fields against the model forecast (solution without data assimilation). Panels (e)-(f): absolute differences between Quasi 4D-Var analyses using no adjiont and using advection adjoint against 4D-Var analysis.



(a) Absolute difference between Quasi 4D-Var gradient using no model adjoint and 4D-Var gradient

(b) Absolute difference between Quasi 4D-Var gradient using advection adjoint and 4D-Var gradient

Figure 6.7: Differences in global ozone gradient fields at 00:00 GMT on August 01, 2006 (start of assimilation window) after the first iteration of Q4D-Var and 4D-Var assimilations with a 2-week assimilation window averaged over the first 10 GEOS-Chem vertical levels.



(a) Computational time versus decrease in error norm for a 5-day assimilation window

(b) Computational time versus decrease in error norm for a 2-week assimilation window

Figure 6.8: Plot of computational time required to perform free model run, suboptimal Kalman Filter, 3D-Var, Q4D-Var with no model adjoint and with advection adjoint, and 4D-Var over a 5-day (top panel) and 2-week (bottom panel) assimilation windows versus mean difference of model predicted ozone and ozonesonde measurements.

6.5 Conclusions

In this chapter we propose the *quasi 4D-Var* (Q4D-Var) data assimilation, a technique which uses approximations of the adjoint model in order to decrease the development

effort, and to reduce the computational time associated with the computation of the 4D-Var gradient. In the context of chemical transport modeling two approximations are discussed: replace the adjoint solution operator by the identity operator, and replace it by the continuous advection adjoint.

Numerical experiments are carried out with a global chemical data assimilation problem using satellite observations and the GEOS-Chem chemical transport model. These experiments reveal that Q4D-Var is only slightly more expensive than 3D-Var. At the same time the accuracy of the resulting analyses is similar to that of 4D-Var for relatively short assimilation windows.

Future work is needed to better understand how the approximations of adjoint gradient impact the convergence of the numerical optimization scheme, and how they impact the accuracy of the resulting analysis. As the conclusions are likely to be problem dependent, one needs to determine the type of adjoint approximations that are sufficient for particular data assimilation problem settings.

Chapter 7

Conclusions and Future Work

This dissertation work has focused on advancing the field of chemical data assimilation. Cost efficient computational methodologies have been developed for modeling the background covariances, for estimating the information content of observations, and for approximating gradients, among other. We have implemented variational data assimilation systems, including full model adjoints, for two of the most widely used chemical transport models; this software is currently being used by the community. We have used these tools to demonstrate for the first time that satellite observations can be used directly to improve estimates of the global tropospheric ozone distribution.

7.1 Conclusions

Adjoint models compute the sensitivity of model outputs with respect to all the model input parameters. Constructing an adjoint of a large scale model is a difficult and error prone task. We have developed adjoints of two widely used chemical transport models: Harvard's global GEOS-Chem v7 and EPA's regional CMAQ v4.5. We employ a variety of strategies for the construction of adjoint models: symbolic preprocessing, derivation of adjoint differential equations and their numerical solution, and automatic differentiation. Validation results show that the adjoints of each science process are within 2% of their finite difference counterparts, with the exception of advection for which the continuous adjoint approach leads to a difference of 7%. Discrete adjoints are known to lead to oscillations at discontinuities and boundaries; this is exemplified with the monotonicity constrained peicewise parabolic method in CMAQ. Using the developed adjoints we have conducted a sensitivity study involving summertime global tropospheric ozone seeking to quantify and locate its sources and sinks. The results show evidence of intercontinental exchanges of pollutants with in a period of three days. Sensitivities with respect to trace gas precursors (NOx, CO, SO2, RCHO) reflect the amount of change in

each of these tracers at 00:00 GMT on July 1, 2006 that cause a change of 1 [v/v] in ozone after three days over a possible TES orbit trajectory. The emission and dry deposition results reflect the changes in the rates at which NO, CO, HNO3 and O3 should enter or leave the system to cause a similar change in ozone. These results are helpful for policy decisions, for designing observing systems, and for identifying the most important emission locations.

We have successfully implemented variational data assimilation (3D-Var and 4D-Var) framework for GEOS-Chem. This system was used to perform assimilation of vertical profile retrievals from TES. We have provided the first direct comparison between 3D-Var, 4D-Var and suboptimal Kalman filter (KF) for chemical data assimilation problems. The comparison accounts for both accuracy and computational cost, and uses an independent data set (ozonesonde measurements) to check the analyses. Sequential assimilation methods – 3D-Var and suboptimal KF – perform similarly over small assimilation windows. For longer assimilation windows suboptimal KF tends to underestimate the ozone concentrations in the lower and mid troposphere, but performs better than 3D-Var in the mid and upper troposphere. 4D-Var provides the best estimates for up to 180 hPa, and performs well for longer simulation time scales. The relative difference between mean ozone estimates and ozonesonde data is reduced by up to 75% for sequential methods (3D-Var, suboptimal KF) and by up to 90% for 4D-Var over a two week simulation. When diagonal background error covariance matrices are used the suboptimal KF is computationally the least expensive, followed by 3D-Var. The memory and computational costs for 4D-Var are the highest as this method requires checkpointing forward variables, and running the adjoint model each iteration. The data assimilation framework will enable GEOS-Chem users to perform data assimilation and obtain better estimates of trace gas concentrations, of emission levels, and of various model parameters such as deposition rates.

Background covariance matrices play a significant role in determining the quality of the assimilation results. We have developed a novel approach for constructing full rank non-diagonal covariance matrices. This structure determines the spatial spread of information from a local observation to neighboring grids.The construction is based on the exponential decay of error correlations. The full multidimensional covariance matrix is never constructed explicitly, instead it is represented as tensor products of one-dimensional correlation matrices. Highly efficient linear algebra operations are possible. The non-diagonal background error covariance matrices have been tested with various correlation lengths within the variational data assimilation system. The computational overhead added when diagonal matrices are replaced with the full rank matrices is minimal. The use of non-diagonal covariances with both the 3D-Var and 4D-Var systems have lead to analyses that are closer to the ozonesonde measurements. In addition, the correction structures for both the assimilation systems were better distributed when compared to the diagonal case.

The reduction in the analysis uncertainty provides a way to quantify the information con-

tent of observations in data assimilation. We have developed an efficient ensemble-based methodology to estimate the value of observations in terms of information theoretic metrics such as the Fisher information matrix, the Shannon information, the relative entropy, the signal information, and the degrees of freedom for signal. The approach was applied to a global ozone data assimilation problem using TES satellite observations and the GEOS-Chem chemical transport model. The (vertical) level-wise information gain from all the observations indicate that the atmospheric region between 200 hPa - 400 hPa benefitted the most from assimilating the satellite data. The observations with the highest signal information content are located roughly along the latitudes $30^\circ S$ and $60^\circ N$. The assimilation results using the top 27% data points (with the highest signal information) are similar to the results obtained using all the data points. The approach is extended to asses the potential benefit of ground level observations, and to select those ground level locations that are associated with the highest information content. The observations with the highest estimated DFS information are located throughout the globe. Assimilation experiments with subsets of data selected based on the degrees of freedom for signal are more difficult to interpret.

We have proposed the *quasi 4D-Var* (Q4D-Var) approach to data assimilation, based on using cost-effective approximations of the adjoint gradient. In the context of chemistry transport models we have used approximations based on the identity operator and on the continuous advection adjoint only. The results for the assimilation of TES data in GEOS-Chem indicate that Q4D-Var is only slightly more expensive than the suboptimal KF and 3D-Var but provides more accurate analyses.

7.2 Future work

Many interesting applications could be tackled using the individual process adjoints developed here. One such application is the quantification of ozone accumulating in the lower and mid troposphere due to troposphere-stratosphere flux exchanges. So far, researchers have studied tropospheric ozone resulting from human and natural processes located in the lower to mid troposphere; the abundance of ozone in the lower stratosphere has not been studied extensively.

A very interesting direction is to couple the adjoints of global GEOS-Chem model with the adjoint of regional CMAQ model. Note that the forward (regular) models are regularly coupled for providing both global coverage, and higher accuracy in the regions of interest. The coupling of the adjoint models developed here would enable the study of long range transport of pollutants, for example the study of the impact Asian emissions have on US air quality.

The developed data assimilation systems currently assimilate only TES profile retrievals. However, the framework could easily be extended to use other instruments by simply

adding their observation operators. In addition, the framework could be extended to carry out assimilations with respect to ground level emissions and other model parameters.

The implementation of the full rank background covariance matrices could easily be extended to include correlations in the vertical direction. Moreover, the formulation allows the user to tune the correlation lengths based on the life time of certain species; this varies significantly with the vertical levels. Another extension is to use non-Gaussian functions to model the spatial decay of correlations. An interesting research topic which has not been pursued yet due to unavailability of full rank covariance matrices is defining correlations among emission sources. To date, it is considered that all emission sources are completely independent.

The proposed information content estimation techniques could be utilized to compare different instruments for optimal design of future sensor systems, assess the effectiveness of assimilation systems, and eliminate erroneous and redundant observations. A direct extension to our work is to quantify the impact that each of these issues (non-linearity, non-normality, approximate posterior distributions, and small samples) has on the accuracy of the information content estimates.

Quasi 4D-Var assimilation systems could be applied to other areas where a full adjoint model is not available; in this case only sequential approaches of data assimilation can be used to date. The success of quasi 4D-Var depends on how well the adjoint gradient can be approximated by easy to compute vectors. A better theoretical framework needs to be developed in order to understand the convergence conditions for quasi-4D-Var converges. Such a framework could be developed starting from previous work on optimization with inexact gradients.

Bibliography

Abramov, R. V. and Majda, A. J., Quantifying uncertainty for non-Gaussian ensemble in complex systems. *SIAM Journal on Scientific Computing*, 2004; **26(2)**, 411-447.

Allen, D. J., Rood, R. B., Thompson, A. M., and Hudson, R. D.: Three-dimensional radon 222 calculations using assimilated meteorological data and a convective mixing algorithm, J. Geophys. Res., 101 (D3), 6871Ű6881, doi:10.1029/95JD03408, 1996.

Anderson, B. D. O., and Moore, J. B.: Optimal Filtering, Information and System Science, Prentice-Hall, ISBN 0-13-638122-7, 1979.

Bartlett, M. S., An introduction to stochastic processes, with special reference to methods and applications. *Cambridge University Press*, 1962.

Beer, R., Glavich, T. A., and Rider, D. M.: Tropospheric emission spectrometer for the Earth Observing System's Aura satellite, Appl. Opt., 40(15), 2356-2367, 2001.

Bei, N., de Foy, B., Lei, W., Zavala, M., and Molina, L. T.: Using 3DVAR data assimilation system to improve ozone simulations in the Mexico City basin, Atmos. Chem. Phys., 8, 7353-7366, doi:10.5194/acp-8-7353-2008, 2008.

Benkovitz, C. M., Scholtz, M. T., Pacyna, J., Tarrasoťn, L., Dignon, J., Voldner, E. C., Spiro, P. A., Logan, J. A., and Graedel, T. E.: Global gridded inventories of anthropogenic emissions of sulfur and nitrogen, J. Geophys. Res., 101(D22), 29,239Ű29,253, 1996.

Bernardo, J. M. and Smith, A. F. M., Bayesian theory. *Wiley, Chichester, UK*, 1994.

Bey, I., Jacob, D. J. , Yantosca, R. M., Logan, J. A., Field, B., Fiore, A. M., Li, Q., Liu, H., Mickley, L. J. and Schultz, M.: Global modeling of tropospheric chemistry with assimilated meteorology: Model description and evaluation, J. Geophys. Res., 106, 23, 073-23,096, 2001.

Binkowski, F. S. and Roselle, S. J.: Models-3 community multiscale air quality (CMAQ) model aerosol component Ű 1. Model description, J. Geophys. Res., 108, 4183, doi:10.1029/2001JD001409, 2003.

Blum, J., Le Dimet, FX., Navon, I. M.: Data Assimilation for Geophysical Fluids, Chapter in Computational Methods for the Atmosphere and the Oceans, Volume 14, Elsevier Science Ltd, New York, ISBN-13: 978-0-444-51893-4, 2009.

Boutahar, J., Lacour, S., Mallet, V., Quélo, D., Roustan, Y., and Sportisse, B.: Development and validation of a fully modular platform for numerical modelling of air pollution: POLAIR, International Journal of Environment and Pollution, 22(1/2):17-28, 2004.

Bowman, K. W., Worden, J., Steck, T., Worden, H. M., Clough, S. and Rodgers, C.: Capturing time and vertical variability of tropospheric ozone: A study using TES nadir retrievals, J. Geophys. Res.,107, (D23), 2007.

Bowman, K. W., Rodgers, C. D., et al: Tropospheric Emission Spectrometer: Retrieval method and error analysis, IEEE Transactions on Geoscience and Remote Sensing, vol. 44, no. 5, May 2006.

Bowman, K. W., Jones, D. B. A., Logan, J. A., Worden, H., Boersma, F., Kulawik, S., Osterman, G., Worden, J. and Chang, R.: Impact of surface emissions to the zonal variability of tropical ozone and carbon monoxide for November 2004, Atmos. Chem. Phys. Disc., 8, 1505–1548, 2008.

Byun, D. W., Ching, J. K. S.: Science Algorithms of the EPA Models-3 Community Multiscale Air Quality (CMAQ) Modeling System, U.S. EPA/600/R-99/030, U.S. Environmental Protection Agency: Research Triangle Park, NC, 1999.

Byun, D. W., Schere, K. L.: Review of the governing equations, computational algorithms, and other components of the Models-3 Community Multiscale Air Quality (CMAQ) modeling system, Appl. Mech. Rev., 59, 51-77, 2006.

Byrd, R. H., Lu, P., Nocedal, J., and Zhu, C.: A limited memory algorithm for bound constrained optimization, Scientific Computing, 16, 1190Ű1208, 1995.

Cacuci, D. G.: Sensitivity theory for nonlinear systems-I: Nonlinear functional analysis approach, J. Math. Phys., 22, 2794-2802, 1981.

Cacuci, D. G.: Sensitivity theory for nonlinear systems-II: Extensions to additional classes of responses, J. Math. Phys., 22, 2803-2812, 1981.

Cacuci, D. G., Schlesinger, M. E.: On the application of the adjoint method of sensitivity analysis to problems in the atmospheric sciences, Atmosfera, 7, 47-59, 1994.

Cacuci, D. G., Ionescu-Bujor, M., Navon, I. M.: Sensitivity and Uncertainty Analysis, Chapman&Hall/CRC Press: Boca Raton, FL, 2003.

Cacuci, D. G., Ionescu-Bujor, M.: Deterministic local sensitivity analysis of augmented systemss I: Theory, Nucl. Sci. Eng., 151, 55-66, 2005.

Cardinali, C., Pezzulli, S., Andersson, E. Influence-matrix diagnostic of data assimilation system. *Quarterly Journal of the Royal Meteorological Societ*; **130**, 2767-2786, 2004.

Carmichael, G. R., Sandu, A., Potra F. A., and Damian-Iordache, V.: Sensitivity analysis for atmospheric chemistry models via automatic differentiation, Atmos. Environ., 31 (3), 475Ű489, 1997.

Carmichael, G. R., Sandu, A., Chai, T., Daescu, D., Constantinescu, E. M. and Tang, Y.: Predicting air quality: Improvements through advanced methods to integrate models and measurements, Journal of Computational Physics, Vol. 227, Issue 7, p. 3540–3571, 2008.

Carter, R. G.: On the global convergence of trust region algorithms using inexact gradient information, SIAM Journal on Numerical Analysis, Volume 28(1), 251-265, ISSN:0036-1429, 1991.

Chai, T., Carmichael, G. R., Sandu, A., Tang, Y. and Daescu, D. N.: Chemical data assimilation of transport and chemical evolution over the Pacific (TRACE-P) aircraft measurements, Journal of Geophysical Research, 111, D02301, doi:10.1029/2005JD005883, 2006.

Chai, T., Carmichael, G. R., Tang, Y., Sandu, A., Hardesty, M., Pilewskie, P., Whitlow, S., Browell, E. V., Avery, M. A., Thouret, V., Nedelec, P., Merrill, J. T. and Thomson, A. M.: Four dimensional data assimilation experiments with ICARTT (International Consortium for Atmospheric Transport and Transformation) ozone measurements, Journal of Geophysical Research, Vol. 112, D12S15, doi:10.1029/2006JD007763, 2007.

Chai, T., Carmichael, G. R., Tang, Y. and Sandu, A.: Regional NO_X emission inversion through a four-dimensional variational approach using SCIAMACHY tropospheric NO_2 column observations, Atmospheric Environment, doi:10.1016/j.atmosenv.2009.06.052, in print, 2009.

Cheng, H., Jardak, M., Alexe, M. and Sandu, A., A hybrid approach to estimating error covariances in variational data assimilation. *Tellus A.* **Vol.** 62, Number 3, May 2010 , pp. 288-297(10).

Clark, H. L., Cathala, M. -L., Teysse'dre, H., Cammas, J. -P. and Peuch.,V. -H.: Cross-tropopause fluxes of ozone using assimilation of MOZAIC observations in a global CTM, Tellus, Ser. A and Ser. B, 59B, 39-49, 2006.

Cohan, D. S., Hu, Y., Hakami, A., Odman, M. T., Russell, A. G.: Implementation of a direct sensitivity method into CMAQ, Models-3 Annual Conference, Chapel Hill, NC, 2002.

Cohen, Y.: Dry Deposition, course notes for Multimedia Environmental Assessment (C118/218), Chapter 12, Univ. of Calif., Los Angeles, 1998.

Cohn, S., Da Silva, A., Guo, J., Sienkiewicz, M., and Lamich, D.: Assessing the Effects of Data Selection with DAO's Physical-space Statistical Analysis System, Monthly Weather Review, 126, 2913-2926, 1998.

Colella, P. and Woodward, P. R.: The Piecewise Parabolic Method (PPM) for Gas-Dynamical Simulations, J. Comp. Phys., 54, 174Ű201, 1984.

Constantinescu, E. M., Chai, T., Sandu, A. and Carmichael, G. R.: Autoregressive models of background errors for chemical data assimilation, Journal of Geophysical Research, Vol. 112, D12309, doi:10.1029/2006JD008103, 2007.

Constantinescu, E. M., Sandu, A., Chai, T. and Carmichael, G. R.: Investigation of ensemble-based chemical data assimilation in an idealized setting, Atmospheric Environment, Vol. 41, Issue 1, p. 18–36, 2007.

Constantinescu, E. M., Sandu, A., Chai, T. and Carmichael, G. R.: Ensemble-based chemical data assimilation. I: general approach, Quarterly Journal of the Royal Meteorological Society, Volume 133, Issue 626, p. 1229–1243, Online ISSN: 1477-870X, Print ISSN: 0035-9009, July 2007 Part A.

Constantinescu, E. M., Sandu, A., Chai, T. and Carmichael, G. R.: Ensemble-based chemical data assimilation. II: covariance localization, Quarterly Journal of the Royal Meteorological Society, Volume 133, Issue 626, p. 1245–1256, Online ISSN: 1477-870X, Print ISSN: 0035-9009, July 2007 Part A.

Cooper, O. R., et al.: Large upper tropospheric ozone enhancements above midlatitude North America during summer: In situ evidence from the IONS and MOZAIC ozone measurement network, J. Geophys. Res., 111, D24S05, doi:10.1029/2006JD007306, 2006.

Cooper, O. R., et al.: Evidence for a recurring eastern North American upper tropospheric ozone maximum during summer, J. Geophys. Res., 112, D23304, doi:10.1029/2007JD008710, 2007.

Courtier, P., Talagrand, O.: Variational assimilation of meteorological observations with the adjoint vorticity equation. 2. Numerical results, Q. J. R. Meteorol. Soc., 113, 1329-1347, 1987.

Courtier, P., Andersson, E., Heckley, W., Pailleux, J., Vasiljevic, D., Hamrud, M., Hollingsworth, A., Rabier, F. and Fisher, M.: The ECMWF implementation of three-dimensional variational assimilation (3D-Var) I: Formulation, Quarterly Journal of the Royal Meteorological Society, 124(550):1783, 1998.

Daescu, D., Carmichael, G. R., Sandu, A.: Adjoint implementation of Rosenbrock methods applied to variational data assimilation problems, J. Comput. Phys., 165, 496-510, 2000.

Daescu, D., Sandu, A., and Carmichael, G.R.: Direct and Adjoint Sensitivity Analysis of Chemical Kinetic Systems with KPP: II - Validation and Numerical Experiments, Atmos. Environ., 37, 5097-5114, 2003.

Daescu, D.N.: On the sensitivity equations of four-dimensional variational (4D-Var) data assimilation, Monthly Weather Review, 136 (8), 3050-3065, 2008.

Daley, R.: Atmospheric Data Analysis, Cambridge University Press, p. 457pp, 1991.

Damian, V., Sandu, A., Damian, M., Potra, F., and Carmichael,G.R.: The Kinetic Pre-Processor KPP - A Software Environment for Solving Chemical Kinetics, Comp. and Chem. Eng., 26, 11, 1567-1579, 2002.

Derber, J. C., Parrish, D. F., Lord, S. J.: The New Global Operational Analysis System at the National Meteorological Center. Weather and Forecasting, 6, 538-547, 1991.

Dickinson, R. P., Gelinas, R. J.: Sensitivity analysis of ordinary differential equation systemssDirect method, J. Comput. Phys., 21, 123-143, 1976.

Duncan, B. N., Martin, R. V., Staudt, A. C., Yevich, R., and Logan, J. A.: Interannual and seasonal variability of biomass burning emissions constrained by satellite observations, J. Geophys. Res., 108(D2), 4100, doi:10.1029/2002JD002378, 2003.

Dunker, A. M.: Efficient calculation of sensitivity coefficients for complex atmospheric models, Atmos. Environ., 15, 1155-1161, 1981.

Dunker, A. M.: The decoupled direct method for calculating sensitivity coefficients in chemical kinetics, J. Chem. Phys., 81, 2385-2393, 1984.

Dunker, A. M., Yarwood, G., Ortmann, J. P., Wilson, G. M.: The decoupled direct method for sensitivity analysis in a threedimensional air quality modelsImplementation, accuracy, and efficiency, Environ. Sci. Technol., 36, 2965-2976, 2002.

Elbern, H., Schmidt, H., and Ebel, A.: Variational data assimilation for tropospheric chemistry modeling, J. Geophys. Res., 102, 15 967Ű15 985, 1997.

Elbern, H. and Schmidt, H.: A four-dimensional variational chemistry data assimilations scheme for Eulerian chemistry transport modeling, J. Geophys. Res., 104, 18 583Ű18 598, 1999.

Elbern, H., Schmidt, H., Talagrand, O., Ebel, A.: 4D-variational data assimilation with an adjoint air quality model for emission analysis, Environ. Modell. Software, 15, 539-548, 2000.

Elbern, H., Schmidt, H.: Ozone episode analysis by four dimensional variational chemistry data assimilation, J. Geophys. Res. [Atmos.], 106, 3569-3590, 2001.

Eller, P., Singh, K., Sandu, A., Bowman, K. W., Henze, D. K. and Lee, M.: Implementation and evaluation of an array of chemical solvers in a global chemical transport model, Geophysical Model Development, Vol. 2, p. 1–7, 2009.

Errico, R. M. and Vukicevic, T.: Sensitivity Analysis Using an Adjoint of the PSU-NCAR Mesoscale Model, Mon.Wea. Rev., 120, 1644Ű1660, 1992.

Evensen, G.: Using the extended Kalman filter with a multi-layer quasigeostrophic ocean model, Journal of Geophysical Research-Ocean, 97(C11):17905Ű17924, 1992b.

Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, J. Geophys. Res, 99, 10143–10162, 1994.

Fisher, R. A., On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London*, 1922; **Series A, 222**, 309-368, URL: http://www.jstor.org/stable/91208.

Fisher, M. and Lary, D. J.: Lagrangian four-dimensional variational data assimilation of chemical species, Quart. J. Roy. Meteorol. Soc., 121, 1681Ű1704, 1995.

Fisher, M.: Estimation of entropy reduction and degrees of freedom for signal for large variational analysis systems. *ECMWF Technical Memoranda*, 2003; **397**.

Geer, A. J., et al.:The ASSET intercomparison of ozone analyses: Method and first results, Atmos. Chem. Phys., 6, 5445Ű5474, 2006.

Gaspari, G., Cohn, S. E.: Construction of correlation functions in two and three dimensions, Quarterly Journal of the Royal Meteorological Society, Vol 125 Issue 554,723-757, 1999.

Gauthier, P., Charette, C., Fillion, L., Koclas, P. and Laroche, S.: Implementation of a 3D Variational Data Assimilation System at the Canadian Meteorological Centre. Part I: The Global Analysis, Atmosphere-Ocean, 37 (2), 103Ű156, 1999.

Gejadze, I. Y., Le Dimet, F. X., and Shutyaev, V., On analysis error covariances in variational data assimilation. *SIAM Journal on Scientific Computing*, 2008; **30(4)**, 1847-1874.

Giering, R. and Kaminski, T.: Recipes for Adjoint code Construction, ACM Trans. Math. Softw., 24, 437Ű474, 1998.

Giering, R., Kaminski, T., Todling, R., Errico, R., Gelaro, R., and Winslow, N.: Generating tangent linear and ajdoint versions of NASA/GMAOŠs Fortran-90 global weather forecast model, in: Automatic Differentiation: Applications, Theory, and Implementations, edited by: Bĺuckerr, H. M., Corliss, G., Hovland, P., Naumann, U., and Norris, B., volume 50 of ŞLecture Notes in Computational Science and EngineeringŤ, pages 275Ű284, Springer,New York, NY, 2005.

Giles, M. B. and Pierce, N. A.: An Introduction to the Adjoint Approach to Design, Flow, Turbulence and Combustion, 65, 393Ű415, 2000.

Gilliland, A. and Abbitt, P. J.: A sensitivity study of the discrete Kalman filter (DKF) to initial condition discrepancies, J. Geophys. Res., 106, 17 939Ű17 952, 2001.

Gou, T., Singh, K., and Sandu, A.: Chemical Data Assimilation with CMAQ: Continuous vs. Discrete Advection Adjoints, Lecture Notes in Computer Science, 5545, 312-321, doi:10.1007/978-3-642-01973-9, 2009.

Griewank, A. and Walther, A.: Algorithm 799: Revolve: An implementation of checkpointing for the reverse or adjoint mode of computational differentiation, ACM Trans. Math. Softw., 26, 19Ű45, 2000.

Hairer, E., Wanner G.: Solving Ordinary Differential Equations II. Stiff and Differential-Algebraic Problems, Springer: Berlin, 1991.

Hack, J. J.: Parameterization of moist convection in the National Center for Atmospheric Research Community Climate Model (CCM2), J. Geophys. Res., 99, 5551Ű5568, 1994.

Hakami, A., Odman, M. T., Russell, A. G.: High-order, direct sensitivity analysis of multidimensional air quality models, Environ. Sci. Technol., 37, 2442-2452, 2003.

Hakami, A., Henze, D. K., Seinfeld, J. H., Chai, T., Tang, Y., Carmichael, G. R. and Sandu, A.: Adjoint inverse modeling of black carbon during ACE-Asia, Journal of Geophysical Research, Vol. 110, D14301, doi:10.1029/2004JD005671, 25 pages, 2005.

Hakami, A., Henze, D. K., Seinfeld, J. H., Singh, K., Sandu, A., Kim, S., Byun, D. W. and Li, Q.: The Adjoint of CMAQ, Environ. Sci. Technol., 41 (22), 7807Ű7817, 2007.

Hakami, A., Seinfeld, J. H., Chai, T. F., Tang, Y. H., Carmichael, G. R., and Sandu, A.: Adjoint sensitivity analysis of ozone nonattainment over the continental United States, Environ. Sci. Technol., 40, 3855Ű3864, 2006.

Hall, M. C. G., Cacuci, D. G., Schlesinger,M. E.: Sensitivity analysis of a radiative-convective model by the adjoint method, J. Atmos. Sci., 39, 2038-2050, 1982.

Hall, M. C. G., Cacuci, D. G.: Physical interpretation of the adjoint functions for sensitivity analysis of atmospheric models, J. Atmos. Sci., 40, 2537-2546, 1983.

Hall, S. J., Matson, P. A., and Roth, P. M.:NOx EMISSIONS FROM SOIL: Implications for Air Quality Modeling in Agricultural Regions, Annual Review of Energy and the Environment, 21, 311-346, doi:10.1146/annurev.energy.21.1.311, 1996.

He, S., Carmichael, G. R., Sandu, A., Hotchkiss, B., Damian-Iordache, V.: Application of ADIFOR for air pollution model sensitivity studies, Environ. Modell. Software, 15, 549-557, 2000.

Henze, D. K, Hakami, A., Seinfeld, J. H.: Development of the adjoint of GEOS-Chem, Atmos. Chem. Phys., 7, 2413-2433, 2007.

Henze, D. K., Seinfeld, J. H., and Shindell, D. T., Inverse modeling and mapping U.S. air quality influences of inorganic PM2.5 precursor emissions with the adjoint of GEOS-Chem. *Atmospheric Chemistry and Physics*, 2009; **9**, 5877-5903.

Hertel, O., Berkowicz, R., Christensen, J., Hov, O.: Test of two numerical schemes for use in atmospheric transport chemistry models, Atmos. Environ., 27, 2591-2611, 1993.

Horowitz, L. W., et al.: A global simulation of tropospheric ozone and related tracers: Description and evaluation of MOZART, version 2, J. Geophys. Res., 108(D24), 4784, doi:10.1029/2002JD002853, 2003.

Horowitz, L. W.: Past, present and future concentrations of tropospheric ozone and aerosols: Methodology, ozone evaluation, and sensitivity to aerosol wet removal, J. Geophys. Res., 111, D22211, doi:10.1029/2005JD006937, 2006.

Houtekamer, P. L., Mitchell, H. L., Pellerin, G., Buehner, M., Charron, M., Spacek, L., and Hansen, B.: Atmospheric Data Assimilation with an Ensemble Kalman Filter: Results with Real Observations, Monthly Weather Review, 133(3):604-620, 2005.

Hudman, R. C., et al.: Surface and lightning sources of nitrogen oxides over the United States: Magnitudes, chemical evolution and outflow, J. Geophys. Res., 112, D12S05, doi:10.1029/2006JD007912, 2007.

Hwang, D., Byun, D. W., and Odman, M. T.: An automatic differentiation technique for sensitivity analysis of numerical advection schemes in air quality models, Atmos. Environ. 31 (6), 879Ű888, 1997.

Jacob, D. J., Liu, H., Mari, C., and Yantosca, B. M.: Harvard wet deposition scheme for GMI, http://gmi.gsfc.nasa.gov/models/jacob wetdep.pdf, 2000.

Jacobson, M. Z., Turco, R. P.: SMVGEAR: A Sparse-Matrix, Vectorized Gear code for atmospheric models, Atmos. Environ., 28, 273-284, 1994.

Jacobson, M. Z.: Computation of Global Photochemistry with SMVGEAR-II, Atmos. Environ., 29, 2541Ű2546, 1995.

Jacobson, M.Z.: Technical Note: Improvement of SMVGEAR II on Vector and Scalar Machines through Absolute Error Tolerance Control, Atmos. Environ., 32, 791-796, 1998.

Jacobson, M. Z.: Fundamentals of Atmospheric Modeling, 2nd ed., Cambridge University Press: New York, 2005.

Jazwinski, A. H., Stochastic processes and filtering theory. *Academic Press, New York*, 1970.

Jones, D. B. A., Bowman, K. W., Palmer, P. I., Worden, J. R., Jacob, D. J., Hoffman, R. N., Bey, I., and Yantosca, R. M., Potential of observations from the Tropospheric Emission Spectrometer to constrain continental sources of carbon monoxide. *Journal of Geophysical Research*, 2003; **108**, D24.

Jones, D. B. A., Bowman, K. W., Logan, J. A., Heald, C. L., Liu, J., Luo, M., Worden, J. and Drummond, J.: Inversion analysis of carbon monoxide emissions using data from the TES and MOPITT satellite instruments, Atmospheric Chemistry and Physics Discussions,7,6, 17625–17662, 2007.

Kalman, R. E., A new approach to linear filtering and prediction problems. *Transaction of the ASME - Journal of basic Engineering*, 1960; **Series D(82)**, 35-45.

Kalnay, E.: Atmospheric modeling, data assimilation and predictability, Cambridge University Press, 2002.

Kelley, C. T., and Sachs, E. W.: Truncated Newton Methods For Optimization With Inaccurate Functions And Gradients, SIAM Journal on Optimization, 10, 43–55, 1999.

Khattatov, B. V., Gille, J. C., Lyjak, L. V., Brasseur, G. P., Dvortsov, V. L., Roche, A. E. and Walters, J.: Assimilation of photochemically active species and a case analysis of UARS data, Journal of Geophysical Research, 104:18715–18737, 1999.

Khattatov, B. V., Lamarque, J. -F., Lyjak, L. V., Menard, R., Levelt, P., Tie, X., Brasseur, G. P. and Gille, J. C.: Assimilation of satellite observations of long-lived chemical species in global chemistry transport models, J. Geophys. Res., 105(D23), 29–135, 2000.

Kopacz, M., Jacob, D. J., Henze, D. K., Heald, C. L., Streets, D. G., and Zhang, Q., A comparison of analytical and adjoint Bayesian inversion methods for constraining Asian sources of CO using satellite (MOPITT) measurements of CO columns. *Journal of Geophysical Research*, 2009; **114**, D04305.

Kopacz, M., Jacob, D. J., Henze, D. K., Heald, C. L., Streets, D. G., and Zhang, Q.: A comparison of analytical and adjoint Bayesian inversion methods for constraining Asian sources of CO using satellite (MOPITT) measurements of CO columns, J. Geophys. Res., doi:0.1029/2007JD009264, 2009.

Kullback, S., Information theory and statistics. *Wiley, New York*, 1968.

Lahoz, W. A., et al.: The Assimilation of Envisat data (ASSET) project, Atmos. Chem. Phys., 7, 1773-1796, 2007.

Lamarque, J.-F., Khattatov, B. V. and Gille, J. C.: Constraining tropospheric ozone column through data assimilation, J. Geophys. Res., 107(D22), 4651, doi:10.1029/2001JD001249, 2002.

Lamb, R. G., Chen, W. H., and Seinfeld, J. H.: Numerico-empirical analysis of atmospheric diffusion theories, J. Atmos. Sci., 32, 1794Ű1807, 1975.

Lanser, D., Verwer, J. G.: Analysis of operator splitting for advection-diffusion-reaction problems from air pollution modeling, Centrum voor Wiskunde en Informatica Report MAS-R9805, 1998.

Laroche, S., Dorval, E. C., Canada, Q. C., Gauthier, P., Tanguay, M., Pellerin, S., and Morneau, J.: Evaluation of the operational 4D-Var at the Meteorological Service of Canada, 21st Conference on Weather Analysis and Forecasting, 14B.3, 2005.

LeDimet, F.-X. and Talagrand, O.: Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects, Tellus 38A, 97-110, 1986.

Li, Y., Navon, I. M., Courtier, P., and Gauthier, P., Variational data assimilation with a semi-Lagrangian semi-implicit global shallow water equation model and its adjoint. *Monthly Weather Review*, 1993; **121(6)**, 1759Ű1769.

Li, Z. and Navon, I. M.: Optimality of variational data assimilation and its relationship with the Kalman filter and smoother, Q. J. R. Meteorol. Soc., 127, pp. 661-683, 2001.

Li, Q., Jacob, D. J., Park, R. J., Wang, Y., Heald, C. L., Hudman, R. C., Yantosca, R. M., Martin, R. V., and Evans, M. J.: North American pollution outflow and the trapping of convectively lifted pollution by upper-level anticyclone, J. Geophys. Res., 110, D10301, doi:10.1029/2004JD005039, 2005.

Lin, S. J. and Rood, R. B.: Multidimensional flux-form semi-Lagrangian transport schemes, Mon. Wea. Rev., 124, 2046Ű2070, 1996.

Lions, J.L., Optimal control of systems governed by partial differential equations. *Springer-Verlag*, 1971.

Liu, Z., Sandu, A.: Analysis of discrete adjoints for upwind numerical schemes, Lecture Notes Computational Science, 3515, 829-836, 2005.

Logan, J. A.: Trends in the vertical distribution of ozone: An analysis of ozonesonde data, J. Geophys. Res., 99(D12), 25,553Ű25,585, 1994.

Logan, J. A.: An analysis of ozonesonde data for the troposphere: Recommendations for testing 3-D models and development of a gridded climatology for tropospheric ozone, J. Geophys. Res., 104(D13), 16,115Ű16,149, 1999.

Lorenc, A. C., Analysis methods for numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society* 1986;**112**, 1177-1194.

Lorenz, E., Predictability: A problem partly solved. *Proceedings of the Seminar on Predictability*, 1996; Shinfield Park, Reading, UK, ECMWF.

Majda, A. J. and Wang, X., Nonlinear dynamics and statistical theories for basic geophysical flows. *Cambridge University Press*, 2006.

Mallet, V. and Sportisse, B.: 3-D chemistry-transport model Polair: numerical issues, validation and automatic-differentiation strategy, Atmos. Chem. Phys. Discuss., 4, 1371Ű1392, 2004.

Mallet, V., Sportisse, B.: A comprehensive study of ozone sensitivity with respect to emissions over Europe with a chemistry-transport model. J. Geophys. Res., [Atmos.], 110, D22302. doi:10.1029/2005JD006234, 2005.

Mallet, V. and Sportisse, B.: Uncertainty in a chemistry-transport model due to physical parameterizations and numerical approximations: An ensemble approach applied to ozone modeling, J. Geophys. Res., 111, D01302, doi:10.1029/2005JD006149, 2006.

Marchuk, G. I.: Numerical Solution of the Problems of the Dynamics of the Atmosphere and the Ocean (in Russian), Gidrometeoizdat: St. Petersburg, 1974.

Marchuk, G. I.: Mathematical Models in Environmental Problems, Elsevier Science Pub. Co.: Amsterdam, 1986.

Marchuk, G. I., Shutyaev, V., Bocharov, G.: Adjoint equations and analysis of complex systems: Application to virus infection modeling, J. Comput. Appl. Math., 184, 177-204, 2005.

Martien, P. T., Harley, R. A., and Cacuci, D. G.: Adjoint Sensitivity Analysis for a three-dimensional photochemical model: implementation and method comparison, Environ. Sci. Technol., 40 (8), 2663Ű2670, doi:10.1021/es0510257, 2006.

Martien, P. T. and Harley, R. A.: Adjoint sensitivity analysis for a three-dimensional photochemical model: Application to Southern California, Environ. Sci. Technol., 40, 4200Ű4210, 2006.

McLinden, C. A., Olsen, S. C., Hannegan, B., Wild, O., Prather, M. J., and Sundet, J.: Stratospheric ozone in 3-D models: A simple chemistry and the cross-tropopause flux, J. Geophys. Res., 105(D11), 14,653Ű14,665, doi:10.1029/2000JD900124, 2000.

Meirink, J. F., Eskes, H. J., and Goede, A. P. H.: Sensitivity analysis of methane emissions derived from SCIAMACHY observations through inverse modelling, Atmos. Chem. Phys., 6, 1275Ű1292, 2006.

Menard, R., Cohn, S. E., Chang, L. -P. and Lyster, P. M.: Assimilation of stratospheric chemical tracer observations using a Kalman Filter I: Formulation, Mon. Weather Rev., 128, 2654-2671, 2000.

Menut, L., Vautard, R., Beekmann, M., and Honore, C.: Sensitivity of photochemical pollution using the adjoint of a simplified chemistry-transport model, J. Geophys. Res., 105, 15 379Ű 15 402, 2000.

Menut, L.: Adjoint modeling for atmospheric pollution process sensitivity at regional scale, J. Geophys. Res., 108, 8562, doi: 10.1029/2002JD002549, 2003.

Miehe, P, Sandu, A.: Forward, tangent linear, and adjoint Runge-Kutta methods in KPP-2.2, Lect. Notes Comput. Sci., 3993, 120-127, 2006.

Miller, R. N., Ghil, M., and Gauthiez, F.: Advanced data assimilation in strongly nonlinear dynamical systems, Journal of the Atmospheric Sciences, 51(8):1037Ű1056, 1994.

Moorthi, S. and Suarez, M. J.: Relaxed Arakawa-Schubert: A parameterization of moist convection for general circulation models, Mon. Weather Rev., 120, 978Ű 1002, 1992.

Muller, J. F. and Stavrakou, T.: Inversion of CO and NOx emissions using the adjoint of the IMAGES model, Atmos. Chem. Phys., 5, 1157Ű1186, 2005.

Munro, R., Siddans, R., Reburn, W. J., and Kerridge, B. J.: Direct measurement of tropospheric ozone distributions from space, Nature, 392(6672), 168Ű171, 1998.

Napelenok, S. L., Cohan, D., Hu, Y., and Russel, A. G.: Decoupled direct 3D sensitivity analysis for particulate matter (DDM-3DPM), Atmos. Environ., 40, 6112Ű6121, 2006.

Nassar, R., Logan, J. A., Worden, H. M., et al.: Validation of Tropospheric Emission Spectrometer (TES) nadir ozone profiles using ozonesonde measurements,J. Geophys. Res., 113, D15S17, doi:10.1029/2007JD008819, 2008.

Navon, I. M., Zou, X., Derber, J., Sela, J.: Variational data assimilation with an adiabatic version of the NMC spectral model, Mon. Weather Rev., 120, 1433-1446, 1992.

Navon, I. M.: Practical and theoretical aspects of adjoint parameter estimation and identifiability in meteorology and oceanography, Dyn. Atmos. Oceans, 27, 55-79, 1998.

Navon, I. M.: Data assimilation for Numerical Weather Prediction: a review, in: Data Assimilation for Atmospheric, Oceanic, and Hydrologic Applications, XVIII, 475 p. 326 illus., Hardcover, ISBN: 978-3-540-71055-4, 2009.

Nester, K. and Panitz, H. J.: Sensitivity analysis by the adjoint chemistry transport model DRAIS for an episode in the Berlin Ozone (BERLIOZ) experiment, Atmos. Chem. Phys., 6, 2091Ű 2106, 2006.

Oltmans, S. J., et al.: Long-term changes in tropospheric ozone, Atmos. Environ., 40, 3156Ű3173, 2006.

Ott, E., Hunt, B. R., Szunyogh, I., Zimin, A.V., Kostelich, E. J., Kostelich, M., Corazza, M., Sauer, T., Kalnay, E., Patil, D. J. and Yorke, J. A: A local ensemble Kalman Filter for Atmospheric Data Assimilation, Tellus, Vol. 56A, pp. 415-428, 2004.

Palmer, P. I., Jacob, D. J., Jones, D. B. A., Heald, C. L., Yantosca, R. M., Logan, J. A., Sachse, G. W. and Streets, D. G.: Observations over the western Pacific, 2003.

Parrington, M., Jones, D. B. A., Bowman, K. W., Horowitz, L. W., Thompson, A. M., Tarasick, D. W. and Witte, J. C.: Estimating the summertime tropospheric ozone distribution over North America through assimilation of observations from the Tropospheric Emission Spectrometer, Journal of Geophysical Research, Vol 113, D18307, doi:10.1029/2007JD009341, 2008.

Parrington, M., Jones, D. B. A., Bowman, K. W., Thompson, A. M., Tarasick, D. W., Merrill, J., Oltmans, S. J., Leblanc, T., Witte, J. C. and Millet, D. B.: Impact of the assimilation of ozone from the tropospheric emission spectrometer on surface ozone across North America, Geophysical Research Letters 36 (4), 2009.

Parrish, D. F. and Derber, J. C.: The national meteorological center's spectral statistical-interpolation analysis system, Monthly Weather Review, (120), p. 1747–1763, 1992.

Pierce, R. B., et al.: Chemical data assimilation estimates of continental U. S. ozone and nitrogen budgets during the Intercontinental Chemical Transport Experiment-North America, J. Geophys. Res., 112, D12S21, doi:10.1029/2006JD007722, 2007.

Pires, C., Vautard, R., and Talagrand, O.: On extending the limits of variational assimilation in nonlinear chaotic systems, Tellus, 48A, 960-121, 1996.

Price, C., and Rind, D.: A Simple Lightning Parameterization for Calculating Global Lightning Distributions, J. Geophys. Res., 97(D9), 9919Ű9933, doi:10.1029/92JD00719, 1992.

Parrington, M., Jones, D. B. A., Bowman, K. W., Horowitz, L. W., Thompson, A. M., Tarasick, D. W., Witte, J. C., Estimating the summertime tropospheric ozone distribution over North America through assimilation of observations from the Tropospheric Emission Spectrometer. *Journal of Geophysical Research*, 2008; **Vol 113**, D18307.

Parrington, M., Jones, D. B. A., Bowman, K. W., Thompson, A. M., Tarasick, D. W., Merrill, J., Oltmans, S. J., Leblanc, T., Witte, J. C., Millet, D. B., Impact of the assimilation of ozone from the Tropospheric Emission Spectrometer on surface ozone across North America. *Geophysical Research Letters*, 2009; **36(4)**, L04802.

Pudykiewicz, J. A.: Application of adjoint tracer transport equations for evaluating source parameters, Atmos. Environ., 32, 3039Ű3050, 1998.

Quelo, D., Mallet, V., Sportisse, B.: Inverse modeling of NOx emissions at regional scale over northern France: Preliminary investigation of the second-order sensitivity, J. Geophys. Res., [Atmos.], 110, D24310. doi:10.1029/2005JD006151, 2005.

Rabier, F., Jarvinen, H., Klinker, E., Mahfouf, J. -F. and Simmons, A.: The ECMWF operational implementation of four-dimensional variational data assimilation I: Experimental results with simplified physics, Quarterly Journal of the Royal Meteorological Society, 126:1143–1170, 2000.

Rabier, F., Fourrie, N., Chafa, D., and Prunet, P., Channel selection methods for Infrared Atmospheric Sounding Interferometer radiances. *Quarterly Journal of the Royal Meteorological Society*, 2002; **128**, 1011-1027.

Reuther, J. J., Jameson, A., Alonso, J. J., Rimlinger, M. J., and Saunders, D.: Constrained multipoint aerodynamic shape optimization using an adjoint formulation and parallel computers, part 1, J. Aircraft, 36, 51Ű60, 1999.

Rodgers, C. D., Retrieval of atmospheric temperature and composition from remote measurements of thermal radiation. *Reviews of Geophysics and Space Physics*, 1976; **14**, 609-624.

Rodgers, C. D., Information content and optimization of high spectral resolution measurements. *Optical Spectroscopic Techniques and Instrumentation for Atmospheric and Space Research*, **SPIE Volume 2830**, 136-147.

Rodgers, C. D., Information content and optimization of high spectral resolution measurements. *Advances in Space Research*, 1998; **21**, 361-367.

Rodgers, C. D., Inverse methods for atmospheric sounding: Theory and Practice. *World Scientific: Singapore*, 2000.

Sandu, A., Verwer, J. G., Blom, J. G., Spee, E. J., Carmichael, G. R., and Potra, F. A.: Benchmarking stiff ODE solvers for atmospheric chemistry problems: II - Rosenbrock solvers, Atmos. Environ., 31, 3459Ű3472, 1997.

Sandu, A., Daescu, D., and Carmichael, G.R.: Direct and Adjoint Sensitivity Analysis of Chemical Kinetic Systems with KPP: I - Theory and Software Tools, Atmos. Environ., 37, 5083-5096, 2003.

Sandu, A., Daescu, D. N., Carmichael, G. R. and Chai, T.: Adjoint sensitivity analysis of regional air quality models, Journal of Computational Physics, Vol. 204, p. 222-252, 2005.

Sandu, A., Sander, R.: Technical note: Simulating chemical systems in Fortran90 and Matlab with the Kinetic PreProcessor KPP-2.1, Atmos. Chem. Phys., 6, 187-195, 2006.

Sandu A. and Zhang L.: Discrete second order adjoints in atmospheric chemical transport modeling, Journal of Computational Physics, 227 (12), 5949–5983, 2008.

Sasaki, Y. K.: An objective analysis based on the variational method, J. Met. Soc. Jap. II(36), 77–88, 1958.

Schichtel, B. A., Malm, W. C., Gebhart, K. A., Barna, M. G., and Knipping, E. M.: A hybrid source apportionment model integrating measured data and air quality model results, J. Geophys. Res., 111, D07301, doi:10.1029/2005JD006238, 2006.

Schmidt, H. and Martin, D.: Adjoint sensitivity of episodic ozone in the Paris area to emissions on the continental scale, J. Geophys. Res., 108, 8561Ű8577, doi:10.1029/2001D001583, 2003.

Segers, A. J., Eskes, H. J., van der A, R. J., van Oss, R. F. and van Velthoven, P. F. J.: Assimilation of GOME ozone profiles and a global chemistry-transport model, using a Kalman Filter with anisotropic covariance, Quarterly Journal of the Royal Meteorological Society, 131, 477-502, 2005.

Seinfeld, J. H., Pandis, S. N.: Atmospheric Chemistry and Physics: From Air Pollution to Climate Change, 2nd ed., J. Wiley: Hoboken, NJ, 2006.

Shannon, C. E. and Weaver, W., The mathematical theory of communication. *University of Illinois Press, Urbana, IL.*, 1949.

Singh, K., Eller, P., Sandu, A., Bowman, K. W., Jones, D. B. A. and Lee, M.: Improving GEOS-Chem model forecasts through profile retrievals from Tropospheric Emission Spectrometer, in: Lecture Notes on Computational Science vol. 5545, p. 302–311, International Conference on Computational Science 2009, Baton Rouge, Louisiana, May 25–27, 2009.

Singh, K., Eller, P., Sandu, A., Henze, D., Bowman, K. W., Kopacz, M. and Lee, M.: Towards the construction of a standard adjoint GEOS-Chem model, High Performance Computing Symposium (HPC 2009) at Spring Simulation Multiconference (SpringSim'09), San Diego, California, March 22–27, 2009.

Singh, K., Jardak, M., Sandu, A., Bowman, K. W., Lee, M., Jones, D.: Construction of non-diagonal background error covariance matrices in global chemical data assimilation, Submitted to *Geophysical Model Development*, 2010.

Sirkes, Z. and Tziperman, E.: Finite difference of adjoint or adjoint of finite difference?, Mon. Wea. Rev., 125, 3373Ű3378, 1997.

Stavrakou, T. and Muller, J. F.: Grid-based versus big region approach for inverting CO emissions using Measurement of Pollution in the Troposphere (MOPITT) data, J. Geophys. Res., 111, D15304, doi:10.1029/2005JD006896, 2006.

Stevenson, D. S., et al.: Multimodel ensemble simulations of present-day and near-future tropospheric ozone, J. Geophys. Res., 111, D08301, doi:10.1029/2005JD006338, 2006.

Stewart, L. M., Dance, S. L., Nichols, N. K., Correlated observation errors in data assimilation. *International Journal for Numerical Methods in Fluids*, 2008; **56**, 1521-1527.

Talagrand, O. and Courtier, P.: Variational assimilation of meteorological observations with the adjoint of the vorticity equations. Part I: Theory, Quart. J. Roy. Meteorol. Soc., 113, 1311Ű1328, 1987.

Tarasick, D. W., Fioletov, V. E., Wardle, D. I., Kerr, J. B., and Davies, J.: Changes in the vertical distribution of ozone over Canada from ozonesondes: 1980Ű2001, J. Geophys. Res., 110, D02304, doi:10.1029/2004JD004643, 2005.

Tellmann, S., Rozanov, V. V., Weber, M., and Burrows, J. P.: Improvements in the tropical ozone profile retrieval from GOME-UV/Vis nadir spectra, Adv. Space Res., 34(4), 739Ű743, 2004.

TES Science Team, TES L2 Data Users Guide, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California. (Available at `http://tes.jpl.nasa.gov/docsLinks/DOCUMENTS/TESL2DataUsersGuidev2.0.pdf`), 2006.

Thacker, W. C., The role of the Hessian matrix in fitting models. to measurements. *Journal of Geophysical Research*, 1989; **94(C5)**, 6177-6196.

Thompson, A. M., et al.(2007a): Intercontinental chemical transport experiment ozonesonde network study (IONS) 2004: 1. Summertime upper troposphere/lower stratosphere ozone over northeastern North America, J. Geophys. Res., 112 D12S12, doi:10.1029/2006JD007441, 2007. Thompson, A. M., et al.(2007b): Intercontinental chemical transport experiment ozonesonde network study (IONS) 2004: 2. Tropospheric ozone budgets and variability over northeastern North America, J. Geophys. Res., 112 D12S13, doi:10.1029/2004JD005359, 2007.

Thuburn, J. and Haine, T. W. N: Adjoints of nonoscillatory advection schemes, J. Comp. Phys., 171, 616Ű631, 2001.

Todling, R., Cohn, S. E.: Suboptimal Schemes for Atmospheric Data Assimilation Based on the Kalman Filter, Monthly Weather Review, 122, 2530-2557, 1994.

Vautard, R., Beekmann, M., and Menut, L.: Applications of adjoint modelling in atmospheric chemistry: sensitivity and inverse modelling, Environ. Modell. Softw., 15, 703Ű709, 2000.

Vukicevic, T. and Hess, P.: Analysis of troposhperic transport in the Pacific Basin using the adjoint technique, J. Geophys. Res., 105, 7213Ű7230, 2000.

Vukicevic, T., Steyskal, M., and Hecht, M.: Properties of advection algorithms in the context of variational data assimilation, Mon. Wea. Rev., 129, 1221Ű1231, 2001.

Wahba, G., Johnson, D. R., Gao, F., and Gong, J.: Adaptive tuning of numerical weather prediction models: Randomized GCV in three and four dimensional data assimilation, Monthly Weather Review, vol. 123, issue 11, p. 3358, 1995.

Worden, J. R., Bowman, K. W. and Jones, D. B. A.: Characterization of atmospheric profile retrievals from Limb Sounding Observations of an inhomogeneous atmosphere, J. Quant. Spectrosc. Radiat. Transfer, 86, (03)00274-7, 2004.

Worden, H. M., et al. (2007), Comparisons of Tropospheric Emission Spectrometer (TES) ozone profiles to ozonesondes: Methods and initial results, J. Geophys. Res., 112, D03309, doi:10.1029/2006JD007258, 2007.

Wu, L., Mallet, V., Bocquet, M. and Sportisse, B.: A comparison study of data assimilation algorithms for ozone forecasts, J. Geophys. Res., 113, D20310, 2008.

Xu, Q., Measuring information content from observations for data assimilation: relative entropy versus Shannon entropy difference. *Tellus, A.* 2006, 198-209.

Yang, Y. J., Wilkinson, J. G., Russell, A. G: Fast, direct sensitivity analysis of multidimensional photochemical models. Environ. Sci. Technol., 31, 2859-2868, 1997.

Yang, Y. J., Odman, M. T., Russell, A. G.: Fast three-dimensional sensitivity analysis of photochemical air quality models: an application to southern California, in: Proceedings of the Air and Waste Management AssociationŠs Annual Meeting and Exhibition 98-WP76A (06):2, 1998.

Yang, Y. J., Wilkinson, J. G., Odman, M. T., Russell, A. G.: Ozone sensitivity and uncertainty analysis using DDM-3D in a photochemical air quality model, Air Pollution Modeling and its Application XIII: Proceedings of the 23rd NATO/CCMS International Technical Meeting on Air Pollution Modelling and Its Application, Varna, Bulgaria:183Ű194, 2000.

Yevich, R., and Logan, J. A.: An assessment of biofuel use and burning of agricultural waste in the developing world, Global Biogeochem. Cycles, 17(4), 1095, doi:10.1029/2002GB001952, 2003.

Zhang, G. J., and McFarlane, N. A.: Sensitivity of climate simulations to the parameterization of cumulus convection in the Canadian Climate Centre general circulation model, Atmos. Ocean, 33 (3), 407Ű446, 1995.

Zhang, L., Constantinescu, E. M., Sandu, A., Tang, Y., Chai, T., Carmichael, G. R., Byun, D., Olaguer, E.: An adjoint sensitivity analysis and 4D-Var data assimilation study of Texas air quality, Atmospheric Environment, Vol. 42, Issue 23, p. 5787–5804, 2008.

Zhu, C., Byrd, R. H., Lu, P., and Nocedal, J.: L-BFGS-B: a limited memory FORTRAN code for solving bound constrained optimization problems, Tech. rep., Northwestern University, 1994.

Zhu, C., Byrd, R. H. and Nocedal, J.: L-BFGS-B: Algorithm 778: L-BFGS-B, FORTRAN routines for large scale bound constrained optimization, ACM Transactions on Mathematical Software, Vol 23, Num. 4, pp. 550 - 560, 1997.

Zou, X., Navon, I. M., Sela, J., Variational data assimilation with moist threshold processes using the NMC spectral model. *Tellus A.*, 1993; **45A**, 370-387.

Zupanski, D., Information measures in ensemble data assimilation. *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications*, 2009.

Appendix A

A.1 Approximate sampling of the posterior distribution in 4D-Var

We use the hybrid of 4D-Var and ensemble approach discussed in [Cheng et al., 2010] to generate our posteriori distribution. Suppose we are given the background state $x_0^B \in \mathbb{R}^n$ and the background error covariance matrix $B \in \mathbb{R}^{nxn}$, the $\mathrm{N_{ens}}$ normally distributed perturbation vectors with zero mean and variance B can be generated as:

$$\Delta x_0^B \in \mathcal{N}(0, B), \quad i = 1, 2, \ldots, \mathrm{N_{ens}}. \tag{A.1}$$

Starting from x_0^B, we save the first k iterates $x_0^{(j)}, j = 1, \ldots, k$, generated by the numerical optimization routine used in the 4D-Var assimilation. The value of k is chosen based on the rate of convergence of the optimization routine. Since the reduction in cost function is fastest during the initial iterations, $k <<$ dimension of the state vector. Let S be the matrix with columns as normalized 4D-Var increments

$$S = \left\{ \frac{x_0^{(j)} - x_0^{(j-1)}}{||x_0^{(j)} - x_0^{(j-1)}||} \right\}, \quad j = 1, 2, \ldots, k$$

with $x_0^{(0)} = x_0^B$. Using the singluar value decomposition of S

$$S = U\Sigma V^T$$

we derive the orthogonal projector onto the orthogonal complement of Range(U) as

$$P = I_{nxn} - UU^T$$

Using P, the ensemble perturbations Δx_0^B are projected from forecast space onto the analysis space

$$\Delta x_0^A = P\Delta x_0^B \tag{A.2}$$

A.2 Properties of random quadratic functions

In Chapter 5 we use the following useful property of random quadratic functions.

Let $\mathbb{Q} = \mathbb{Q}^T$ be a symmetric positive semidefinite matrix and ζ a random vector with $\mathbb{E}[\zeta] = \mu$ and $\mathrm{cov}[\zeta] = \mathbb{C}$. Then the quadratic function $\zeta^T \mathbb{Q} \zeta$ has the following statistics:

$$\mathbb{E}\left[\zeta^T \cdot \mathbb{Q} \cdot \zeta\right] = \mathrm{trace}\left(\mathbb{Q}\mathbb{C}\right) + \mu^T \cdot \mathbb{Q} \cdot \mu \,, \tag{A.3a}$$

$$\mathrm{var}\left[\zeta^T \cdot \mathbb{Q} \cdot \zeta\right] = \mathrm{trace}\left(\mathbb{Q}\mathbb{C}\mathbb{Q}\mathbb{C}\right) + 4\mu^T \cdot \mathbb{Q}\mathbb{C}\mathbb{Q} \cdot \mu \,. \tag{A.3b}$$

If $\mathbf{x} \in \mathcal{N}\left(\mathbf{x}_0^A, \mathbb{A}_0\right)$ then $\mathbf{x} - \mathbf{x}_0^A \in \mathcal{N}\left(0, \mathbb{A}_0\right)$ and

$$\mathbb{E}^{\mathrm{A}}\left[\frac{1}{2}\left(\mathbf{x} - \mathbf{x}_0^A\right)^T \mathbb{A}_0^{-1}\left(\mathbf{x} - \mathbf{x}_0^A\right)\right] = 0 + \frac{1}{2}\,\mathrm{trace}\left(\mathbb{A}_0^{-1}\,\mathbb{A}_0\right) = \frac{n}{2} \,. \tag{A.4}$$

Similarly, $\mathbf{x} - \mathbf{x}_0^B \in \mathcal{N}\left(\mathbf{x}_0^A - \mathbf{x}_0^B, \mathbb{A}_0\right)$ and

$$\mathbb{E}^{\mathrm{A}}\left[\frac{1}{2}\left(\mathbf{x} - \mathbf{x}_0^B\right)^T \mathbb{B}_0^{-1}\left(\mathbf{x} - \mathbf{x}_0^B\right)\right] = \frac{1}{2}\left(\mathbf{x}_0^A - \mathbf{x}_0^B\right)^T \mathbb{B}_0^{-1}\left(\mathbf{x}_0^A - \mathbf{x}_0^B\right) + \frac{1}{2}\,\mathrm{trace}\left(\mathbb{B}_0^{-1}\,\mathbb{A}_0\right) \,. \tag{A.5}$$

A.3 4D-Var data assimilation with linear models, linear observation operators, and Gaussian errors

In this section we consider the case where the model dynamics is linear

$$\mathcal{M}_{t_0 \to t_i}\left(\mathbf{x}_0\right) = M_i\,\mathbf{x}_0 \tag{A.6}$$

and the observation operator is also linear, %

$$\mathcal{H}\left(\mathbf{x}_i\right) = H_i\,\mathbf{x}_i \,. \tag{A.7}$$

In addition, we assume that the background errors and the observation errors are both normally distributed. In this case the 4D-Var cost function is:

$$\begin{aligned}
\mathcal{J}^B(\mathbf{x}_0) &= \frac{1}{2}\left(\mathbf{x}_0 - \mathbf{x}_0^B\right)^T \mathbb{B}^{-1}\left(\mathbf{x}_0 - \mathbf{x}_0^B\right) \\
\mathcal{J}^{\mathrm{obs}}(\mathbf{x}_0) &= \sum_{i=1}^{N} \mathcal{J}_i^{\mathrm{obs}}(\mathbf{x}_0) \\
\mathcal{J}_i^{\mathrm{obs}}(\mathbf{x}_0) &= \frac{1}{2}\left(H_i\,\mathbf{x}_i - \mathbf{y}_i\right)^T \mathbb{R}_i^{-1}\left(H_i\,\mathbf{x}_i - \mathbf{y}_i\right) \\
&= \frac{1}{2}\left(H_i\,M_i\,\mathbf{x}_0 - \mathbf{y}_i\right)^T \mathbb{R}_i^{-1}\left(H_i\,M_i\,\mathbf{x}_0 - \mathbf{y}_i\right)
\end{aligned}$$

The posterior distribution is Gaussian $\mathcal{P}^A(\mathbf{x}_0) = \mathcal{N}\left(\mathbf{x}_0^A, \mathbb{A}_0\right)$. The posterior covariance matrix $\mathbb{A}_0$ satisfies

$$\mathbb{A}_0^{-1} = \mathbb{B}_0^{-1} + \sum_{i=0}^{N} M_i^T \, H_i^T \, \mathbb{R}_i^{-1} \, H_i \, M_i \tag{A.8}$$

and the analysis initial condition x_0^A obtained by solving the linear system

$$\mathbb{A}_0^{-1} \cdot \left(\mathbf{x}_0^A - \mathbf{x}_0^B\right) = \sum_{i=0}^{N} M_i^T \, H_i^T \, \mathbb{R}_i^{-1} \, \left(\mathbf{y}_i - H_i \, M_i \, \mathbf{x}_0^B\right) \; . \tag{A.9}$$