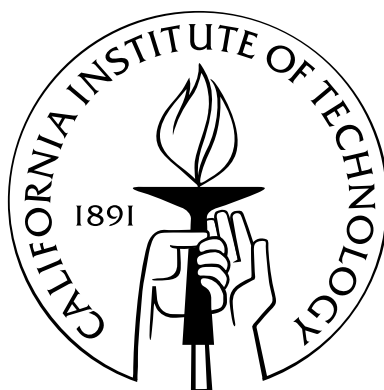# Forward and Inverse Analysis of Chemical Transport Models

Thesis by

Daven K. Henze

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2007

(Defended May 23, 2007)

To Nina, my grandmother who never doubted

Acknowledgements

First, I would like to thank John Seinfeld, whose fundamental approach to science drew me to Caltech. I greatly appreciate a number of others who also encouraged, inspired, and advised me, in particular, Adrian Sandu, Greg Carmichael, Dylan Jones, Tapio Schneider, Rick Flagan, and Daniel Jacob. Additional thanks to Qinbin Li and Kevin Bowman, collaborators at JPL, who also allowed generous use of their computational facilities. I owe Amir Hakami for many valuable discussions and insights into sensitivity analysis, and Philip Stier, for making sure that I read important climate papers. I would like to thank Sally Ng and Jesse Kroll for being willing to bridge the gap between Keck and Spalding to have a quick chat about SOA on occasion. Everyone in the group was a pleasure to work with. Even Shane.

I would like to thank the William H. Corcoran Memorial Fellowship for providing funding for my first year at Caltech, and William and Sonya Davidow for supporting me during my last year. I wish to recognize support from the U.S. Environmental Protection Agency and the National Science Foundation, the latter also for providing access to TeraGrid resources at the National Center for Supercomputing Applications.

I owe my particularly wonderful experience at Caltech to an incredible range of people. My gratitude to Alex, Stephane, Cesar, C-Boys and Girl, Jesse, Corinne, Gwyn, Fady, Min, Troy, Marta (for realizing I needed a turtle), and to all who ever gave me a ride or lent me their cell phone. Thank you Emma and ML for keeping me well fed, thank you Jora and Edgar for making me relax, and thank you Elena, for being everything to me.

Historically, though there were many others, I credit Scott Olson for my interest in computational science. I owe much to my parents, for supporting me unwaveringly, and to my sister, for simply being the best.

Daven

Abstract

Assessing the discrepancy between modeled and observed distributions of aerosols is a persistent problem on many scales. Tools for analyzing the evolution of aerosol size distributions using the adjoint method are presented in idealized box model calculations. The ability to recover information about aerosol growth rates and initial size distributions is assessed given a range of simulated observations of evolving systems. While such tools alone could facilitate analysis of chamber measurements, improving estimates of aerosol sources on regional and global scales requires explicit consideration of many additional chemical and physical processes that govern secondary formation of atmospheric aerosols from emissions of gas-phase precursors. The adjoint of the global chemical transport model GEOS-Chem is derived, affording detailed analysis of the relationship between gas-phase aerosol precursor emissions (SO_x, NO_x and NH_3) and the subsequent distributions of sulfate - ammonium - nitrate aerosol. Assimilation of surface measurements of sulfate and nitrate aerosol is shown to provide valuable constraints on emissions of ammonia. Adjoint sensitivities are used to propose strategies for air quality control, suggesting, for example, that reduction of SO_x emissions in the summer and NH_3 emissions in the winter would most effectively reduce non-attainment of aerosol air quality standards. The ability of this model to estimate global distributions of carbonaceous aerosol is also addressed. Based on new yield data from environmental chamber studies, mechanisms for incorporating the dependence of secondary organic aerosol (SOA) formation on NO_x concentrations are developed for use in global models. When NO_x levels are appropriately accounted for, it is

demonstrated that sources such as isoprene and aromatics, previously neglected as sources of aerosol in global models, significantly contribute to predicted SOA burdens downwind of polluted areas (owing to benzene and toluene) and in the free troposphere (owing to isoprene).

Contents

7 Conclusions 177

List of Figures

xiii

List of Tables

Chapter 1

Introduction

The combined effects of air pollution and climate change pose a tremendous challenge to stewardship of the environment in the coming century. Aerosols, or particulate matter (PM), are particles small enough to remain aloft in the atmosphere for as long as several days. Either emitted directly as particles, or formed in the atmosphere by condensation of trace gases, aerosols generated by human activities play an important role in determining both air quality and climate. Despite attempts to regulate such activity, concentrations of particulate matter presently exceed recommended thresholds throughout much of the industrialized northern hemisphere (*WHO*, 2003). Exposure to excessive aerosol concentrations is associated with increased chances of cardiovascular diseases, inhibited lung development and premature death (*Burnett et al.*, 2000; *Dominici et al.*, 2006; *Gauderman et al.*, 2004; *Pope*, 2000; *Pope et al.*, 2002). Fine particulates are also responsible for reduced visibility in national parks and scenic areas (*Malm et al.*, 2000). Additionally, these particles often travel far from their sources, where ultimately, when removed from the atmosphere, the reactive chemicals they have amassed can be hazardous to surface ecosystem, threatening biodiversity (*Stevens et al.*, 2004).

While the prevalence of these problems alone warrants further air quality regulations, increases in global industrialization in the coming decades make formulation of even more aggressive mitigation strategies an imperative. Of growing concern is that traditional

strategies for improving air quality will fail in the future owing to enhanced intercontinental transport of pollution. Despite the success of current emissions regulations in stabilizing local pollution levels in parts of North America and Europe, increased export of pollution from rapidly developing regions such as India and China may soon begin to negate such progress. Intercontinental transport of aerosol already hinders local efforts at restoring pristine environments (*Park et al.*, 2004, 2006). How much will emissions of pollutants and their precursors have to increase before it becomes impossible to maintain local air quality with purely local control strategies? Formulation of effective mitigation strategies for the future must necessarily account for increasingly delocalized ramifications of anthropogenic activity on global air quality.

That aerosols have played an important role in affecting current climate has also become increasingly clear. Since preindustrial times, anthropogenic aerosols have caused a direct radiative forcing estimated at -0.1 to -0.8 W m^{-2} (*Schulz et al.*, 2006; *Yu et al.*, 2006) and an indirect radiative forcing likely even larger (*Lohmann and Feichter*, 2005). For comparison, the total radiative forcing from anthropogenic carbon dioxide (CO_2) that has accumulated in the atmosphere over the same time period is $+1.66 \pm 0.17$ W m^{-2} (*IPCC*, 2007). Unlike assessing future trends of CO_2, a long lived inert tracer, predicting how global burdens of aerosol will respond to changes in emissions in the coming century is complicated by gas-phase reactions and thermodynamic transformations that govern their production and subsequent influence on both the chemical and physical state of the atmosphere.

Aerosols can also play an important role in determining the production and lifetime of gases such as ozone (O_3), which is in turn itself of critical concern for both air quality and climate. It has been shown that heterogeneous chemistry significantly influences radiative forcing of O_3, particularly at high latitudes where aerosol burdens are large and concentrations of OH low (*Liao et al.*, 2004). Overall, most future emission scenarios for the coming decades result in changes to global burdens of pollutants such that global warming is enhanced. Even if concentrations of greenhouse gases such as O_3 and methane are

dramatically reduced, commensurate reduction in sulfate levels would result in a net positive forcing, as current concentrations of sulfate provide substantial cooling (*Unger et al.*, 2006b). In turn, exacerbated climate change will alter the formation of such air pollution owing to changes in transport, precipitation, and temperature. While these physical climate effects are predicted to dampen future increases in global burdens of O_3 and sulfate, surface level concentrations are predicted to be regionally enhanced in many heavily populated areas (*Liao et al.*, 2006; *Murazaki and Hess*, 2006; *Unger et al.*, 2006b), while global burdens of secondary organic aerosol may even increase (*Liao et al.*, 2006). In short, assessment of the ability of such species to affect climate change must take into account these forms of coupling; considering species individually will not suffice (*Stier et al.*, 2006; *Unger et al.*, 2006a).

Clearly an attempt to design successful aerosol controls first requires comprehensive understanding of the processes that govern formation and distribution of gas and particle-phase atmospheric pollutants. Chemical transport models are essential tools for establishing links between anthropogenic activity and air quality by estimating the chemical state of the atmosphere for a given set of meteorological conditions and emissions inventories. However, discrepancies between model predictions and observations are still a persistent problem on many scales. On regional scales, estimates of nitrate aerosol are often widely inconsistent with surface measurements (*Park et al.*, 2006; *Liao et al.*, 2007). Globally, organic species comprise a significant fraction of the mass of fine particles, although the actual composition of most of this mass is largely unidentified (*Kanakidou et al.*, 2005). Simulations of secondary organic aerosol (SOA) in environments ranging from urban centers to the free troposphere have recently been shown to be lower than observed by as much as one to two orders of magnitude (*Heald et al.*, 2005; *Volkamer et al.*, 2006; *de Gouw et al.*, 2005).

To improve the quality of model predictions, one can take the bottom-up approach, wherein results from new fundamental laboratory experiments are used to revise or improve

the treatment of various processes simulated in models, building new theory, or adding new levels of detail. This approach is well suited for situations where the underlying physics or chemistry is the source of uncertainty and only a limited amount of observations are available for comparison to 3-dimensional model estimates. For example, estimates of the global budgets of SOA are difficult to verify owing to a paucity of measurements. However, key findings from recent field measurements and environmental chamber studies continue to drive model development.

When presented with a reasonable model and a set of ample observations, another approach for minimizing discrepancies between models and observations owing to uncertain model parameters is inverse modeling. Measurements themselves are used to directly constrain the set of believable model parameter values by rejecting parameters whose conditional probability given resulting estimates of observations is small (*Tarantola*, 2006). Networks of surface air quality monitoring stations, new space-born observations and coordinated field campaigns present tremendous opportunity for constraining estimates of sources and distributions of aerosols. However, integrated analysis and interpretation of these data is a formidable challenge as measurements range from vertical profiles of reactive gasses to optical properties of long lived particulate species. Corresponding model predictions are dependent upon numerous parameters, such as emissions inventories, injection heights, convective scavenging efficiencies, deposition rates and heterogeneous reaction probabilities.

A powerful approach to addressing challenges of both source attribution and model performance is the adjoint method. Originating from optimal control theory and well established in fields such as meteorology and oceanography, this approach has been applied to problems in atmospheric chemistry only relatively recently. Constructed as a functional transpose of the chemical transport model itself, an adjoint model is an efficient tool for evaluating the sensitivity of a scalar model response function with respect to numerous model parameters. For inverse modeling (data assimilation), the model response is defined

as a metric of the overall mismatch between predictions and observations. The adjoint gradients are then used to optimize control parameters to minimize this mismatch while simultaneously staying within reasonable bounds. By quantifying the relative importance of many uncertain factors, the adjoint method can be used to efficiently provide insight into the origins of the initial discrepancy between the predictions and observations, rather than simply adjusting the predicted chemical state to match the observations. A significant advantage of an adjoint model is that it calculates sensitivities on the resolution of the forward model itself, thus avoiding the common practice of lumping sources into large geographical regions for the sake of minimizing computational expense, as the latter approach smoothes out important variability in the sources of pollution (*Kopacz et al.*, submitted; *Stavrakou and Muller*, 2006).

However, even with a model that reproduces observations reasonably well, the complex nonlinear chemical and physical processes that govern formation of pollution from precursor emissions often preclude simple assessment of the precise changes in emissions required to affect a specified pollution reduction. Formulation of regulatory measures requires the use of receptor based models that quantify the origins of hazardous pollution concentrations (*Marmur et al.*, 2006). An adjoint model can also be used for this purpose. If the response function is a metric of air quality non-attainment in a select region, and the variable parameters are the emissions inventories used by the forward model, then the resulting adjoint model sensitivities give the linear estimate of the efficacy of reducing emissions from each source on achieving attainment (*Hakami et al.*, 2006).

Overall, both forward and inverse model analysis are important steps toward understanding distributions of atmospheric aerosols well enough to successfully address the regulatory challenges of controlling air quality while accounting for climate change. With these goals in mind, issues of model sensitivity, data assimilation, and source evaluation are assessed on a broad range of scales. In Chapter 2, equations are derived for analyzing the evolution of multicomponent aerosol size distributions in idealized box model calcula-

tions using the adjoint method. The ability to recover information about aerosol growth rates and initial size distributions are assessed given a range of simulated observations of evolving systems. Extension of this approach to inclusion of coagulation, with further attention to numerical techniques, can be found in *Sandu et al.* (2005). Chapter 3 presents the adjoint of a global 3-dimensional chemical transport model (GEOS-Chem) including detailed tropospheric chemistry, thermodynamic aerosol partitioning, heterogeneous chemistry, transport, and depositional losses. In Chapter 4, the adjoint of GEOS-Chem is used to evaluate sources of secondary inorganic aerosols over the United States. Application of the adjoint of GEOS-Chem to constrain source estimates of carbon monoxide (CO) using satellite measurements from MOPPIT is described in *Kopacz et al.* (submitted). Consequences of the recent revelation that photooxidation of isoprene forms significant amounts of SOA in low-NO_x environments (*Claeys et al.*, 2004b; *Kroll et al.*, 2006) on global carbon aerosol budgets are considered in Chapter 5; comparisons of total predicted organic aerosol with surface measurements in the Unites States can be found in *Liao et al.* (2007); *Zhang et al.* (submitted). Examination of the magnitude of this substantial biogenic source in comparison to recent re-evaluation of the yield of SOA from aromatic compounds (i.e. anthropogenic sources) is given in Chapter 6. The final chapter concludes with discussion of key findings and future work.

Bibliography

Burnett, R. T., J. Brook, T. Dann, C. Delocla, O. Philips, S. Cakmak, R. Vincent, M. S. Goldberg, and D. Krewski, Association between particulate- and gas-phase components of urban air pollution and daily mortality in eight Canadian cities, *Inhal. Tox.*, *12*, 15–39, 2000.

Claeys, M., et al., Formation of secondary organic aerosols through photooxidation of isoprene, *Science*, *303*(5661), 1173–1176, 2004.

de Gouw, J. A., et al., Budget of organic carbon in a polluted atmosphere: Results from the New England Air Quality Study in 2002, *J. Geophys. Res.*, *110*(D16), D16305, doi: 10.1029/2004JD005623, 2005.

Dominici, F., R. D. Peng, M. L. Bell, L. Pham, A. McDermott, S. L. Zeger, and J. M. Samet, Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases, *J. A. M. A.*, *295*(10), 1127–1134, 2006.

Gauderman, W. J., et al., The effect of air pollution on lung development from 10 to 18 years of age, *N. Engl. J. Med.*, *351*(11), 1057–1067, 2004.

Hakami, A., J. H. Seinfeld, T. F. Chai, Y. H. Tang, G. R. Carmichael, and A. Sandu, Adjoint sensitivity analysis of ozone nonattainment over the continental United States, *Environ. Sci. Technol.*, *40*(12), 3855–3864, 2006.

Heald, C. L., D. J. Jacob, R. J. Park, L. M. Russell, B. J. Huebert, J. H. Seinfeld, H. Liao,

and R. J. Weber, A large organic aerosol source in the free troposphere missing from current models, *Geophys. Res. Lett.*, *32*, L18809, doi:10.1029/2005GL023831, 2005.

Intergovernmental Panel on Climate Change (IPCC), Summary for policymakers, *IPCC WGI Fourth Assessment Report*, 2007.

Kanakidou, M., et al., Organic aerosol and global climate modelling: A review, *Atmos. Chem. Phys.*, *5*, 1053–1123, 2005.

Kopacz, M., D. Jacob, D. K. Henze, C. L. Heald, D. G. Streets, and Q. Zhang, A comparison of analytical and adjoint Bayesian inversion methods for constraining Asian sources of CO using satellite (MOPITT) measurements of CO columns, submitted.

Kroll, J. H., N. L. Ng, S. M. Murphy, R. C. Flagan, and J. H. Seinfeld, Secondary organic aerosol formation from isoprene photooxidation, *Environ. Sci. Technol.*, *40*(6), 1869–1877, doi:10.1021/es0524301, 2006.

Liao, H., J. H. Seinfeld, P. J. Adams, and L. J. Mickley, Global radiative forcing of coupled tropospheric ozone and aerosols in a unified general circulation model, *J. Geophys. Res.*, *109*(D16), D16207, doi:10.1029/2003JD004456, 2004.

Liao, H., W.-T. Chen, and J. H. Seinfeld, Role of climate change in global predictions of future tropospheric ozone and aerosols, *J. Geophys. Res.*, *111*, D12304, doi:10.1029/2005JD006852, 2006.

Liao, H., D. K. Henze, J. H. Seinfeld, S. Wu, and L. J. Mickley, Biogenic secondary organic aerosol over the United States: Comparison of climatological simulations with observations, *J. Geophys. Res.*, *112*, D06201, doi:10.1029/2006JD007813, 2007.

Lohmann, U., and J. Feichter, Global indirect aerosol effects: A review, *Atmos. Chem. Phys.*, *5*, 715–737, 2005.

Malm, W. C., M. L. Pitchford, M. Scruggs, J. F. Sisler, R. Ames, S. Copeland, K. A. Gebhart, and D. E. Day, Spatial and seasonal patterns and temporal variability of haze and its constituents in the United States: Report III, 0737-5352-47, *Tech. rep.*, Coop. Inst. for Res. in the Atmos., Colo. State Univ., 2000.

Marmur, A., S. K. Park, J. A. Mulholland, P. E. Tolbert, and A. G. Russell, Source apportionment of $PM_{2.5}$ in the southeastern United States using receptor and emissions-based models: Conceptual differences and implications for time-series health studies, *Atmos. Environ.*, *40*(14), 2533–2551, 2006.

Murazaki, K., and P. Hess, How does climate change contribute to surface ozone change over the United States?, *J. Geophys. Res.*, *111*(D5), 2006.

Park, R. J., D. Jacob, B. D. Field, R. Yantosca, and M. Chin, Natural and transboundary pollution influences on sulfate-nitrate-ammonium aerosols in the United States: Implications for policy, *J. Geophys. Res.*, *109*, D15204, doi:10.1029/2003JD004473, 2004.

Park, R. J., D. J. Jacob, N. Kumar, and R. M. Yantosca, Regional visibility statistics in the United States: Natural and transboundary pollution influences, and implications for the Regional Haze Rule, *Atmos. Environ.*, *40*(28), 5405–5423, 2006.

Pope, C. A., Review: Epidemiological basis for particulate air pollution health standards, *Aerosol. Sci. Tech.*, *32*(1), 4–14, 2000.

Pope, C. A., R. T. Burnett, M. J. Thun, E. E. Calle, D. Krewski, K. Ito, and G. D. Thurston, Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution, *J. A. M. A.*, *287*(9), 1132–1141, 2002.

Sandu, A., W. Liao, G. R. Carmichael, D. K. Henze, and J. H. Seinfeld, Inverse modeling of aerosol dynamics using adjoints: Theoretical and numerical considerations, *Aerosol. Sci. Tech.*, *39*(8), 677–694, 2005.

Schulz, M., et al., Radiative forcing by aerosols as derived from the AeroCom present-day and pre-industrial simulations, *Atmos. Chem. Phys.*, *6*, 5225–5246, 2006.

Stavrakou, T., and J. F. Muller, 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.

Stevens, C. J., N. B. Dise, J. O. Mountford, and D. J. Gowing, Impact of nitrogen deposition on the species richness of grasslands, *Science*, *303*(5665), 1876–1879, 2004.

Stier, P., J. Feichter, S. Kloster, E. Vignati, and J. Wilson, Emission-induced nonlinearities in the global aerosol system: Results from the ECHAM5-HAM aerosol-climate model, *J. Climate*, *19*(16), 3845–3862, 2006.

Tarantola, A., Popper, Bayes and the inverse problem, *Nat. Physics*, *2*(8), 492–494, 2006.

Unger, N., D. Shindell, D. Koch, and D. Streets, Cross influences of ozone and sulfate precursor emissions changes on air quality and climate, *Proc. Natl. Acad. Sci.*, *103*, 4377–4380, doi:doi:10.1073pnas.0508769103, 2006a.

Unger, N., D. T. Shindell, D. M. Koch, M. Amann, J. Cofala, and D. G. Streets, Influences of man-made emissions and climate changes on tropospheric ozone, methane, and sulfate at 2030 from a broad range of possible futures, *J. Geophys. Res.*, *111*(D12), D12313, doi:10.1029/2005JD006518, 2006b.

Volkamer, R., J. L. Jimenez, F. San Martini, K. Dzepina, Q. Zhang, D. Salcedo, L. T. Molina, D. R. Worsnop, and M. J. Molina, Secondary organic aerosol formation from anthropogenic air pollution: Rapid and higher than expected, *Geophys. Res. Lett.*, *33*(17), L17811, doi:10.1029/2006GL02689, 2006.

World Health Organization (WHO), Health aspects of air pollution with particulate matter, ozone, and nitrogen dioxide, Rep. EUR/03/5042688, Bonn, 2003.

Yu, H., et al., A review of measurement-based assessments of the aerosol direct radiative effect and forcing, *Atmos. Chem. Phys.*, *6*, 613–666, 2006.

Zhang, Y., J.-P. Huang, D. K. Henze, and J. H. Seinfeld, The role of isoprene in secondary organic aerosol formation on a regional scale, submitted.

Chapter 2

Inverse Modeling of Aerosol Dynamics: Condensational Growth[1]

2.1   Introduction

In recent years, data assimilation techniques have been used to increase one's ability to predict and characterize atmospheric chemical phenomena by providing valuable estimates of surface emissions, improved model sensitivities, and optimized measurement strategies. By enforcing closure between model predictions and experimental observations, these methods constrain the variance of chemical transport models (CTMs) to produce optimal representations of the state of the atmosphere. As the number of variables used to describe the state of the atmosphere increases, the process of integrating models and measurements becomes increasingly difficult. Fortunately, advances in algorithm efficiency, computational resources, and the theory of inverse modeling have facilitated extension of these techniques to systems of increasing complexity. Anticipating the point at which all main features of sophisticated atmospheric CTMs are endowed with an inverse, this work examines the possibilities of extending data assimilation studies to include explicit consideration of size

---

[1]Henze, D. K., J. H. Seinfeld, W. Liao, A. Sandu, and G. R. Carmichael (2004), Inverse modeling of aerosol dynamics: Condensational Growth, *J. Geophys. Res., 109*, D14201, doi:1029/2004JD004593

and composition aerosol dynamics.

Although the actual implementation of data assimilation methods can be quite different, in general all techniques utilize some observational data set to provide an improved model representation of the system in question. Many previous studies on inverse modeling have utilized the Kalman filter, wherein propagation of the error covariance matrix is used to retain consistency between the model and the measurements (*Lyster et al.*, 1997; *Khattatov et al.*, 2000; *Stajner et al.*, 2001; *Palmer et al.*, 2003a). While using a Kalman filter has the distinct advantage that model error is explicitly included in the analysis, the large computational cost of this approach has historically been the prime motivation for development of alternative methods. As an alternative approach, the adjoint method was first suggested as an efficient technique for performing variational data assimilations in atmospheric transport models by *Marchuk* (1974). Originating from the mathematics of systems optimization and control theory (*Cacuci*, 1981a,b) and well established in the fields of fluid mechanics (*Pironneau*, 1974), meteorology (*Talagrand and Courtier*, 1987) and oceanography (*Tziperman and Thacker*, 1989), the adjoint method has only been applied to CTMs relatively recently (*Fisher and Lary*, 1995; *Elbern et al.*, 1997; *Errera and Fonteyn*, 2001). The treatment, while successful, has been limited to the assimilation and recovery of gas phase species.

The inclusion of detailed aerosol chemistry and physics has become requisite in atmospheric CTMs. Future implementation of 4D-Var assimilation techniques will likewise require the inclusion of aerosols in the adjoint models. To lay the groundwork for this endeavor, the fundamental capabilities (and limitations) of applying such techniques to aerosols need to be investigated. In this paper, we apply the first inverse models of multi-component aerosol dynamics, and evaluate their performance under conditions designed to facilitate incorporation of these routines into existing adjoint CTMs. A paper presenting derivations of the necessary equations for several other forms of inverse aerosol models, and evaluation of these for a simple, single component aerosol has also been submitted (*Sandu*

et al., 2005). These works differ substantially from the only previous data assimilation study involving aerosols (*Collins et al.*, 2001) in that the aerosol distribution is allowed to evolve according to the aerosol dynamic equation (*Pilinis*, 1990) and that the inversion is performed using the adjoint technique. In the study by *Collins et al*, the aerosols were represented as growing via empirical correlations and growth rates, and the total aerosol optical depth was assimilated sequentially using a Kalman filter.

With the above goal in mind, adjoint aerosol models are developed and are tested using simulated observations (commonly known as an identical twin experiment). The (forward) aerosol model used is a simplified, yet numerically and physically consistent, version of the aerosol submodel currently employed in several 4-D CTMs (*Meng et al.*, 1998; *Song and Carmichael*, 2001). As operator splitting is used in such models to isolate all aerosol processes into a single 0-D (box) routine which is called within each cell of the discretized 3-D spatial field, it is sufficient to use a forward box model that does not include gas-phase chemistry or spatial advection. Within this forward box model, emphasis is placed on gas-to-particle conversion, wherein gas-phase transport is the rate-limiting step for particle growth. The details of the forward model are given in Section 2.

An immediate application of an inverse aerosol model is to infer the size distributions of aerosol sources using surface, airborne, or possibly even satellite measurements. This involves reconstructing back trajectories of the distribution by repetitive calls to the adjoint box model from within the overall adjoint 4-D CTM, asking each time to recover the shape of the distribution at a previous time step. Therefore, an important capability of the aerosol adjoint routine is to recover an initial size distribution based upon knowledge of the distribution at some later time(s). The length of the assimilation period will depend upon the temporal resolution of the forward model and the frequency of the observations; herein we consider periods ranging from several minutes to a few hours.

In addition to recovering initial distributions, an inverse aerosol model can be used to estimate physical properties key to the dynamic evolution of the distribution by treating

these quantities as variable parameters. The growth of aerosol particles due to condensation / evaporation is heavily influenced by the thermodynamic properties of the transferring species. A significant fraction of organic aerosol particles are comprised of chemical compounds whose thermodynamic properties in the particulate phase are not well characterized. Better estimates of such properties would not only increase the accuracy of CTMs, but would also aid in interpretation of laboratory studies of aerosol dynamics. Hence another desired capability of an adjoint aerosol model is to provide estimates of the thermodynamic properties of the aerosol species.

The aerosol adjoint models can also help refine experimental measurement strategies. Conditions can be simulated in which either individual species are not measured, or the size distribution is only partially sampled. Comparison of the assimilations between these scenarios leads to sampling schemes that provide an optimum balance between data recoverability and observational burden.

One of the primary reasons for choosing the adjoint method to construct an inverse aerosol model is the computational efficiency of this approach. As variations in the actual implementation of this methodology affect the overall computational requirements, it is beneficial to consider different approaches to constructing the adjoint models, of which there are two generally recognized types—continuous and discrete (*Giles and Pierce*, 2000; *Tziperman and Thacker*, 1989). The first method is to derive the continuous adjoint equations from the governing equations, and then solve these numerically. The second approach is to cast the forward equations into a numerical discretized form, and then take the adjoint of this discretized formula. Numerical discretization and adjoint operations do not commute in general, therefore the continuous and discrete approaches lead to final gradients that differ in accuracy and computational expense; hence, it is desirable to assess both tactics when introducing the adjoint method to a new field (*Sandu et al.*, 2005).

2.2 Multicomponent gas-to-particle conversion (the forward model)

We consider a multicomponent aerosol that is growing/evaporating as a result of gas-to-particle conversion. The continuous governing equation for a 0-D, multicomponent, internally mixed aerosol distribution is then (*Pilinis*, 1990; *Meng et al.*, 1998)

$$\frac{\partial p_i(\mu, t)}{\partial t} = H_i(\mu, p_1, \ldots, p_n, t)p(\mu, t) - \frac{1}{3}\frac{\partial(Hp_i)}{\partial \mu}. \tag{2.1}$$

The boundary conditions are

$$p_i(\mu = \mu_{min}, t) = 0 \ , \quad p_i(\mu = \mu_{max}, t) = 0 \ , \quad p_i(\mu, t = t^0) = p_i^0(\mu) \ ,$$

and the terms are

$$p(\mu, t) = \sum_{i=1}^{n} p_i(\mu, t) \ , \quad H(\mu, p_1, p_2, \cdots, p_n, t) = \sum_{i=1}^{n} H_i(\mu, p_1, p_2, \cdots, p_n, t) \ ,$$

where p is the total mass distribution, p_i is the mass distribution of the ith species, n is the number of species, μ is the log of the particle diameter over a reference diameter, H_i is the condensation/evaporation rate of a single species, and H is the total condensation/evaporation rate. H_i is given by the expression (*Wexler and Seinfeld*, 1990)

$$H_i = \frac{1}{m}\frac{dm_i}{dt} = \frac{2\pi D_p D_i}{m(1 + \frac{2\ell}{\alpha D_p})}(g_i - c_i), \tag{2.2}$$

where D_p is the diameter of the aerosol particle, D_i is the molecular diffusivity of species i in air, m_i is the mass of species i in a particle of diameter D_p, m is the total mass of the particle, ℓ is the mean free path, α is the sticking coefficient, g_i is the concentration of species i in the gas phase and c_i is the surface concentration of species i.

To solve (2.1), the aerosol distribution is discretized using a sectional approach (*Gelbard*

and Seinfeld, 1980; *Gelbard et al.*, 1980). The discrete form of the equation is solved using operator splitting techniques (*Yanenko*, 1971) and a modified Bott advection scheme (*Bott*, 1989; *Dhaniyala and Wexler*, 1996) in which the growth term is calculated before the advection term in order to avoid particles being left behind in the lower bins (*Dabdub and Seinfeld*, 1994; *Zhang et al.*, 1999).

2.3 The inverse problem

The goal of inverse modeling is to estimate model parameters which, when implemented in the forward model, yield solutions that are in optimal agreement with a set of observational data. The first step is to calculate a trial solution of the forward model (2.1) using a background (first guess) value for the model parameters, χ. The discrepancy between the trial solution and what is known from observations is measured by the cost function, which can be represented in general form as

$$\mathcal{J}(p_i, \chi) = \int_{t^0}^{T} \int_{\mu_{min}}^{\infty} J_0(p_i(\mu, t)) d\mu dt. \tag{2.3}$$

More specifically, for data assimilation problems, the cost function $\mathcal{J}$ is given as

$$\mathcal{J}(p_i, \chi) = \frac{1}{2}(\chi - \chi_b)^T B^{-1}(\chi - \chi_b) + \frac{1}{2}\sum_i^n \sum_{k \in \Omega} (y^k - h(p_i^k))^T R_k^{-1}(y^k - h(p_i^k)). \tag{2.4}$$

where Ω is the set of discrete time points t^k for which data are known, y^k are the observations at time t^k, h maps the solution from the model space to the observational space, χ_b is the apriori (background) estimate of χ, the matrix B is the error covariance associated with the background term, and the R_k are error covariances of the observations. The optimal model solution and parameters are found by solving the minimization problem

$$\min_{\chi} \mathcal{J}(p_i, \chi),$$

where $\mathcal{J}_{min}$ is found using the gradient resulting from taking the derivative of (2.3) with respect to χ. The difficulty lies in the fact that there is typically no single equation relating the model parameters to the model solution, as $\mathcal{J}$ depends on χ implicitly through the dependency of p_i on χ given by the forward model. In order to determine $\nabla_\chi \mathcal{J}$, an inverse model must be constructed which can calculate the derivative of the forward solution with respect to the model parameters.

2.3.1 The adjoint method

The adjoint method uses a single backward integration of the model (with the state variable during the backward integration being the derivative of the cost function with respect to the original forward state variables) from the final time to the initial conditions in order to determine all elements of the gradient simultaneously. Compared to forward sensitivity analysis (*Hoffman*, 1986), in which the gradient is determined by consecutively propagating perturbations of each parameter individually through the model, the dependence of the calculation's complexity on the number of variable parameters is greatly reduced (*Talagrand and Courtier*, 1987). Not only does this approach afford application to detailed models, it also facilitates the simultaneous estimation of large numbers of parameters. One drawback to the adjoint approach is that for nonlinear problems, trajectories from the forward integration must be available for the backward integration. This leads to large storage requirements; however, multiple level checkpointing schemes can be implemented to reduce this demand. A limitation of the adjoint method itself is that estimates from the solution of the inverse problem are subject to the same systematic and random errors present in the forward model. Unlike the Kalman filter approach, these factors can not be treated explicitly. Although the method can be used to improve systematic error induced by model parameters, sound application is limited to models for which random errors in the forward solution are small, or at least well characterized.

In the following subsections, we give the equations for $\nabla_\chi \mathcal{J}$ derived using both the

continuous and discrete adjoint methods. While there is no formal advantage of one method over another in any general sense, one approach may be better suited to a given application. Typically, the discrete approach yields analytical gradients by implementing in reverse order the exact numerical code used to calculate the forward model, thereby capturing the variable dependencies and nonlinearities that are included in the discretized forward model. Furthermore, if the governing equation is solved using an explicit numerical algorithm, it can be possible to generate the discrete adjoint codes easily and quickly using automatic differentiation software. Alternatively, to derive the continuous adjoint equations by hand, one must linearize the equations first, leading to gradients that can be highly approximate. On the other hand, deriving the continuous adjoint equations often provides insight into the physical meanings of the adjoint variables and boundary conditions, and the solution to these equations can usually be implemented more efficiently than automatically generated adjoints of the discretized model.

We present the continuous adjoint equation first. Then we consider the adjoint of the discretized governing equation as is generated by the Tangent Adjoint Model Compiler (TAMC) (*Giering and Kaminski*, 1998). In Section 4 we compare the results of each approach using a sample system representative of atmospheric aerosols.

2.3.2   Continuous adjoint equations

For the continuous adjoint equations, we consider the case where the model parameters are simply the initial distributions of each species,

$$\chi = p_i(\mu, 0) = p_i^0.$$

The equation adjoint to (2.1) is

$$\frac{\partial \lambda_i}{\partial t} = -\sum_{j=1}^{n} \lambda_j H_j - p \sum_{j=1}^{n} \lambda_j \frac{\partial H_j}{\partial p_i} - \frac{1}{3} \sum_{j=1}^{n} p_j \frac{\partial \lambda_j}{\partial \mu} \frac{\partial H}{\partial p_i} - \frac{H}{3} \frac{\partial \lambda_i}{\partial \mu} - \frac{\partial J_0}{\partial p_i}, \qquad (2.5)$$

the derivation of which is given in Appendix A. The adjoint equation is integrated backward in time from the "initial conditions"

$$\lambda(\mu, T) = 0$$

to the "final conditions"

$$\lambda_i(\mu, t^0) = \nabla_{p_i^0} \mathcal{J} \tag{2.6}$$

to solve for the adjoint variable $\lambda(\mu, t)$ at $t = 0$, which we see from (2.6) is the gradient of the cost function with respect to the initial distribution.

Although we have derived the adjoint equation (2.5) in continuous form, the continuous method is, in practice, still a hybrid of continuous and discrete calculations. The nonlinear dependence of H upon $p_i(\mu, t)$ for growth laws such as that given by (2.2) makes the $\frac{\partial H}{\partial p_i}$ term of the adjoint equation (2.5) difficult to evaluate using continuous equations; therefore, automatic differentiation is used to calculate this term. (This nonlinearity also makes it difficult to distinguish between those variations in H caused by variations of parameters within the growth law, and those caused by variations in $p_i(\mu, t)$, which is why we have limited the scope of the continuous analysis to $\chi = p_i^0$.) In addition, both continuous forward and adjoint equations are eventually integrated numerically, further blurring the distinction between the continuous and discrete approaches.

2.3.3 Discrete adjoint equations

In this section we explicitly derive the discrete adjoint formulas to illustrate the differences between the continuous and discrete approaches. The actual formulas used were created automatically using TAMC. A complete explanation of the theory and algorithms used in TAMC is given by *Giering and Kaminski* (1998).

We begin with a discretized form of the governing equation, which we shall represent

below as

$$[p_i]_j^k = F_j(p_i^{k-1}, g_i^{k-1}) , \quad k = 1, \ldots, N , \quad i = 1, \ldots, n , \quad j = 1, \ldots, s . \qquad (2.7)$$

where $[p_i]_j^k$ is the concentration of species i in the jth bin at time step k, p_i^k is the vector of all particulate concentrations, g_i^k is the vector of all gas concentrations, and F_j represents the numerical operator describing gas / particle transport and advection in diameter space. An informative example to consider is when the observations are simply the concentrations at the final time step, and the only recoverable parameters are the initial conditions. In this case, $\Omega = \{N\}$, h is simply an identity, and, ignoring background terms, the cost function can be written as

$$\mathcal{J}(p_i^0) = \frac{1}{2} \sum_{j=1}^{s} \sum_{i=1}^{n} ([y_i]_j^N - [p_i]_j^N)^T R_N^{-1} ([y_i]_j^N - [p_i]_j^N). \qquad (2.8)$$

The desired quantity to be computed is the derivative of the cost function with respect to changes in the vector of initial conditions,

$$\nabla_{p_i^0} \mathcal{J} = \frac{\partial \mathcal{J}(p_i^N)}{\partial p_i^0}. \qquad (2.9)$$

Using the chain rule (in its transposed form), one can expand the RHS of (2.9)

$$\nabla_{p_i^0} \mathcal{J} = \left[\frac{\partial p_i^1}{\partial p_i^0} \right]^T \left[\frac{\partial p_i^2}{\partial p_i^1} \right]^T \cdots \left[\frac{\partial p_i^N}{\partial p_i^{N-1}} \right]^T \left[\frac{\partial \mathcal{J}(p_i^N)}{\partial p_i^N} \right] \qquad (2.10)$$

Evaluation of the RHS of (2.10) from right to the left corresponds to calculating $\nabla_{p_i^0} \mathcal{J}$ via the adjoint method, while calculating this series of matrix products from left to right constitutes a forward sensitivity calculation. Careful consideration of the number of required scalar multiplications shows that the computational demands of the adjoint method are significantly less than those of the forward method when the dimension of $\mathcal{J}$ is smaller

than the dimension of p (*Kaminski et al.*, 1999; *Sandu et al.*, 2003). Since in this case $\mathcal{J}$ is a scalar and p has $n \times s$ elements, calculating this series of matrix products in reverse is preferable.

Defining the discrete adjoint variable as

$$\lambda^k = \left[\frac{\partial p_i^N}{\partial p_i^k}\right]^T \left[\frac{\partial \mathcal{J}(p_i^N)}{\partial p_i^N}\right] = \nabla_{p_i^k} \mathcal{J} \tag{2.11}$$

and initializing λ^k as $\lambda^N = \nabla_{p_i^N}\mathcal{J}$, $\lambda^0 = \nabla_{p_i^0}\mathcal{J}$ can be found iteratively (beginning with $k = N$ and ending with $k = 1$) using the following expression

$$\lambda^{k-1} = \left[\frac{\partial p_i^k}{\partial p_i^{k-1}}\right]^T \lambda^k. \tag{2.12}$$

In this manner, the adjoint method is reduced to calculating $\frac{\partial p_i^k}{\partial p_i^{k-1}} = \frac{\partial F_j(p_i^k, g_i^k)}{\partial p_i^{k-1}}$ at each step. $F_j(p_i^k, g_i^k)$ is implemented using standard FORTRAN constructs such as loops, conditionals, basic functions and algebraic manipulations, for which algorithms for calculating the derivatives are known (*Giering and Kaminski*, 1998; *Giles et al.*, 2003), hence the adjoint code can be constructed automatically. One potentially problematic routine in $F_j(p_i^k, g_i^k)$ is the Bott-advection scheme: the positive-definite constraints contain many evaluations of min / max statements, whose derivatives are undefined if the arguments are equal. To avoid this problem, we use double precision floating point numbers and resign ourselves to arbitrarily choosing the path of dependence in the rare case that the arguments are exactly equal.

Due to the nonlinear nature of $F_j(p_i^k, g_i^k)$ introduced by the dynamic time step and nonlinearities in the growth law, $\frac{\partial F_j(p_i^k, g_i^k)}{\partial p_i^{k-1}}$ will depend upon p_i^k and g_i^k, hence their values from the forward trajectories will be required at each step of the iteration. This can lead to significant storage requirements and read / write demands for full-scale models with many components in many cells. Similar situations have been handled gracefully by checkpointing

schemes which minimize these types of computational demands (for example *Elbern and Schmidt* (1999), or the distributed scheme implemented for a parallel model of *Sandu et al.* (submitted 2003)); these techniques could be applied to the aerosol adjoint model as well.

2.4   Inverse modeling of aerosol size-composition dynamics

In order to assess the various adjoint models, we perform multiple twin experiments on a test system that consists of three species whose properties are designed to be representative of conditions commonly encountered in atmospheric aerosols. Observations are sampled from the reference, or true, solution generated using the forward model. The simulation is repeated with perturbed values of the parameters, and the reference values are recovered through inverse modeling. The adjoint method is used to calculate the gradient of the cost function with respect to the initial distributions and/or pure species vapor concentrations. The cost function is then minimized using a limited memory BFGS algorithm (*Byrd et al.*, 1995; *Zhu et al.*, 1994), providing optimized estimates of the desired quantities.

To simplify the calculations, the components of the test system are assumed to have ideal thermodynamic properties. Ignoring surface tension and non-ideal effects, Raoults law and the ideal gas equation can be used to express the surface vapor concentration as a function of the particle composition,

$$c_i = x_i c_i^\circ$$

where x_i is the aerosol phase mole fraction and c_i° is the pure component vapor concentration of species i. If we assume, for simplicity, that each species has equal molecular mass, then the fractions are equivalent to the mass fractions, and the growth rate can be written as

$$H_i = \frac{2\pi D_p D_i}{m(1 + \frac{2\ell}{\alpha D_p})}(g_i - \frac{p_i}{\sum_{i=1}^{n} p_i} c_i^\circ) \tag{2.13}$$

The initial conditions for the reference (true) solution used throughout this study are given in Table 2.1, and the physical properties of the aerosols are $\alpha = 0.1$, $\ell = 65$ nm, and $D_i = 1 \times 10^{-5}$ m^2/s. In the aerosol phase, each species is initially log-normally distributed: species 1 is located in the smaller bins, species 2 in the larger bins, and species 3 across all bins. The gas-phase concentrations and pure component vapor concentrations are selected such that species 1 condenses and species 2 evaporates, while the third species is nonvolatile. Frames (a) - (c) of Figure 2.1 show the reference run at $t = 0$, 15 min, and 2.5 h, respectively. Most of the progress towards an equilibrium distribution is made during the first 15 min. Frame (d) shows the time evolution of the gas phase concentrations. Species 1 condenses before species 2 evaporates because gas / particle transport takes longer for the larger particles. The initial decrease in the vapor concentration of species 2 occurs because its mole fraction is very low in the smaller particles, causing the effective surface vapor concentration for these particles to be lower than the surrounding gas concentration.

For use with the discrete adjoint model, the time step for the forward numerical simulation is adjusted dynamically to be as long as possible while still meeting the following criteria: it always satisfies the Courant stability condition, and it is sufficiently small to justify operator splitting. After an initial brief period during which most of species 1 condenses, the time step levels off to a value of ~ 18 s, leading to a simulation in which 50 steps span ~ 15 min.

The continuous adjoint equation (2.5) for the forward model is solved using finite differences. Due to the nonlinearity of (2.1), solving the adjoint equation requires values from the forward solution. Rather than allow each integration to have a different time stepping scheme, and then attempt to match the trajectories by interpolating, it is preferable to use a static time step for both forward and backward runs. In order to avoid the possibility of either solution becoming unstable, the time step is fixed at 5.0 s. Consequently, the number of time steps required to run the continuous model is almost four times greater than that required to run the discrete model.

Multiple assimilation studies were performed using the test system described above. The studies were grouped into four scenarios according to how much information was initially known and how observations were used to recover the unknown data. As the primary interest was investigation of formulation of the inverse modeling problem, we did not explore variations in the complexity of the aerosol distribution in order to keep the forward model consistent from case to case. Discrete adjoint codes were generated using TAMC for each scenario. Reconstructing the adjoint model for each set of dependent and independent variables did not present a major challenge, as the calculation of an adjoint model of this system using TAMC takes less than a few minutes.

Table 2.3 summarizes the conditions and results of each of the cases considered. The RECOVER column lists which parameters were being assimilated; the numbers refer to species whose initial distribution (p_i^0) or pure surface concentrations (c_i°) were unknown. The initial guesses for these unknown parameters are given in the GUESS column. The notation $\times(a, b, c)$ indicates that the initial guess was equal to the true value multiplied by a factor of a, b, c for the 1st, 2nd and 3rd species, respectively, while $+(a, b, c)$ implies that the true values were amended by these amounts. The extent to which details of the reference solution were included as observations is summarized by the three columns under the OBSERVE heading. The numbers in the bin column indicate which of the bins were observed (terms like $\overline{12}$ indicate that only the total concentration in bins 1 through 2 was known), and the numbers in the species column indicate which species were measured. The ratio in the time column is the time between observations over the total simulation time (both in minutes). The R column gives the results of each test. A scalar measure of the relative success of the data assimilation is the percent of the error in the initial guess that is still present after optimization,

$$R(z) = \left[\frac{\sum_{bins/species} (z_{optimized} - z_{true})^2}{\sum_{bins/species} (z_{guess} - z_{true})^2} \right]^{1/2} \tag{2.14}$$

where z is either p_i^0 or c_i°. Low values of R imply that either the initial guess was extremely bad or the assimilation converged to the true value.

As the entire assimilation procedure depends critically upon the minimization of $\mathcal{J}$, it is worth digressing momentarily to discuss some features of the cost function that arise in inverse aerosol modeling. Consider the full cost function given in (2.4). Rigorous treatment of the cost function for the test problem would require generation of fictitious error covariance such that R_k and B can be defined. However, realistic values of R_k and B will be highly case dependent in any assimilation involving real data, hence they will be implemented herein less formally in order to focus on construction of the adjoint model in general. Within the twin experiment framework, the observations can be considered to be exact and independent, hence R_k reduces to the identity matrix. For assimilation cases in which we limit ourselves to observations in only a subset of the species or bins, the corresponding diagonal element of R_k will be zero. Since all the weight factors are then either zero or one, this can be equivalently represented by writing the summations in (2.4) over only the observed species / bins. In most cases considered, the cost function does not penalize departure from the background estimates since we know that the observations are correct while the initial guesses are wrong. Leaving out the first term of (2.4) is equivalent to letting B go to ∞. Exceptions to this arise in Case 4, where we have reason to believe that the background estimate of the initial distribution is functionally more appropriate than the converged solution. For such cases, which are, in general, underdetermined, preconditioning of the cost function (i.e. including the penalty term $\frac{1}{2}(\chi - \chi_b)^T B^{-1}(\chi - \chi_b)$) may be appropriate.

Finally, let us consider issues that arise for real aerosol inverse modeling. Even for inverse modeling studies of real systems, R_k and B are commonly taken to be diagonal (*Mendoza-Dominguez and Russell*, 2000, 2001). Furthermore, it is often assumed that all elements of B are equal so that the entire matrix can be characterized by a single parameter, the so called ridge regression parameter. Much attention has been given to methods for

estimating the optimal value of this parameter. For aerosol inverse modeling these assumptions may not be valid. Significant observational error covariance will exist between species which are not measured independently, but are inferred on the basis of charge equilibrium (for example, nitrate concentrations are often inferred from the measured amounts of sulfate and ammonium). Furthermore, it will be likely that the background terms for some species (for example sulfates) will be known with relatively small variance, while others will have very large variance (SOA), hence B will likely not be simply a scalar multiple of the identity matrix. Overall, inverse aerosol problems are likely to be ill-conditioned due to the model resolution in the size domain being much more refined than the observational resolution. One possible alternative which avoids having to introduce additional bias via R_k and B is to simply halt the optimization process before the cost function is completely minimized, as conjugate gradient methods will minimize along the largest regular vectors first.

2.4.1   Case 1: Recovery of initial distributions

The most important aspect of the data assimilation is the ability to recover the initial distribution, as determination of other parameters is dependent upon the adjoint of the concentration variable. Case 1a is the easiest test, with all 3 species being measured in all 8 bins and all the surface concentrations considered known. Cases 1a-c.i used the discrete adjoint model while cases 1a-c.ii used the continuous adjoint model. The reference, guessed, and optimized initial distributions for cases 1a.i and 1a.ii are shown in Figure 2.2. Both adjoint models recover the true distribution very well, and the continuous model converges more completely than the discrete model in this case. Considering a longer assimilation period (40 min), yet still only making an observation at the final time, the results of Case 1b.i and 1b.ii (given in Table 2.3 but not plotted) show that in this situation the discrete model optimizes to a more accurate set of initial distributions. In Case 1c, the simulation time is 2.5 h, but observations are still taken every $\sim$ 15 min. Figure 2.3 shows

that the optimized p_i^0 are greatly improved over the initial guess, yet still noticeably far from the true distribution. Overall, when the interval between consecutive observations is relatively short ($\sim$15 min), the continuous method provides better estimates than the discrete method; however, the opposite becomes true as the distribution of observations becomes increasingly sparse. Given only a single observation over a period of 2.5 h, the discrete model performs much better than the continuous model (Case 1d, see Figure 2.4).

While even the longer assimilation periods considered here are much shorter (temporally) than can be expected for assimilations involving actual data and real species, this is only an artifact of the arbitrary environmental conditions used for this test case. A more relevant (and general) measure of the assimilation period is the number of numerical integration steps taken between observations. Examining assimilation intervals of 50, 150 and 500 time steps over a length of up to 500 steps covers a wide range of potential models and sets of observational data. For example, a local urban aerosol model that is run for a few days typically employs time steps on the order of minutes and is compared to observations taken during intervals on the order of hours. For large scale regional models that are run for months, time steps are typically on the order of hours and observation intervals on the order of days.

In order to test the validity of the tangent linear approximations inherent in the adjoint model over the assimilation period, the gradient was also calculated using finite differences with a perturbation of 10^{-9}. Figure 2.5 shows the relative reduction in the cost function after the first optimization step, $\Delta = \frac{\mathcal{J}^0 - \mathcal{J}^1}{\mathcal{J}^0} \times 100\%$, as a function of the total number of steps in the assimilation period. The adjoint gradient becomes increasingly inaccurate beyond ~ 250 steps. As the aerosol distribution approaches equilibrium, the assimilation becomes increasingly difficult.

In addition to comparing the ability of the two types of adjoint models to recover the initial distributions, it is important to compare the computational expense of each approach. The total optimization expense ratio is η_{tot}, where

$$\eta_{tot} = \frac{\text{Total computational time (discrete)}}{\text{Total computational time (continuous)}} \tag{2.15}$$

Let t_f be the time for the forward calculation, t_b be the time for the backward calculation, and N_J be the number of cost function evaluations during minimization. Noting that the total computational time for each test is approximately equal to $N_J * (t_f + t_b)$, this ratio can be further broken down into a product of ratios which are fairly consistent in magnitude throughout each test, and whose smallness indicates the degree to which the discrete calculation is preferable.

$$\eta_{tot} = \eta_J \eta_f \eta_b \tag{2.16}$$

where

$$\eta_J = \frac{N_J(d)}{N_J(c)}$$

$$\eta_f = \frac{t_f(d)}{t_f(c)}$$

$$\eta_b = \frac{\left(1 + \frac{t_b(d)}{t_f(d)}\right)}{\left(1 + \frac{t_b(c)}{t_f(c)}\right)}$$

The values of each ratio are given in Table 2.2. Considering t_b/t_f to be a measure of the efficiency of the backward calculation with respect to the forward calculation, the large values of η_b indicate that the backward calculation is much more efficient for the continuous model than the discrete model. However, as indicated by η_J, the gradients from the continuous model are not as accurate as those from the discrete model. Both these results are consistent with what one would expect from these two types of models. Simplifications made to derive the adjoint equations in continuous form lead to faster calculations that are more approximate in nature.

In addition to analyzing the fundamental capabilities of the adjoint method in this test system, we would like to make recommendations for the direction of future work involving

more sophisticated aerosol models. As the complexity of the model increases, a continuous derivation will require an increasingly large number of approximations, leading to adjoint times that are faster, yet gradients that are not as accurate; hence, we speculate that η_f will decrease and η_b will increase. If, to a first order, these effects cancel each other out, the overall efficiency of a more complex aerosol model will depend upon η_f. In this simple model, η_f is $\sim 1/4$ because the average time step taken in the discrete model is about four times as long as the static time step set in the continuous model. For detailed aerosol models, the range of the dynamic time step can span several orders of magnitude. Using a static time step will force the forward calculation for the continuous model to be much slower than the forward calculation for the discrete model, causing η_f, and likely η_{tot}, to be less than unity by several orders of magnitude. To avoid this, one could use dynamic time steps for both forward and backward runs of the continuous model; however, the interpolation process required to utililize data from the forward trajectory when solving the adjoint equation may increase the error in the resulting gradient. While there are no inherent restrictions on the types of time steps that can be used to solve the continuous equations, these issues can complicate their implementation. In short, the discrete adjoint formulation appears to be the more viable method.

2.4.2 Case 2: Recovery of pure species vapor concentrations

The next set of tests examines the situation in which the initial distributions of all the components are known, but the pure component surface vapor concentrations are not. The value of $R(c_i^\circ)$ for Case 2a is 0.00 because the true values of c_i° are recovered to at least six significant digits. For example, the optimized value of c_1° is 1.0000028. Case 2b considers the situation in which the initial guesses for c_i° are such that the overall transport of each species is in the opposite direction than in the true solution. For example, with c_1° = 20 μg/m^3, species 1 evaporates instead of condensing. Again, the optimized c_i° matches the true value to at least six significant digits, indicating that c_i° can be recovered even

when the overall direction of the mass transport is not known before the initial analysis.

2.4.3   Case 3: Recovery of initial distribution and vapor concentrations

The third scenario addresses a common question encountered in aerosol measurement —
based upon accurate information of a subset of the aerosol components, what can be
inferred about an unmeasured species? In this set, no information about species 1 is used
in performing the assimilation, and the cost function is

$$\mathcal{J}(p_i^0) = \frac{1}{2} \sum_{j=1}^{s} \sum_{i=2}^{n} ([y_i]_j^N - [p_i]_j^N)^2.$$

Results for Case 3a indicate that both p_1^0 and c_1° can be recovered simultaneously. While
these results look promising, to say that "nothing" was known about species 1 is perhaps
misleading in that the initial guess for p_1^0 had the same shape as the true solution, greatly
facilitating the assimilation. This being said, it is interesting to note that it is not necessary
to precondition the cost function in order to converge to the correct distribution because
the problem is overdetermined in this case.

   To determine how much the success of the assimilation depends upon the shape of the
initial guess, Case 3b starts with p_1^0 being a constant value of 5 μg/m^3 throughout the size
distribution. Not surprisingly, with such a poor initial guess, the performance is drastically
decreased, as indicated by $R(p_1^0) = 0.49$. However, a plot of the initial distribution shows
that the assimilation is very successful for all parameters except the concentrations in the
two largest size bins (Figure 2.6(i)). To understand why this would be the case, it is useful
to recall that the driving term for the discrete adjoint model is $\frac{\partial \mathcal{J}(p_i^N)}{\partial p_i^N}$. In other words,
the adjoint model is forced by the difference in the concentration of the observed species
between the guessed and the reference solutions at the time when the observations were
made. For Case 3b, the simulation results at $t = 15$ min are shown in Figure 2.6(ii), and we
see optimization of p_1^0 in bins 7 and 8 was stopped prematurely because there was no longer

any driving force for the adjoint model; the optimized solution had already converged to the true value. Since the characteristic time for condensation / evaporation in bin 7 is several hours, the concentrations in the larger bins had yet to change significantly after only 15 min. In this situation, as confirmed by the results of Case 3c, it is advantageous to run the simulation longer before taking an observation in order to provide ample forcing for the adjoint model. On the other hand, if the observation time is delayed too long, the assimilation would become impossible (imagine trying to determine the initial condition for an aerosol that has equilibrated to an evenly distributed profile), as indicated by Figure 2.5.

2.4.4 Case 4: Recovery from partial distributions

In addition to considering variations in the observation frequency and species detection, it is of interest to examine the performance of the data assimilation when only portions of the size distribution are measured. Scenario 4a addresses the situation in which observations are made only in the smaller four size bins,

$$\mathcal{J}(p_i^0) = \frac{1}{2} \sum_{j=1}^{s/2} \sum_{i=1}^{n} ([y_i]_j^N - [p_i]_j^N)^2$$

Based upon this information, the initial concentrations in the larger bins were determined and are shown in Figure 2.7(a). At first glance, the results appear to be fairly poor; however, one must take into account the direction that each species is advecting. Considering the initial guess as a perturbation of the reference solution, the effect that this perturbation has on the concentrations in the smaller four bins is the driving force for the adjoint model. For species 1, the lower half of the distribution is largely invariant to perturbations in the upper four bins because this component is growing. However, for species 2, particles are evaporating and advection is bringing information about the contents of bins 5-8 to bins 1-4, hence we would expect the assimilation to have performed better for species 2 than

for species 1. Indeed this is the case. Providing further forcing by running the simulation longer also leads to better results (Case 4b), and not surprisingly, if distributions 1 and 3 are considered known, then the assimilation of species 2 is even more improved (Case 4c, Figure 2.7(c)).

Tests 4d-f address cases in which the observed concentrations are actually sums over two or more adjacent size bins. Since the observations are no longer exactly equivalent to the state variables, this averaging is represented by the function h in the cost function,

$$\mathcal{J}(p_i^0) = \frac{1}{2} \sum_{\bar{j}=1}^{2} ([y]_{\bar{j}}^N - h_{\bar{j}}([p]_j^N))^2,$$

where $\bar{j}$ is the index of the lumped bins. In Case 4d, each pair of adjacent bins is averaged, while in 4e the observed distribution is of only two bins—one that contains particles whose diameter is smaller than 2.76 μm, and one that contains particles that are larger. The adjoint method is only able to resolve the initial distributions to a level consistent with the resolution of the initial guess. Given an initial guess that is resolved on the scale of an 8-bin distribution, the assimilations are fairly successful. However, the optimized distributions become increasingly featureless as the resolution of the initial guess is decreased, see Figure 2.8. In order to avoid optimizing to erroneously smooth or jagged distributions, the solution can be constrained by including the penalty term in the cost function. While this approach biases the final estimate, this may be appropriate when there is sufficient information known about the true distribution to quantitatively estimate the error covariance matrix B of the initial guess.

2.5 Conclusions

As part of a broad effort to better the understanding of the state of the atmosphere using inverse modeling techniques, this paper focused on the specific goal of incorporating multicomponent, size resolved aerosols in data assimilation studies. The adjoint method has

been explored as a means of recovering parameters of an aerosol distribution evolving by condensation / evaporation. Within the field of adjoint modeling, we have explored two general tactics for creating the inverse model—discrete and continuous. Evaluating these methods with a simplified, yet representative, model of an atmospheric aerosol, we have attempted to recover parameters of the distribution by assimilating observations that are sparse in time, size and / or chemical resolution.

Intricacies of what was still a simple test model (compared to the aerosol routines implemented in detailed CTMs) limited the feasibility of formulating the adjoint equations in an entirely continuous fashion. In particular, nonlinearities introduced by the particle growth rate limits the extent to which the continuous equations can be derived in full. Nonetheless, the results of problems that have been addressed using the continuous approach are comparable to those found the using discrete approach. However, the flexibility of discrete adjoint models, combined with the ease of creating them automatically using programs such as TAMC, makes them the more viable method for solving inverse problems involving increasingly complex aerosol systems.

In the test problem considered, we attempted to recover parameters such as the initial distribution and the species' pure surface concentrations. Either of these were easily recovered for all three species when at least one observation of the entire distribution was known sufficiently prior to equilibration. Additionally, if both of these properties for a single species were unknown, and this species was never even observed, the adjoint calculations allowed us to adequately infer this information from measurements of the dynamic evolution of the other two species. The most difficult task attempted was the recovery of initial distributions when observations were known in only a subset of the size range, or when the initial estimates were exceptionally poor. For understandable reasons, this type of assimilation required the most observational information in order to yield decent estimates of the aerosol parameters. Overall, we demonstrated that given ample observations and reasonable initial estimates, the adjoint method can be used to recover information about

a dynamic, size and chemically resolved aerosol distribution under a variety of conditions.

Appendix 2.A   Derivation of continuous adjoint equation

We will use the Lagrangian multiplier method to derive the continuous adjoint derivations. The cost function is defined as

$$
\begin{aligned}
\mathcal{J} &= \int_{t^0}^{T} \int_{0}^{\infty} J_0(p_1(\mu, t), p_2(\mu, t), \ldots, p_n(\mu, t)) \, d\mu \, dt \\
&\quad - \sum_{i=1}^{n} \int_{t_0}^{T} \int_{0}^{\infty} \lambda_i(\mu, t)(LHS_{p_i} - RHS_{p_i}) \, d\mu \, dt,
\end{aligned}
\tag{A-17}
$$

Here LHS_{p_i} and RHS_{p_i} refer to the left side and right side of (2.1), respectively. J_0 is the local cost function component,

$$
J_0 = \frac{1}{2} \sum_{i=1}^{n} (y_i - h(p_i))^T R_k^{-1} (y_i - h(p_i)) \delta(t - t^k),
\tag{A-18}
$$

where $t^k \in \Omega$, and Ω is the set of discrete time points t^k for which data are known. Taking the variation of (A-17), we get

$$
\begin{aligned}
\delta \mathcal{J} &= \int_{t^0}^{T} \int_{0}^{\infty} \sum_{i=1}^{n} \frac{\partial J_0}{\partial p_i} \delta p_i(\mu, t) \, d\mu \, dt \\
&\quad - \int_{t^0}^{T} \int_{0}^{\infty} \sum_{i=1}^{n} \delta \lambda_i(\mu, t)(LHS_{p_i} - RHS_{p_i}) \, d\mu \, dt \\
&\quad - \int_{t^0}^{T} \int_{0}^{\infty} \sum_{i=1}^{n} \lambda_i(\mu, t) \delta(LHS_{p_i} - RHS_{p_i}) \, d\mu \, dt
\end{aligned}
\tag{A-19}
$$

Inserting the expressions of LHS_{p_i} and RHS_{p_i}, (A-19) can be written as

$$
\delta \mathcal{J} = \int_{t^0}^{T} \int_{0}^{\infty} \sum_{i=1}^{n} \frac{\partial J_0}{\partial p_i} \delta p_i(\mu, t) \, d\mu \, dt
$$

$$-\int_{t^0}^{T}\int_0^\infty \sum_{i=1}^{n} \delta\lambda_i(\mu,t)(LHS_{p_i}-RHS_{p_i})\,d\mu\,dt$$

$$-\int_{t^0}^{T}\int_0^\infty \sum_{i=1}^{n} \lambda_i(\mu,t)\delta(\frac{\partial p_i(\mu,t)}{\partial t} - H_i(\mu,p_1,p_2,\cdots,p_n,t)p(\mu,t)+\frac{1}{3}\frac{\partial}{\partial\mu}\,(Hp_i))\,d\mu\,dt$$

$$(A\text{-}20)$$

Then we can re-write (A-20) as

$$
\begin{aligned}
\delta\mathcal{J} \;=\; & \sum_{i=1}^{n}\int_{t^0}^{T}\int_0^\infty \delta p_i(\mu,t)\frac{\partial J_0(\mu,p,t)}{\partial p_i}\,d\mu\,dt \\
& -\sum_{i=1}^{n}\int_{t^0}^{T}\int_0^\infty \delta\lambda_i(\mu,t)(LHS_{p_i}-RHS_{p_i})\,d\mu\,dt \\
& -\sum_{i=1}^{n}\int_{t^0}^{T}\int_0^\infty \lambda_i(\mu,t)\frac{\partial(\delta p_i(\mu,t))}{\partial t}\,d\mu\,dt \\
& +\sum_{i=1}^{n}\int_{t^0}^{T}\int_0^\infty \lambda_i(\mu,t)H_i(\mu,p,t)\sum_{j=1}^{n}\delta p_j(\mu,t)\,d\mu\,dt \\
& +\sum_{i=1}^{n}\int_{t^0}^{T}\int_0^\infty \lambda_i(\mu,t)\sum_{j=1}^{n}\frac{\partial H_i(\mu,p,t)}{\partial p_j}\delta p_j p(\mu,t)\,d\mu\,dt \\
& -\frac{1}{3}\sum_{i=1}^{n}\int_{t_0}^{T}\int_0^\infty \lambda_i(\mu,t)\frac{\partial}{\partial\mu}\left(\sum_{j=1}^{n}\frac{\partial H(\mu,p,t)}{\partial p_j}\delta p_j p_i + H(\mu,p,t)\delta p_i\right)d\mu\,dt
\end{aligned}
$$

$$(A\text{-}21)$$

If we choose the final condition $\lambda(\mu,T)=0$, and integrate the third term on the righthand side of equation (A-21) by parts, this term becomes

$$\sum_{i=1}^{n}\int_0^\infty \lambda_i(\mu,t_0)\delta p_i(\mu,t_0)\,d\mu + \sum_{i=1}^{n}\int_{t^0}^{T}\int_0^\infty \delta p_i(\mu,t)\frac{\partial(\lambda_i(\mu,t))}{\partial t}\,d\mu\,dt \qquad (A\text{-}22)$$

Likewise, letting $\lambda_i(0, t) = 0$, $p_i(+\infty, t) = 0$, the sixth term on the right-hand side of equation (A-21) can be written as

$$\frac{1}{3} \sum_{i=1}^{n} \int_{t_0}^{T} \int_{0}^{\infty} \frac{\partial \lambda_i(\mu, t)}{\partial \mu} \left(p_i \sum_{j=1}^{n} \frac{\partial H(\mu, p, t)}{\partial p_j} \delta p_j + H(\mu, p, t) \delta p_i \right) d\mu \, dt \tag{A-23}$$

If $p(\mu, t)$ is the solution of (2.1), $LHS_{p_i} - RHS_{p_i} = 0$, then

$$\begin{aligned}
\delta \mathcal{J} &= \sum_{i=1}^{n} \int_{t^0}^{T} \int_{0}^{\infty} \delta p_i(\mu, t) \frac{\partial J_0(\mu, t)}{\partial p_i} \, d\mu \, dt \\
&+ \sum_{i=1}^{n} \int_{0}^{\infty} \lambda_i(\mu, t^0) \delta p_i(\mu, t^0) \, d\mu \\
&+ \sum_{i=1}^{n} \int_{t^0}^{T} \int_{0}^{\infty} \frac{\partial \lambda_i(\mu, t)}{\partial t} \delta p_i(\mu, t) \, d\mu \, dt \\
&+ \sum_{i=1}^{n} \int_{t^0}^{T} \int_{0}^{\infty} \lambda_i(\mu, t) H_i(\mu, p, t) \sum_{j=1}^{n} \delta p_i(\mu, t) \, d\mu \, dt \\
&+ \sum_{i=1}^{n} \int_{t^0}^{T} \int_{0}^{\infty} \lambda_i(\mu, t) \sum_{j=1}^{n} \frac{\partial H_i(\mu, p, t)}{\partial p_j} p(\mu, t) \delta p_j \, d\mu \, dt \\
&+ \frac{1}{3} \sum_{i=1}^{n} \int_{t_0}^{T} \int_{0}^{\infty} \frac{\partial \lambda_i(\mu, t)}{\partial \mu} \sum_{j=1}^{n} \frac{\partial H(\mu, p, t)}{\partial p_j} \delta p_j p_i \, d\mu \, dt \\
&+ \frac{1}{3} \sum_{i=1}^{n} \int_{t_0}^{T} \int_{0}^{\infty} \frac{\partial \lambda_i(\mu, t)}{\partial \mu} H(\mu, p, t) \delta p_i \, d\mu \, dt
\end{aligned}$$

$$\tag{A-24}$$

Assigning the coefficient in front of δp_i to 0 results in the adjoint equation,

$$\frac{\partial \lambda_i}{\partial t} = -\sum_{j=1}^{n} \lambda_j H_j - p \sum_{j=1}^{n} \lambda_j \frac{\partial H_j}{\partial p_i} - \frac{1}{3} \sum_{j=1}^{n} p_j \frac{\partial \lambda_j}{\partial \mu} \frac{\partial H}{\partial p_i} - \frac{H}{3} \frac{\partial \lambda_i}{\partial \mu} - \frac{\partial J_0}{\partial p_i} \tag{A-25}$$

Bibliography

Bott, A., A positive definite advection scheme obtained by nonlinear renormalization of the advective fluxes, *Mon. Weather Rev.*, *117*, 1006, 1989.

Byrd, R. H., P. Lu, J. Nocedal, and C. Zhu, A limited memory algorithm for bound constrained optimization, *Scientific Computing*, *16*(5), 1190–1208, 1995.

Cacuci, D. G., Sensitivity theory for nonlinear systems. I. Nonlinear functional analysis approach, *J. Math. Phys.*, *22*(12), 2794, 1981a.

Cacuci, D. G., Sensitivity theory for nonlinear systems. II. Extensions to additional classes of responses, *J. Math. Phys.*, *22*(12), 2803, 1981b.

Collins, W., P. Rasch, B. Eaton, B. Khattatov, J.-F. Lamarque, and C. Zender, Simulating aerosols using a chemical transport model with assimilation of satellite aerosol retrievals: Methodology for INDOEX, *J. Geophys. Res.*, *106*(D7), 7313–7336, 2001.

Dabdub, D., and J. H. Seinfeld, Numerical advective schemes used in air quality models–sequential and parallel implementation, *Atmos. Environ.*, *28*(20), 3369–3385, 1994.

Dhaniyala, S., and A. S. Wexler, Numerical schemes to model condensation and evaporation of aerosols, *Atmos. Environ.*, *30*(6), 919–928, 1996.

Elbern, H., and H. Schmidt, A four-dimensional variational chemistry data assimilation scheme for eulerian chemistry transport modeling, *J. Geophys. Res.*, *104*(D15), 18,583–18,598, 1999.

Elbern, H., H. Schmidt, and A. Ebel, Variational data assimilation for tropospheric chemistry modeling, *J. Geophys. Res.*, *102*, 15,967–15,985, 1997.

Errera, Q., and D. Fonteyn, Four-dimensional variational chemical assimilation of crista stratospheric measurements, *J. Geophys. Res.*, *106*(D11), 12,253–12,265, 2001.

Fisher, M., and D. J. Lary, Lagrangian four-dimensional variational data assimilation of chemical species, *Q. J. R. Meteorol. Soc.*, *121*, 1681–1704, 1995.

Gelbard, F., and J. H. Seinfeld, Simulation of multicomponent aerosol dynamics, *J. Colloid Interface Sci.*, *78*(2), 485, 1980.

Gelbard, F., Y. Tambour, and J. H. Seinfeld, Sectional representation for simulating aerosol dynamics, *J. Colloid Interface Sci.*, *76*(2), 541, 1980.

Giering, R., and T. Kaminski, Recipes for adjoint code construction, *ACM Trans. Math. Software*, *24*(4), 437, 1998.

Giles, M. B., and N. A. Pierce, An introduction to the adjoint approach to design, *Flow Turbul. Combust.*, *65*, 393–415, 2000.

Giles, M. B., M. C. Duta, J.-D. Müller, and N. A. Pierce, Algorithm developments for discrete adjoint methods, *AIAA Journal*, *41*(2), 198, 2003.

Hoffman, R. N., A four-dimensional analysis exactly satisfying equations of motion, *Mon. Weather Rev.*, *114*, 338–397, 1986.

Kaminski, T., M. Heimann, and R. Giering, A coarse grid three-dimensional global inverse model of the atmospheric transport, *J. Geophys. Res.*, *104*(D15), 18,535–18,553, 1999.

Khattatov, B., J.-F. Lamarque, L. v. Lyjak, R. Menard, P. F. Levelt, X. X. Tie, G. P. Brasseur, and J. C. Gille, Assimilation of satellite observations of long-lived chemical species in global chemistry-transport models, *J. Geophys. Res.*, *105*(D23), 29,135–29,144, 2000.

40

Lyster, P. M., S. E. Cohn, R. Menard, L.-P. Chang, S.-J. Lin, and R. Olsen, An implementation of a two-dimensional filter for atmospheric chemical constituent assimilation on massively parallel computers, *Mon. Weather Rev.*, *125*, 1674–1686, 1997.

Marchuk, G., Numerical solution of the problems of the dynamics of the atmosphere and the ocean (in Russian), *Gidrometeoizdat*, 1974.

Mendoza-Dominguez, A., and A. G. Russell, Iterative inverse modeling and direct sensitivity analysis of a photochemical air duality model, *Environ. Sci. Technol.*, *34*(23), 4974–4981, 2000.

Mendoza-Dominguez, A., and A. G. Russell, Emission strength validation using four-dimensional data assimilation: application to primary aerosol and precursors to ozone and secondary aerosol, *J. Air & Waste Manage. Assoc*, *51*, 1538–1550, 2001.

Meng, Z., D. Dabdub, and J. H. Seinfeld, Size-resolved and chemically resolved model of atmospheric aerosol dynamics, *J. Geophys. Res.*, *102*(D3), 3419–3435, 1998.

Palmer, P. I., D. J. Jacob, D. B. A. Jones, C. L. Heald, R. M. Yantosca, J. A. Logan, G. W. Sachse, and D. G. Streets, Inverting for emissions of carbon monoxide from Asia using aircraft observations over the western Pacific, *J. Geophys. Res.*, *108*(D21), 2003a.

Pilinis, C., Derivation and numerical solution of the species mass distribution equations for multicomponent particulate systems, *Atmos. Environ.*, *24A*(7), 1923–1928, 1990.

Pironneau, O., On optimum design in fluid mechanics, *J. Fluid Mech*, *64*, 97–110, 1974.

Sandu, A., D. N. Daescu, and G. R. Carmichael, Direct and adjoint sensitivity analysis of chemical kinetic systems with KPP: I—Theory and software tools, *Atmos. Environ.*, *37*, 5083–5096, 2003.

Sandu, A., D. N. Daescu, G. R. Carmichael, and T. Chai, Adjoint sensitivity analysis of regional air quality models, submitted 2003.

Song, C. H., and G. R. Carmichael, A three-dimensional modeling investigation of the evolution processes of dust and sea-salt in East Asia, *J. Geophys. Res.*, *106*, 18,131–18,154, 2001.

Stajner, I., L. Riishojgaard, and R. Rood, The GEOS ozone data assimilation system: specification of error statistics, *Q. J. R. Meteorol. Soc.*, *127*(573), 1069–1094, 2001.

Talagrand, O., and P. Courtier, Variational assimilation of meteorological observations with the adjoint of the vorticity equations. Part I. Theory, *Q. J. R. Meteorol. Soc.*, *113*, 1311–1328, 1987.

Tziperman, E., and W. C. Thacker, An optimal-control / adjoint-equations approach to studying the oceanic general circulation, *J. Phys. Oceanogr.*, *19*, 1471, 1989.

Wexler, A. S., and J. H. Seinfeld, The distribution of ammonium salts among a size and composition dispersed aerosol, *Atmos. Environ.*, *24A*(5), 1231–1246, 1990.

Yanenko, N. N., *The Method of Fractional Steps*, Springer, New York, 1971.

Zhang, Y., C. Seigneur, J. H. Seinfeld, M. Z. Jacobson, and F. S. Binkowski, Simulation of aerosol dynamics: A comparative review of algorithms used in air quality models, *Aerosol Sci. Tech.*, *31*, 487–514, 1999.

Zhu, C., R. H. Byrd, P. Lu, and J. Nocedal, L-BFGS-B: A limited memory FORTRAN code for solving bound constrained optimization problems, *Tech. rep.*, Northwestern University, 1994.
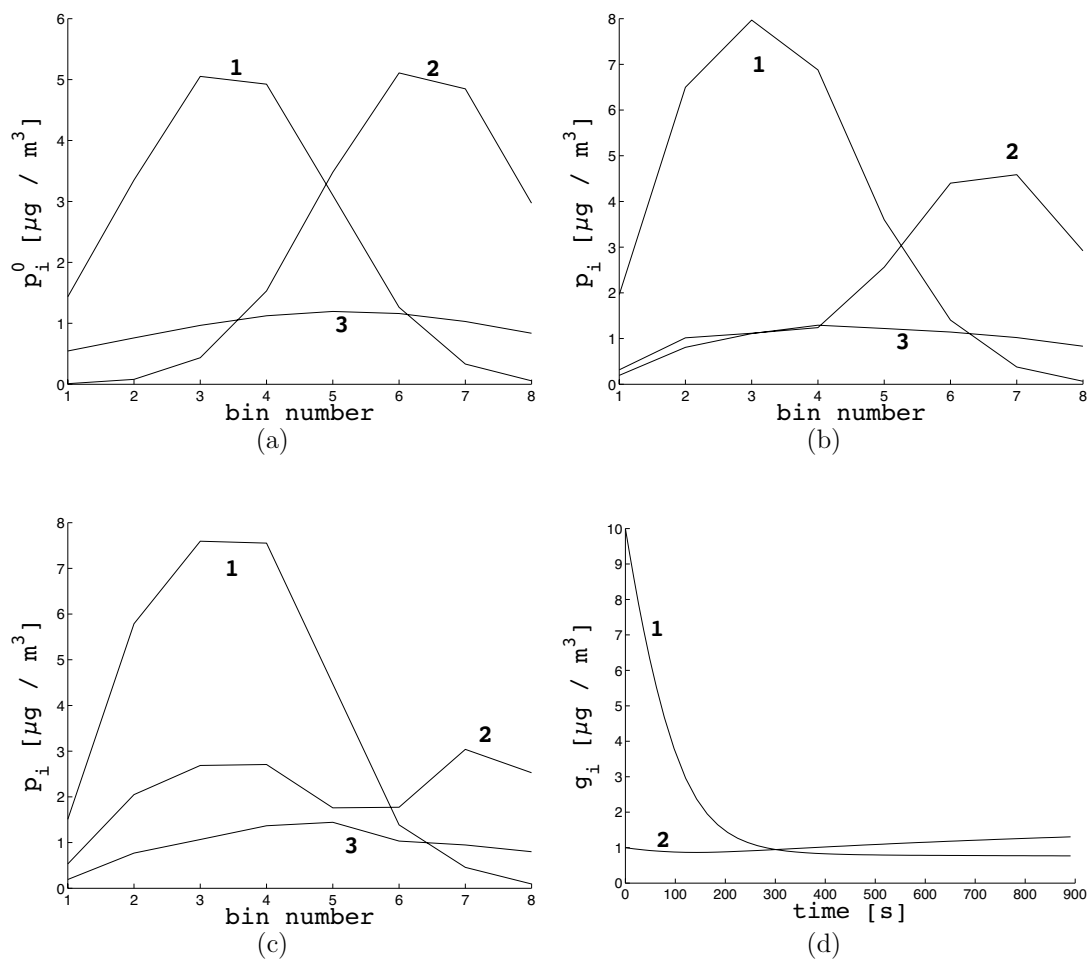
Figure 2.1: The forward model calculation (reference solution). Species 1 is condensing, species 2 is mostly evaporating, and species 3 is inert. Plotted are the aerosol size distributions at t = 0 (a), t = 15 min (b), t = 2.5 h (c), and the gas-phase concentrations as a function of time (d).
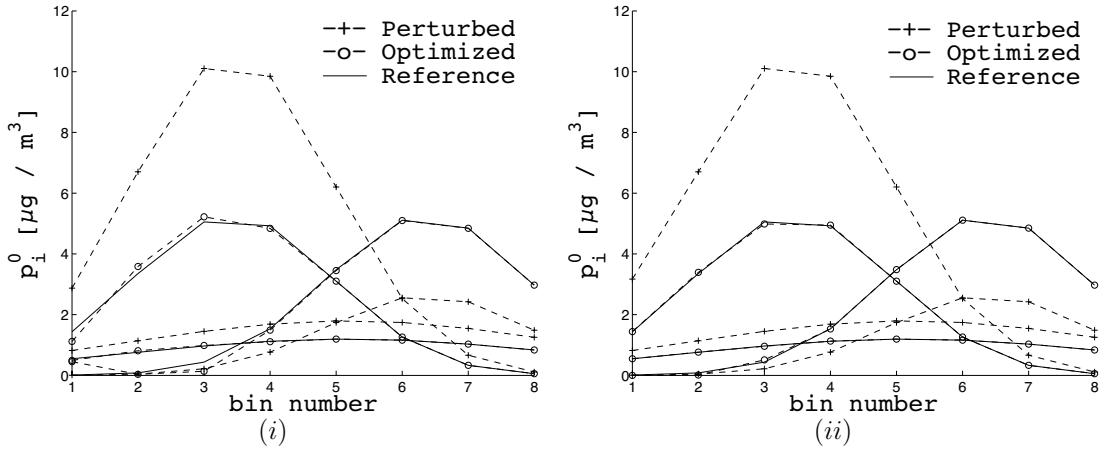
Figure 2.2: Case 1a: Simultaneous recovery of the initial distribution of all three species from an observation at the final time (15 min) using the discrete adjoint model (*i*) and the continuous adjoint model (*ii*). The continuous model performs slightly better, primarily in the lower bins for species 1 and 3.

Figure 2.3: Case 1c: Simultaneous recovery of the initial distributions of all three species from 10 observations taken every 15 min over the course of 2.5 h using the discrete adjoint model (*i*) and the continuous adjoint model (*ii*). Overall performance is similar between the two approaches.
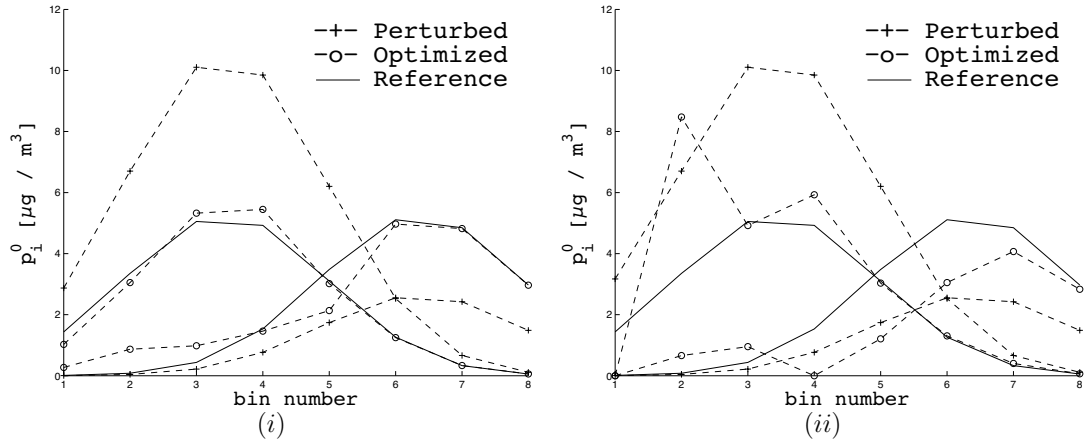
Figure 2.4: Case 1d: Simultaneous recovery of the initial distributions of all three species using only one observation after 2.5 h. Results are shown for the discrete adjoint model (i) and the continuous adjoint model (ii), from which the superior performance of the former for this case is quit evident. Species 3 is omitted from the plots for clarity.
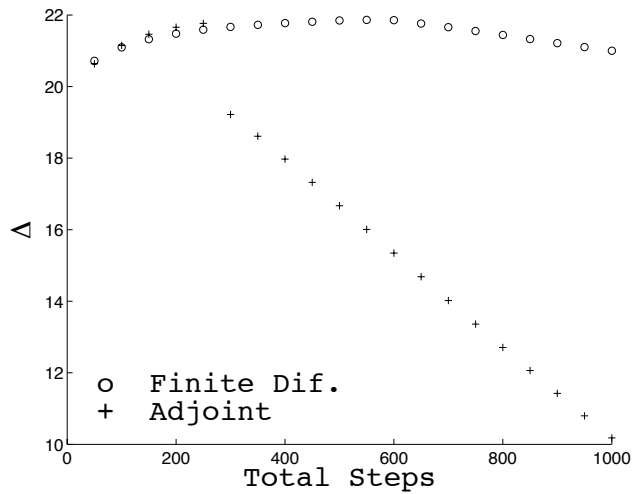
Figure 2.5: Case 1d: Simultaneous recovery of the initial distributions of all three species using only one observation at the final time step (x-axis). Plotted is the relative reduction of the cost function after the first minimization step, $\Delta = \frac{\mathcal{J}^0 - \mathcal{J}^1}{\mathcal{J}^0} \times 100\%$, as a function of the assimilation period. The accuracy of the gradient computed using the adjoint method ($+$) is seen to decay in comparison to that from the finite difference calculation (o) as the distribution approaches equilibrium.
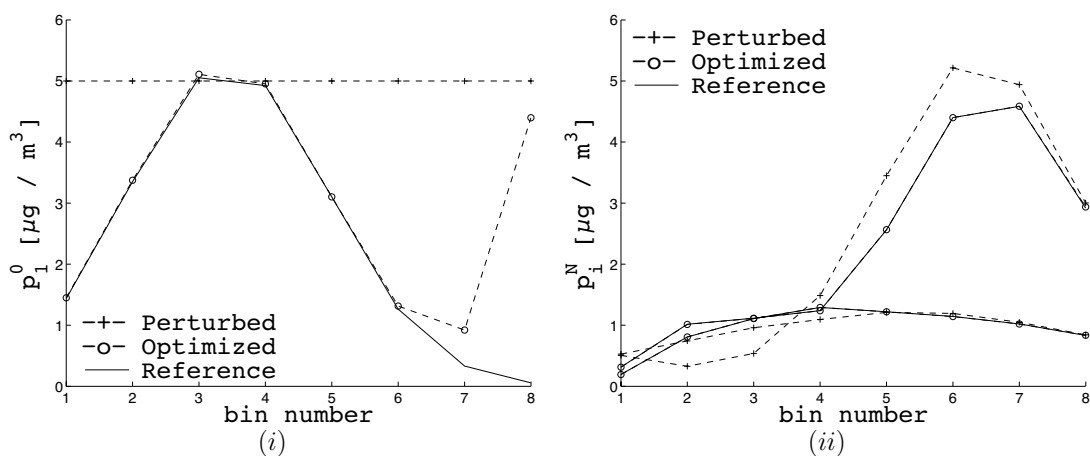
Figure 2.6: Case 3b: Recovering the initial distribution of species 1 from an observation of species 2 and 3 at the final time (15 min). Shown are the aerosol size distributions of species 1 at t = 0 (*i*) and of species 2 and 3 at t = 15 min (*ii*).
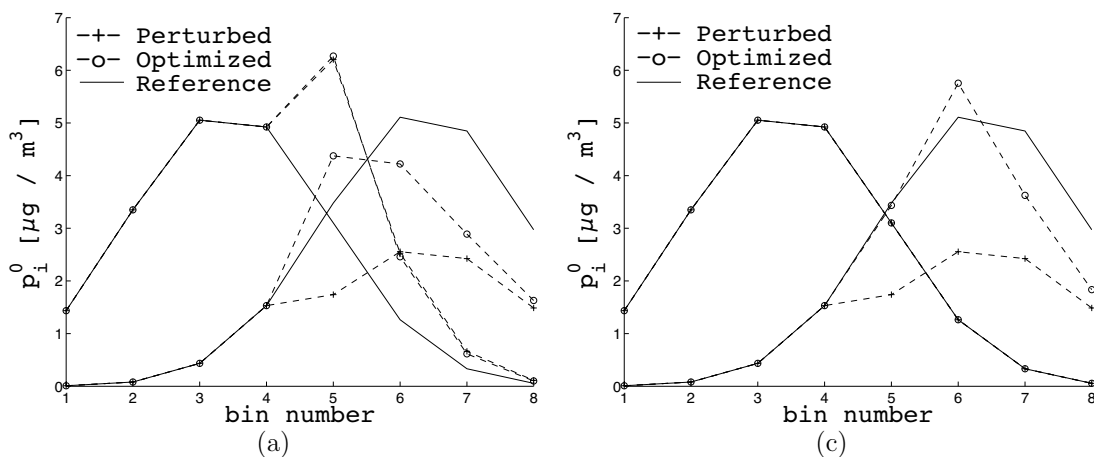
Figure 2.7: Case 4: Recovering initial distributions using only data from the smaller four size bins. The results for Case 4a (recovery of all three initial distributions simultaneously) and Case 4c (recovery of only the initial distribution of species 2) are shown in frames (a) and (c), respectively. Species 3 is omitted from the plots for clarity.
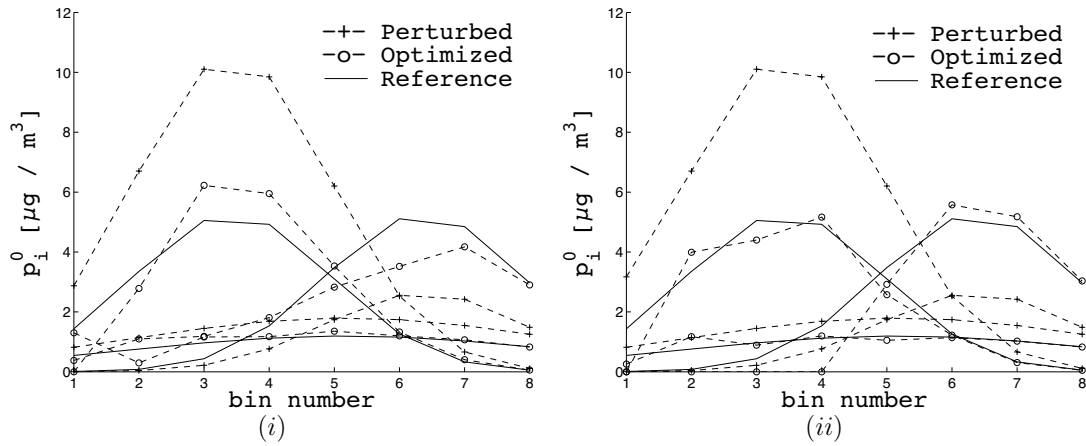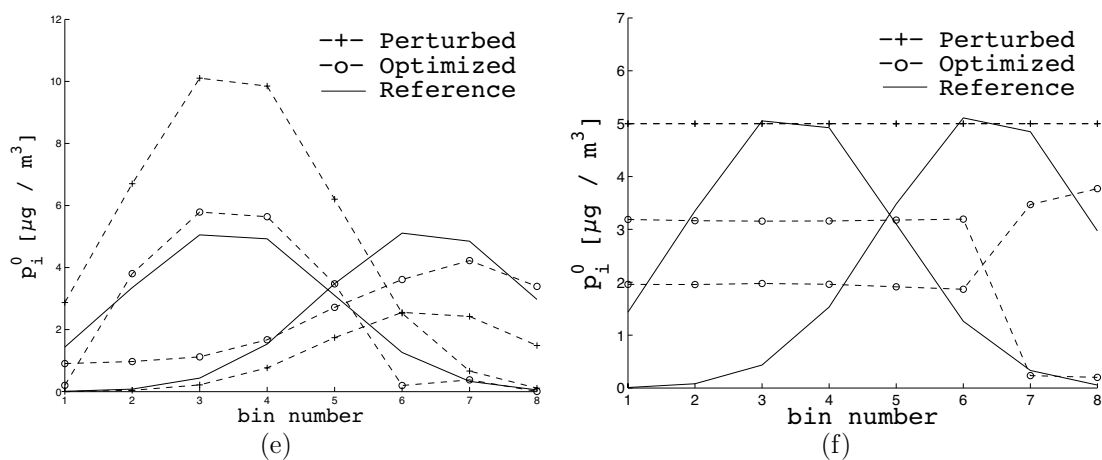
Figure 2.8: Case 4: Simultaneous recovery of the initial distribution of all three species from observations of the total particulate concentration in bins 1 - 6 and bins 7 - 8 using a log normal initial guess and (e) a flat initial guess (f). Species 3 is omitted from the plot for clarity.

Table 2.1: Test problem specifications. Initial gas phase concentrations (g_i), pure component surface vapor concentrations (c_i°), and parameters of the initial log-normal distribution: total concentration (p_i), mean particle diameter ($\bar{D}_p$), and standard devation (σ).

Species	g_i [μg/m^3]	c_i° [μg/m^3]	p_i [μg/m^3]	$\bar{D}_p$ [μm]	σ
1	10.0	1.0	20.0	0.3	2.8
2	1.0	10.0	20.0	2.3	2.8
3	0.0	0.0	10.0	1.0	10.0

Table 2.2: Timing ratios for comparing the discrete to the continuous adjoint model, as defined by eq (2.16). Values less than one indicate the discrete model is preferable.

Case	η_{tot}	η_J	η_f	η_b
1a	2.2	0.8	0.3	8.3
1b	1.4	0.9	0.2	8.3
1c	1.6	0.6	0.3	8.2

Table 2.3: Conditions and results of assimilation tests using notation outlined in section 4

CASE	RECOVER		OBSERVE			GUESS		$\underline{R}$	
	p_i^0	c_i°	bins	species	time	p_i^0	c_i°	p_i^0	c_i°
1a.i	1-3	-	1-8	1-3	15/15	$\times(2,\,0.5,\,1.5)$	-	0.07	-
1a.ii	"	-	"	"	"	"	-	0.01	-
1b.i	"	-	"	"	40/40	"	-	0.19	-
1b.ii	"	-	"	"	"	"	-	0.26	-
1c.i	"	-	"	"	15/150	"	-	0.34	-
1c.ii	"	-	"	"	"	"	-	0.27	-
1d.i	"	-	"	"	150/150	"	-	0.21	-
1d.ii	"	-	"	"	"	"	-	0.68	-
2a	-	1-3	1-8	1-3	15/15	-	$\times 2,\,\times 5,\,+1$	-	0.00
2b	"	"	"	"	"	-	$+20,\,-10,\,+5$	-	0.00
3a	1	1	1-8	2-3	15/15	$\times(2,-,-)$	$\times(10,-,-)$	0.11	0.01
3b	"	"	"	"	"	$5,-,-$	"	0.49	0.02
3c	"	"	"	"	46/46	"	"	0.01	0.00
4a	1-3	-	1-4	1-3	15/15	$\times(2,\,0.5,\,1.5)$	-	0.84	-
4b	"	-	"	"	46/46	"	-	0.63	-
4c	2	-	"	"	"	$\times(-,0.5,-)$	-	0.46	-
4d	1-3	-	$\overline{12},\,\overline{34},\,\overline{56},\,\overline{78}$	1-3	15/15	$\times(2,\,0.5,\,1.5)$	-	0.13	-
4e	"	-	$\overline{123456},\,\overline{78}$	"	"	"	-	0.31	-
4f	"	-	"	"	"	$5,5,5$	-	0.31	-

Chapter 3

Development of the Adjoint of GEOS-Chem[1]

3.1 Introduction

Chemical transport models (CTMs) enhance our ability to understand the chemical state of the atmosphere and allow detailed analysis of issues ranging from intercontinental pollution transport to the coupling of anthropogenic processes, regional pollution and climate change. Of particular interest in these realms is explicit consideration of the role of aerosols, the importance of which is well documented. Given the substantial uncertainty that remains in many aspects of detailed aerosol simulations, it is critical to further examine how the numerous parameters in such models steer their predictions, especially estimates of emissions inventories for aerosols and their precursors. The complexity of the thermodynamic and photochemical processes that govern secondary formation of aerosols precludes simple assessment of the dependence of model predictions on such parameters. Working to arrive at CTMs that more reliably reproduce observations, adjoint modeling is often employed as a method for determining the sensitivity of model predictions to input parameters and for optimizing these parameters to enforce agreement between the model predictions and an

---

[1]Henze, D. K., A. Hakami and J. H. Seinfeld, Development of the Adjoint of GEOS-Chem, *Atmos. Chem. Phys.*, *7*, 2413-2433

observational data set.

Several inverse modeling studies have analyzed sources of aerosols and aerosol precursors on regional scales. As of yet, most studies have been fairly coarse, limited to optimization of a few scaling factors for emissions inventories spanning large domains. *Park et al.* (2003) used multiple linear regression to estimate annual mean sources of seven types of primary carbonaceous aerosol over the United States. A Kalman filter approach was used to estimate improved monthly emissions scaling factors for NH_3 emissions over the United States using observations of ammonium wet deposition in works by *Gilliland and Abbitt* (2001) and *Gilliland et al.* (2003, 2006). *Mendoza-Dominguez and Russell* (2000, 2001) optimized domain-wide emissions scaling factors for eight species over the eastern Unites States using observations of gas-phase inorganic and organic species and speciated fine particles. Source apportionment models have also been refined using inverse modeling (*Knipping et al.*, 2006; *Schichtel et al.*, 2006).

Data from satellite observations offer tremendous potential for inverse modeling of aerosols (*Collins et al.*, 2001; *Kahn et al.*, 2004). In order to best exploit these, and other, large data sets, it is desired to extend inverse analysis of aerosol models to global scales and to finer decomposition of the emissions domains. Such goals require consideration of inverse modeling methods designed for large sets of variable parameters. The adjoint method is known to be an efficient means of calculating model sensitivities that afford examination of numerous parameters, where these values can subsequently be used in tandem with an observational data set for data assimilation. First appearing in the field of atmospheric science in the early 1970s (*Marchuk*, 1974; *Lamb et al.*, 1975), the method later came to be applied extensively in meteorology, e.g., *Talagrand and Courtier* (1987); *Errico and Vukicevic* (1992). In the last decade, the adjoint approach has expanded to include ever more detailed CTMs, beginning with the abbreviated Lagrangian stratospheric model of *Fisher and Lary* (1995) and the Lagrangian tropospheric model of *Elbern et al.* (1997). *Vukicevic and Hess* (2000) used the adjoint method to perform a sensitivity study of an

inert gas-phase tracer over the Pacific, while *Elbern and Schmidt* (1999) presented the first adjoint of a 3-D Eulerian CTM to include chemistry. These initial works have been followed more recently by similar development and application of adjoint models of several CTMs: CHIMERE (*Vautard et al.*, 2000; *Menut et al.*, 2000; *Schmidt and Martin*, 2003), IMAGES (*Muller and Stavrakou*, 2005; *Stavrakou and Muller*, 2006), Polair (*Mallet and Sportisse*, 2004, 2006), TM4 (*Meirink et al.*, 2005), the California Institute of Technology urban-scale model (*Martien et al.*, 2006; *Martien and Harley*, 2006), and DRAIS (*Nester and Panitz*, 2006). The adjoint of the regional model STEM also has been developed (*Sandu et al.*, 2005a) and deployed (*Hakami et al.*, 2005, 2006; *Chai et al.*, 2006).

Of all the previous 3-D adjoint modeling studies, none includes detailed treatment of aerosols, likely owing to the difficult prospect of deriving the adjoint of the model routines dealing with aerosol thermodynamics. The study of *Hakami et al.* (2005) deals only with inert carbonaceous aerosols, and the work of *Dubovik et al.* (2004), though global in scale, does not include full chemistry or aerosol thermodynamics. Detailed adjoint modeling of aerosols began with the theoretical investigations of *Henze et al.* (2004) and *Sandu et al.* (2005). However, these are preliminary studies performed on idealized box model systems. In the current work we present the first adjoint of a global CTM that includes dynamics, full tropospheric chemistry, heterogeneous chemistry, and aerosol thermodynamics. We demonstrate the potential value of this tool for quantifying and constraining factors that govern global secondary inorganic aerosol formation. In addition, we note the general usefulness of the adjoint model of GEOS-Chem for a wide variety of applications, such as constraining CO emissions using satellite data (Kopacz et al., 2007[2]).

---

[2]Kopacz, M., Jacob, D., 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, submitted, 2007.

3.2 Forward and inverse models

The GEOS-Chem model is used to simulate global aerosol distributions (version 6.02.05 with a horizontal resolution of $4° \times 5°$ and 30 layers up to $0.01\,\text{hPa}$, GEOS-3 meteorological fields). This version of the model includes detailed gas-phase chemistry coupled with heterogeneous reactions, inorganic aerosol thermodynamics, and oxidative aging of carbonaceous aerosols (*Park et al.*, 2004). A few of the specific equations for various model processes are given in Sect. 3.3.3, along with their corresponding adjoints. We note here that gaseous SO_2 and primary sulfate are co-emitted in GEOS-Chem using a single emissions inventory, referred to as SO_x, which is partitioned between the two species on a regional basis, with sulfate comprising 5% of SO_x emissions in Europe, 1.7% in North America, and 3% elsewhere (*Chin et al.*, 2000).

The standard model has been modified to facilitate the specific inverse modeling goals of the present study. We neglect stratospheric chemistry, which over the course of the short simulations considered here should not have a substantial impact. The standard GEOS-Chem tropospheric chemical mechanism comprises 87 species and 307 reactions integrated using the SMVGEARII solver of *Jacobson* (1995). We retain this standard chemical mechanism; however, we implement a different numerical solver. The details of this are given in Appendix A. To summarize, we implement a 3rd order Rosenbrock solver that not only facilitates construction of the adjoint model, but also improves forward model efficiency. We also consider using offline concentrations of sulfate aerosol for calculation of photolysis rates and heterogeneous reaction probabilities, see Sect. 3.3.5.

3.2.1 Inverse modeling

An adjoint model is used to calculate the gradient of a cost function, J, with respect to a set of model parameters, $\mathbf{p}$, $\nabla_{\mathbf{p}} J$. For data assimilation applications, the cost function is

defined to be

$$J = \frac{1}{2} \sum_{\boldsymbol{c} \in \Omega} (\boldsymbol{c} - c_{\text{obs}})^T \mathbf{S}_{\text{obs}}^{-1} (\boldsymbol{c} - c_{\text{obs}}) + \frac{1}{2} \gamma_r (\boldsymbol{p} - p_a)^T \mathbf{S}_p^{-1} (\boldsymbol{p} - p_a) \tag{3.1}$$

where $\boldsymbol{c}$ is the vector of species concentrations mapped to the observation space, $\boldsymbol{c}_{\text{obs}}$ is the vector of species observations, $\mathbf{S}_{\text{obs}}$ is the observation error covariance matrix, $\boldsymbol{p}$ is a vector of active model parameters throughout the model domain, $\boldsymbol{p}_a$ is the initial estimate of these parameters, $\mathbf{S}_p$ is the error covariance estimate of these parameters, γ_r is a regularization parameter, and Ω is the domain (in time and space) over which observations and model predictions are available. We will sometimes use the notation c and p to represent single elements of the vectors $\boldsymbol{c}$ and $\boldsymbol{p}$. Using the variational approach, the gradient $\nabla_{\boldsymbol{p}} J$ is supplied to an optimization routine and the minimum of the cost function is sought iteratively. At each iteration, improved estimates of the model parameters are implemented and the forward model solution is recalculated. In this study, the magnitude of each variable parameter is adjusted using a scaling factor, σ, such that $p = \sigma p_a$. We use the L-BFGS-B optimization routine (*Byrd et al.*, 1995; *Zhu et al.*, 1994), which affords bounded minimization, ensuring positive values for the scaling factors.

Alternatively, for sensitivity analysis, the cost function can be defined as simply a set of model predictions,

$$J = \sum_{g \in \Omega_s} g(\mathbf{c}) \tag{3.2}$$

where Ω_s is the set of times at which the cost function is evaluated. The desired gradient values are the sensitivities of this set of model predictions to the model parameters.

3.2.2   Adjoint modeling

Equations for calculating the desired gradients using the adjoint method can be derived from the equations governing the forward model or from the forward model code. The prior approach leads to the continuous adjoint, while the latter leads to the discrete adjoint

(*Giles and Pierce*, 2000). The continuous adjoint equations for CTMs have been derived previously, using methods based upon the Lagrange duality condition (*Vukicevic and Hess*, 2000; *Pudykiewicz*, 1998; *Schmidt and Martin*, 2003) or Lagrange multipliers (*Elbern et al.*, 1997). Continuous adjoint gradients may differ from the actual numerical gradients of J, and continuous adjoint equations (and requisite boundary/initial conditions) for some systems are not always readily derivable; however, solutions to continuous adjoint equations can be more useful for interpreting the significance of the adjoint values. Many previous studies have also described the derivation of discrete adjoints of such systems (*Sandu et al.*, 2005a; *Muller and Stavrakou*, 2005). An advantage of the discrete adjoint model is that the resulting gradients of the numerical cost function are exact, even for nonlinear or iterative algorithms, making them easier to validate. Furthermore, portions of the discrete adjoint code can often be generated directly from the forward code with the aid of automatic differentiation tools. Here we present a brief description of the discrete adjoint method for the sake of defining a self-consistent set of notation for this particular paper; we refer the reader to the cited works for further derivations and discussions of continuous and discrete adjoints.

The GEOS-Chem model can be viewed as a numerical operator, F, acting on a state vector, $\boldsymbol{c}$

$$\boldsymbol{c}^{n+1} = F(\boldsymbol{c}^n) \tag{3.3}$$

where $\boldsymbol{c}$ is the vector of all K tracer concentrations, $\boldsymbol{c}^n = [c_1^n, \ldots, c_k^n, \ldots, c_K^n]^T$ at step n. In practice, F comprises many individual operators representing various physical processes. For the moment we will simply let F represent a portion of the discrete forward model which advances the model state vector from step n to step $n+1$.

For simplicity, we consider a cost function evaluated only at the final time step N with no penalty term. We wish to calculate the gradient of the cost function with respect to

the model state vector at any step in the model,

$$\nabla_{\boldsymbol{c}^n} J = \frac{\partial J(\boldsymbol{c}^N)}{\partial \boldsymbol{c}^n} \tag{3.4}$$

We define the local Jacobian around any given step as

$$\frac{\partial \boldsymbol{c}^{n+1}}{\partial \boldsymbol{c}^n} = \frac{\partial F(\boldsymbol{c}^n)}{\partial \boldsymbol{c}^n} = F_{\boldsymbol{c}}^n \tag{3.5}$$

Using the chain rule, we can expand the right hand side of Eq. (3.4) to explicitly show the calculation of $\boldsymbol{c}^N$ from $\boldsymbol{c}^n$,

$$\nabla_{\boldsymbol{c}^n} J = (F_{\boldsymbol{c}}^n)^T (F_{\boldsymbol{c}}^{n+1})^T \cdots (F_{\boldsymbol{c}}^{N-1})^T \frac{\partial J(\boldsymbol{c}^N)}{\partial \boldsymbol{c}^N} \tag{3.6}$$

Evaluating the above equation from left to right corresponds to a forward sensitivity calculation, while evaluating from right to left corresponds to an adjoint calculation. When K is larger than the dimension of J, which in this case is a scalar, the adjoint calculation is much more efficient (*Giering and Kaminski*, 1998).

For the adjoint calculation, we define the adjoint state variable $\boldsymbol{\lambda}_c^n$,

$$\boldsymbol{\lambda}_c^n = \frac{\partial J(\mathbf{c}^N)}{\partial \mathbf{c}^n}. \tag{3.7}$$

This can also be expanded,

$$\boldsymbol{\lambda}_c^n = \left[\frac{\partial \mathbf{c}^{n+1}}{\partial \mathbf{c}^n}\right]^T \frac{\partial J(\mathbf{c}^N)}{\partial \mathbf{c}^{n+1}} \tag{3.8}$$

$$= (F_{\mathbf{c}}^n)^T \frac{\partial J(\mathbf{c}^N)}{\partial \mathbf{c}^{n+1}}. \tag{3.9}$$

The equation above suggests how to solve for the adjoint variable iteratively. Initializing

the adjoint variable at the final time step

$$\boldsymbol{\lambda}_c^N = \frac{\partial J(\mathbf{c}^N)}{\partial \mathbf{c}^N} \tag{3.10}$$

we solve the following equation iteratively from $n=N,\ldots,1$,

$$\boldsymbol{\lambda}_c^{n-1} = (F_{\mathbf{c}}^n)^T \boldsymbol{\lambda}_c^n \tag{3.11}$$

The value of $\boldsymbol{\lambda}_c^0$ is then the sensitivity of the cost function with respect to the model initial conditions,

$$\boldsymbol{\lambda}_c^0 = \nabla_{c^0} J \tag{3.12}$$

The scheme above shows why calculating the adjoint variable is often referred to as "reverse integration" of the forward model, as we step from the final time to the initial time. This should not be confused with simply integrating the forward model equations backwards in time.

In order to calculate the sensitivity of J with respect to other model parameters, such as emissions, similar analysis (see, for example, *Sandu et al.* (2003)) shows that the gradient of the cost function with respect to these parameters,

$$\boldsymbol{\lambda}_p^0 = \nabla_p J \tag{3.13}$$

can be found by iteratively solving the following equation,

$$\boldsymbol{\lambda}_p^{n-1} = (F_p^n)^T \boldsymbol{\lambda}_c^n + \boldsymbol{\lambda}_p^n \tag{3.14}$$

where the subscripts c and p indicate sensitivity with respect to $\mathbf{c}$ and $\mathbf{p}$, respectively, and

$$F_p^n = \frac{\partial F^n}{\partial p} \tag{3.15}$$

When a penalty term is included in the cost function, the gradient becomes

$$\nabla_p J = \boldsymbol{\lambda}_p^0 + \gamma_r \mathbf{S}_p^{-1}(\mathbf{p} - \mathbf{p_a}) \qquad (3.16)$$

3.3   Constructing and validating the adjoint of GEOS-Chem

Here we present the derivation of the adjoint of GEOS-Chem. While the adjoint of the advection scheme is based upon the continuous approach, the remainder of the adjoint model is based upon the discrete formulation, using automatic differentiation tools for assistance. We use the Tangent and Adjoint Model Compiler (TAMC (*Giering and Kaminski*, 1998)), a freeware multipurpose program, and the Kinetic PrePocessor (KPP (*Sandu et al.*, 2003; *Damian et al.*, 2002)), a public domain numerical library for constructing the adjoint of chemical mechanisms. Always some, if not significant, manual manipulation of the code is required to use such tools. We often combine automatically generated adjoint code with manually derived discrete adjoint code to improve efficiency and transparency of the adjoint model.

Validation of the adjoint model is an important part of introducing an adjoint model of this size and complexity. Discrete portions of the adjoint code have the advantage of being easily validated via comparison of adjoint gradients to forward model sensitivities calculated using the finite difference approximation. The hybrid approach adopted here (discrete and continuous) requires detailed inspection of the adjoint gradients on a component-wise basis as discrepancies owing to the continuous portion are anticipated to obscure such comparisons for the model as a whole. Additional motivations exit for checking the gradients of subprocesses in the model separately and collectively. For large CTMs, it is not feasible to compare adjoint and finite difference gradients for each control parameter, as the finite difference calculation requires an additional forward model evaluation per parameter. However, component-wise analysis affords simultaneous examination of large numbers of sensitivities throughout the model domain, a much better approach to revealing

potential errors than performing validation checks in only a few locations. Furthermore, as GEOS-Chem has many routines common to other models, it behooves us to consider the adjoint of these routines separately.

Forward model sensitivities, Λ, are calculated using the finite difference (brute force) method. For component-wise tests of nonlinear routines, Λ is calculated using the two-sided formula,

$$\Lambda = \frac{J(\sigma + \delta\sigma) - J(\sigma - \delta\sigma)}{2\delta\sigma} \tag{3.17}$$

while for testing the full model, the more approximate one-sided finite difference equation,

$$\Lambda = \frac{J(\sigma + \delta\sigma) - J(\sigma)}{\delta\sigma} \tag{3.18}$$

is used in order to minimize the number of required forward model function evaluations. The latter method is also adequate for testing linear components of the model. We use $\delta\sigma$=0.1–0.01 for most tests, which experience showed to be an optimal balance between truncation and roundoff error. For most of these validation tests, it suffices to use a simplified cost function that does not depend on any observational data set, as in Eq. (3.2), defining g to be a predicted tracer mass, either gas- or aerosol-phase, in a single grid cell, or the total mass burden over a larger spatial domain.

3.3.1 Aerosol thermodynamics

The equilibrium thermodynamic model MARS-A (*Binkowski and Roselle*, 2003) is used to calculate the partitioning of total ammonia and nitric acid between aerosol and gas phases. While it is a relatively simple treatement compared to others such as SCAPE (*Kim et al.*, 1993) or ISORROPIA (*Nenes et al.*, 1998), the MARS-A model is still fairly complex. It uses an iterative algorithm to find equilibrium concentrations, considering two primary regimes defined by the ionic ratio of ammonium to sulfate and several sub-regimes defined by conditions such as relative humidity.

Several factors have historically prevented rigorous treatment of aerosol thermodynamics from inclusion in adjoint modeling studies of CTMs, or even adjoint studies of aerosol dynamics (*Henze et al.*, 2004; *Sandu et al.*, 2005). Division of the possible thermodynamic states into distinct regimes causes many discontinuities in the derivatives, precluding easy derivation of continuous adjoint equations and raising doubts to the value of such sensitivities. Furthermore, several coding tactics often employed in these types of models render them intractable for direct treatment using automatic differentiation tools.

We develop the adjoint of MARS-A in pieces, separating the model into several subprograms, the adjoints of which are then created using TAMC. Tracking variables are added to the forward model routine to indicate which of these subroutines to call during the adjoint calculation. Initial unequilibrated concentrations at the beginning of each external time step are saved in checkpoint files during the forward calculation. Intermediate values are recalculated from these during the adjoint integration. This type of two-level checkpointing strategy has been shown to optimally balance storage, memory and CPU requirements (*Griewank and Walther*, 2000; *Sandu et al.*, 2005a).

The accuracy of the resulting adjoint code is tested by comparing adjoint gradients to finite difference gradients calculated using Eq. (3.17) with $\delta\sigma = 0.1$. These comparisons can be made directly throughout the entire model domain by turning off all transport processes. Figure 3.1 shows comparisons for the sensitivity of surface level nitrate aerosol mass with respect to scaling factors for emissions of surface level anthropogenic SO_x and NH_3 after a week-long simulation. The gradients agree quite well, confirming the accuracy of the thermodynamic adjoint code. Discussion of values of model sensitivities is given in Sect. 3.4.

3.3.2 Chemistry

KPP (v2.2) (*Sandu et al.*, 2003; *Damian et al.*, 2002) is used to automatically generate code for the adjoint of the tropospheric chemistry solver, which calculates gradients with

respect to the initial species concentrations. We are also interested in the gradient with respect to the emission rates for those species whose emissions are incorporated into the chemical mechanism itself, such as NO_x, (as opposed to those that are simply injected into the model grid cells at intermediate times, such as SO_x). The additional equations for calculating discrete adjoint gradients with respect to reaction rate constants are derived in Appendix B. Though these equations have not been presented previously, KPP does provide the necessary subroutines for solving them.

To assess the accuracy of the adjoints of the chemistry routine, we calculate the sensitivity of the species concentrations at the end of a single chemistry time step (1 h) with respect to the emissions of NO_x (emitted as NO) in a box model test. For this test, the chemical environment is that of a polluted, urban grid cell in the afternoon. Figure 3.2 shows the ratio $\boldsymbol{\lambda}_{ENO_x}/\Lambda_{ENO_x}$ for three separate cases. Using a two-sided finite difference calculation (Eq. (3.17)) with $\delta\sigma_{ENOx}=0.1$ leads to agreement within a few percent. The dependence of the internal time step on species concentrations is a feedback not accounted for in the adjoint algorithm; hence, also holding the internal time step fixed at 60 s results in ratios of nearly 1.000 for all species. For comparison, the ratios when Eq. (3.18) is used for Λ_{ENO_x} are also shown, which can differ as much as 8% from unity, demonstrating the nonlinearity of such chemical systems.

The above test was reassuring, yet limited in scope for a global CTM. To test our adjoint model over a wide variety of chemical conditions, we also compare the accuracy of the adjoint derivatives of the chemical mechanism in global simulations over much longer time scales. We turn off all transport related processes in the model and calculate the adjoint and finite difference sensitivities of surface level tracer masses with respect to NO_x emissions in each location after a week-long simulation. As lack of transport leads to unrealistically extreme concentrations, emissions are reduced by an order of magnitude to prevent the chemical systems from becoming too stiff. Many chemical changes associated with aerosols are treated separately from the main tropospheric chemistry mechanism in

GEOS-Chem, such as aqueous reactions, dry deposition, chemical aging, and emission of SO_x and NH_3 (*Park et al.*, 2004). The adjoints of these processes are constructed separately (manually and with TAMC) and included in the following tests.

Figure 3.3 shows the adjoint and finite difference sensitivities of several species with respect to surface level, anthropogenic NO_x emissions scaling factors. We choose to show sensitivities of species such as acetone and methacrolein to NO_x emissions to also highlight the potential value of the adjoint model for analysis of non-aerosol species. We see from these, and similar tests for other active species (not shown), that the sensitivities calculated using the adjoint model consistently agree with those using the finite difference method over a wide range of conditions.

The code generated by KPP allows computation of either the continuous or discrete adjoints of the chemical mechanism. The continuous adjoint equation can be solved faster than the discrete adjoint equation at a given tolerance level, as calculation of the latter requires recalculation of intermediate values from the forward integration and computation of the Hessian during the adjoint integration, see Appendix B. At tight tolerance levels (i.e. very small internal time steps), the results of these methods should converge. However, for tolerance levels appropriate for global modeling, the continuous adjoint is only approximate, as $\boldsymbol{\lambda}+\delta\boldsymbol{\lambda}$, where $||\delta\boldsymbol{\lambda}|| < C \cdot Tol$. Given that the computational expense of the Rosenbrock solver increases substantially for tighter tolerance levels (see Appendix A), it is more efficient to use the discrete adjoint, even though this requires an additional forward integration. This is in contrast to the approach of *Errera and Fonteyn* (2001), who chose to approximate the necessary intermediate values by linearly interpolating from values stored at each external time step, an approach likely more appropriate for their stratospheric chemistry application.

GEOS-Chem accounts for the effect of aerosol concentrations on the radiation available for photolysis reactions and on the available surface area for the heterogeneous reactions included in the main chemical mechanism. The influence of the concentration of sulfate-

ammonium-nitrate aerosols on such rates is not currently accounted for in the adjoint model. We assume such an effect is less than 5% (*Liao et al.*, 1999; *Martin et al.*, 2003), especially as the absorbing aerosols (black carbon, mineral dust) are not active variables during these tests. The general agreement between $\boldsymbol{\lambda}$ and Λ, only the latter of which accounts for this effect, indicates this assumption is adequate, at least for simulations of this length. Further tests indicate that this assumption is valid for most, though not all, cases, see Sect. 3.3.5.

3.3.3 Convection, turbulent mixing, and wet removal

Wet removal of tracers in GEOS-Chem is generally treated as a first-order process, leading to discrete forward model equations of the form,

$$c_k^{n+1} = c_k^n e^{-r_{w,k} \triangle t} \tag{3.19}$$

Since the loss rate $r_{w,k}$ for most species does not depend on any active variables (*Jacob et al.*, 2000), the corresponding adjoint is simply

$$\lambda_k^n = \lambda_k^{n+1} e^{-r_{w,k} \triangle t} \tag{3.20}$$

The adjoints of these routines are generated using hand-created code, retaining efficiency and legibility. However, the in-cloud formation and cycling of sulfate aerosol from SO_2 is decidedly nonlinear, as the soluble fraction of SO_2 is limited by availability of H_2O_2, and a fraction of the SO_2 is reintroduced into the gas phase as sulfate when droplets evaporate (*Park et al.*, 2004). Such nonlinearities that span multiple program modules are treated both manually and with the help of TAMC, requiring additional recalculation and checkpointing of intermediate values.

Turbulent mixing in the boundary layer in the forward model is calculated according to

a mass-weighted mixing algorithm applied every dynamic time step (30 min for our case),

$$\mu_{k,j}^{n+1} = \frac{\sum_{l=1}^{L} m_l \mu_{k,l}^{n}}{m_T} \tag{3.21}$$

where $\mu_{k,j}$ is the mixing ratio (c/ρ, ρ is the density of air) of tracer k in layer j, m_l is the air mass in a single layer l, m_T is the total air mass in the boundary layer column, and L is the number of layers in the boundary layer. Rewritten in matrix form, this equation reads,

$$\begin{bmatrix} \mu_{k,1} \\ \vdots \\ \mu_{k,L} \end{bmatrix}^{n+1} = \begin{bmatrix} \frac{m_1}{m_T} & \cdots & \frac{m_L}{m_T} \\ \vdots & \ddots & \vdots \\ \frac{m_1}{m_T} & \cdots & \frac{m_L}{m_T} \end{bmatrix} \cdot \begin{bmatrix} \mu_{k,1} \\ \vdots \\ \mu_{k,L} \end{bmatrix}^{n} \tag{3.22}$$

Direct application of Eq. (3.11) yields the corresponding adjoint equation,

$$\begin{bmatrix} \lambda_{\mu_{k,1}} \\ \vdots \\ \lambda_{\mu_{k,L}} \end{bmatrix}^{n} = \begin{bmatrix} \frac{m_1}{m_T} & \cdots & \frac{m_1}{m_T} \\ \vdots & \ddots & \vdots \\ \frac{m_L}{m_T} & \cdots & \frac{m_L}{m_T} \end{bmatrix} \cdot \begin{bmatrix} \lambda_{\mu_{k,1}} \\ \vdots \\ \lambda_{\mu_{k,L}} \end{bmatrix}^{n+1} \tag{3.23}$$

which can be simply written as,

$$\lambda_{\mu_{k,j}}^{n} = \frac{m_j \sum_{l=1}^{L} \lambda_{\mu_{k,l}}^{n+1}}{m_T} \tag{3.24}$$

Deep convection is calculated in the forward model using cumulus cloud fluxes and an RAS type algorithm, see Appendix A of *Allen et al.* (1996). We calculate the discrete adjoint of this scheme using TAMC, noting that TAMC initially generates code that is accurate, yet several orders of magnitude slower than necessary due to several superfluous loops that have to be removed manually. The numerical scheme for the forward calculation iteratively solves a set of essentially linear equations, with an internal time step of five minutes. If we neglect a single conditional statement that checks only for rare floating point exceptions,

then storage or recalculation of the intermediate values is not required for the adjoint calculation.

The adjoint model performance for a simulation including convection, turbulent mixing, and wet deposition is tested by comparison of finite difference sensitivities to the adjoint sensitivities of concentrations of a soluble tracer with respect to its initial concentrations in a location exhibiting strong convection, deposition, and mixing. Horizontal transport, chemistry, and aerosol thermodynamics are turned off for these tests. We use a perturbation of one percent for the finite difference calculation. The ratio λ_c/Λ_c for simulations that are 6 h, 1 d and 3 d in length are 0.9998, 1.0002 and 1.0003, from which we see consistent satisfactory agreement between the two methods. Performance is similar in other tested locations.

3.3.4 Advection

We implement the adjoint of the continuous advection equations. GEOS-Chem nominally employs a monotonic piecewise parabolic (PPM) advection routine (*Colella and Woodward*, 1984; *Lin and Rood*, 1996). Below we briefly show how this scheme can be used to solve the continuous adjoint advection equations and afterwards address some of the issues wedded to this approach. We consider the 1-D example of the advection equation for a tracer in mass concentration units,

$$\frac{\partial c}{\partial t} = -\frac{\partial (uc)}{\partial x} \tag{3.25}$$

where u is the wind velocity in the x-direction. The forward numerical model actually solves the flux form of Eq. (3.25) in terms of the mixing ratio (*Lin and Rood*, 1996),

$$\frac{\partial (\rho\mu)}{\partial t} = -\frac{\partial (\rho\mu u)}{\partial x} \tag{3.26}$$

Assuming that the continuity equation for ρ is satisfied, this can be rewritten in the advection form,

$$\frac{\partial \mu}{\partial t} = -u\frac{\partial \mu}{\partial x} \tag{3.27}$$

Applying the adjoint variable as a Lagrange multiplier and integrating by parts (see, for example, Appendix A of *Sandu et al.* (2005a)), the continuous adjoint of Eq. (3.27) is

$$-\frac{\partial \lambda_\mu}{\partial t} = \frac{\partial (\lambda_\mu u)}{\partial x} \tag{3.28}$$

where λ_μ is the adjoint of the mixing ratio. Note that we have assumed that the winds (or any other met fields) are not active variables; taking the adjoint with respect to the meteorology is another task in itself (see, for example, *Giering et al.* (2004)). Applying the simple transform $\hat{\lambda}_\mu = \lambda_\mu / \rho$, and substituting this into Eq. (3.28), we arrive at the following adjoint equation,

$$-\frac{\partial (\rho \hat{\lambda}_\mu)}{\partial t} = \frac{\partial (\rho \hat{\lambda}_\mu u)}{\partial x} \tag{3.29}$$

which is similar in form to Eq. (3.26). If we assume that ρ is relatively constant over a single dynamic time step and that the advection is linear, then we can simply solve Eq. (3.29) using the same numerical code that was used to solve Eq. (3.26) in the forward model, scaling the adjoint by $1/\rho$ before and re-scaling by ρ afterwards, which is equivalent to solving Eq. (3.28).

While the continuous approach was in part adopted for reasons of practicality (the discrete advection algorithm in the forward model not being directly amenable for use with automatic differentiation tools), subsequent investigation indicates that the continuous approach is suitable, if not preferable. This is not surprising, as it is well documented that discrete adjoints of sign preserving and monotonic (i.e. nonlinear and discontinuous) advection schemes are not well behaved and can contain undesirable numerical artifacts, see for example *Thuburn and Haine* (2001), *Vukicevic et al.* (2001), and Liu and Sandu,

2006^3.

To illustrate the benefits of the continuous adjoint approach for our system, the following numerical test is performed. The sensitivity of aerosol concentrations with respect to concentrations in a neighboring cell six hours earlier are calculated for a meridional cross section of the northern hemisphere. To afford simultaneous calculation of finite difference and adjoint sensitivities throughout this domain, only horizontal advection in the E/W direction is included in these tests. Figure 3.4 shows finite difference sensitivities calculated using Eq. (3.18) for several values of $\delta\sigma$ as well as the adjoint gradients. The undesirable nature of the finite difference sensitivities is indicated by negative sensitivities that have no physical meaning. That negative values become more prevalent as $\delta\sigma \rightarrow 0$ indicates such values are caused by discontinuities in the discrete algorithm (*Thuburn and Haine*, 2001). We can expect that adjoint sensitivities of the discrete advection algorithm would contain similar features, which, despite being numerically precise gradients of the cost function, can result in convergence to undesirable local minimums for data assimilation (*Vukicevic et al.*, 2001). Given the importance of transport for analysis of aerosols, use of the continuous approach is deemed preferable to implementing a linear transport scheme with well-behaved discrete adjoints at the cost of forward model performance.

3.3.5   Combined performance

Again we compare the gradients calculated using the adjoint model to those calculated using the finite difference method, this time including all model processes. We calculate the sensitivity of global aerosol distributions of sulfate, ammonium, and nitrate to surface emissions of anthropogenic SO_x, NO_x and NH_3 in select locations. As noted previously, such comparisons are quite time consuming to perform on a global scale owing to the expense of the finite difference calculations. Attempting to cover a wide range of conditions, while keeping the number of required calculations within reason, we choose to analyze ten

---

[3]Liu, Z. and Sandu, A.: Analysis of Discrete Adjoints of Numerical Methods for the Advection Equation, Int. J. Numer. Meth. Fl., submitted, 2006.

locations for each set of emissions considered, see Fig. 3.5. The simulations are one day in length, and the cost function (Eq. 3.2) is evaluated only once at the end of the day. We use a perturbation of $\delta\sigma$=0.1 and Eq. (3.18) for the finite-difference calculations.

Figure 3.6 shows the adjoint gradients compared to the finite difference gradients for each of nine relationships. From visual inspection of the scatter plots, it is clear that the agreement is generally within reason given the fact that using a continuous adjoint for advection is expected to cause some amount of discrepancy. Regression lines, slopes, and R^2 values are given for each set of comparisons. The absolute difference between the two methods is often more substantial for the larger values. As the gradients in a given set usually span several orders of magnitude, many of the slopes are biased by a few such larger values and are not representative of the overall fit. However, accounting for such heteroscedasticity by re-scaling the gradients by $1/\mathbf{p}$ or performing weighted regressions that place less emphasis on the larger values still leads to the same general results. Picking twice as many test cells, different test cells, or a different value of $\delta\sigma$ also was not found to substantially alter the overall comparisons.

Initial comparison (not shown) of gradients for five of the 90 tests showed underestimation of adjoint sensitivities by more than an order of magnitude. Four of these tests were for the sensitivity of sulfate with respect to NH_3 emissions while one was for the sensitivity of nitrate with respect to SO_x emissions. Using offline concentrations for calculation of the contribution of sulfate aerosol to photolysis rates and heterogeneous reaction probabilities in the main tropospheric chemical mechanism for these tests alleviated the discrepancy, demonstrating that while this feedback is generally negligible, it is occasionally quite strong. Future work will extend the adjoint model to account for this feedback.

Napelenok et al. (2006) performed a complementary analysis on a regional scale, calculating the sensitivities of local aerosol distributions with respect to domain-wide precursor emissions over the United States with a forward sensitivity method (DDM-3D), using finite-difference calculations to check their results. While they found similarly good agreement for

the more direct relationships (such as sensitivity of sulfate with respect to SO_2 emissions, or ammonium with respect to NH_3 emissions), they had difficulty verifying the variability in the sensitivities of some of the more indirect relationships (such as the sensitivity of sulfate to NH_3 emissions or nitrate to SO_2 emissions). Granted, they used the more complex and rigorous thermodynamic model ISORROPIA; they suggested that such discrepancies were due to numerical diffusion, with spatial oscillations of the sensitivities indicative of errors due to transport.

In our tests, transport does not drastically degrade the consistency of the correlation between the two approaches; all of the R^2 are near unity. There is, however, some amount of bias in the comparisons, as indicated by slopes ranging from 0.8 to 1.3, and this does appear to be a result of transport. Figure 3.7 contains scatter plots of the sensitivities of sulfate with respect to NO_x emissions for several additional tests. Panel (a) shows the results when advection is turned off. This leads to improved agreement, m=1.03, compared to the center left panel of Fig. 3.6; hence, the source of this bias is presumably advection. As shown in Fig. 3.4, the adjoint gradients are likely smoother and more physically meaningful than the finite difference sensitivities.

To assess the extent to which using the continuous adjoint of advection hinders this approach to validating the adjoint model as a whole, we perform additional tests, the results of which are shown in Figure (3.7). Including advection, but evaluating the cost function only in a single location, rather than globally, leads to a very unsmooth adjoint field and triggers many nonlinear and discontinuous aspects of the numerical scheme in a manner inconsistent with advection of the relatively smooth concentration field in the forward model; hence, agreement between adjoint and finite difference gradients under these conditions is worse, see panel (b). All of the tests so far have been based on a single evaluation of the cost function at the end of a day-long simulation. The effects of changing the assimilation window (the time between consecutive evaluations of the cost function) and the total simulation length are shown in panels (c) and (d). Doubling both the simulation

length and the assimilation window to two days leads to an increased discrepancy, panel (c). Again, such behavior is likely owing to discrepancies between the finite difference and adjoint sensitivities of the advection scheme that can accumulate when integrating such sensitivities over several other nonlinear processes. Doubling only the simulation length but maintaining a one-day assimilation window improves the agreement, panel (d), as forcing from additional observations outweighs spurious discrepancies from advection.

Finally, we consider a more realistic example. Model predictions are compared to measurements of aerosol nitrate from the IMPROVE network of monitoring stations (http://vista.cira.colostate.edu/improve/). The sensitivities of the error weighted squared difference between predicted and observed nitrate aerosol with respect to natural NH_3 emissions scaling factors are shown in Fig. 3.8. The cost function is evaluated regionally only on the U.S. East Coast (72.5° W – 82.5° W), and the model is run for ten days starting Jan 1, 2002. Daily average measurements are assimilated during three of the ten days. Also shown is a comparison between the adjoint sensitivities and finite difference sensitivities evaluated for the same domain. That the overall discrepancy is not much different from the simple 24 h tests (Fig. 3.6, or Fig. 3.7, panel (b)) increases our confidence in the ability of short tests to diagnose the model's performance in practical applications.

Overall, we find the accuracy of the adjoint gradients to be satisfactory. The adjoint model clearly captures the dependence of inorganic aerosol burdens on the chemical and thermodynamic interactions that lead to their formation. While using the continuous adjoint of advection makes this verification process more laborious, we have characterized the discrepancies for future reference.

3.3.6 Computational efficiency

Here we report computational resource requirements for running the adjoint model of GEOS-Chem on a Linux workstation with dual Intel Itanium 1.5 GHz processors and 4 GB of RAM. The adjoint model utilizes multiple processors on shared memory architectures

as efficiently as the forward model. It requires $16\,\mathrm{KB}$ of checkpoint storage space per simulated day per grid cell; this amounts to $11\,\mathrm{GB}$ of storage space per week with the current model configuration. This is comparable to the storage requirements of other adjoint models of CTMs such as STEM, $40\,\mathrm{KB}$ per day per cell (*Sandu et al.*, 2005a), or the CIT model, $100\,\mathrm{KB}$ per day per cell (*Martien et al.*, 2006), taking into account that the time step is $30\,\mathrm{min}$ in GEOS-Chem (for this study), $15\,\mathrm{min}$ for STEM, and $3\,\mathrm{min}$ for the CIT model. The computational cost of the adjoint model (backward only) of GEOS-Chem is 1.5 times that of the forward model, requiring $2.5\,\mathrm{h}$ for a week long iteration (forward and backward). Adjoint models of other CTMs report this ratio as: STEM: 1.5, CHIMERE: 3–4, IMAGES: 4, Polair: 4.5–7, CIT: 11.75. We see that the adjoint of GEOS-Chem is quite efficient; in general, adjoint codes that are derived by hand or use specialized tools such as KPP are most efficient. Such efficiency is the trade-off for the labor involved in manually constructing an adjoint model of this size and complexity.

3.4   Sensitivity analysis

In this section we demonstrate how the adjoint model can be used as an efficient method of investigating the sensitivity of modeled aerosol concentrations to their precursor emissions. Sensitivity calculations for the full model are performed for a week-long simulation. Figure 3.9 shows the sensitivity of global burdens of sulfate, nitrate and ammonium aerosol to surface level emissions of anthropogenic SO_x, NO_x and NH_3. The cost function is evaluated once daily. Other results retrieved from the same calculations (not shown) are sensitivities of these species with respect to the following emissions: stack SO_x, stack NO_x, biofuel SO_2, biomass burning SO_2, ship SO_2, biofuel NH_3, biomass burning NH_3, and natural NH_3.

The sensitivities in Fig. 3.9 encompass a wide range of relationships between aerosols and their primary precursors. Some of these relationships are practically intuitive, such as the sensitivities of sulfate to SO_x emissions or of nitrate to NO_x emissions, both of which are generally large and positive. The sensitivity of ammonium to emission of NH_3 is also

positive, and the sensitivities of ammonium to SO_x and NO_x emissions are always positive, owing to uptake of NH_3 on inorganic aerosol by sulfate and nitrate.

Some of the relationships in Fig. 3.9 are less obvious, such as the negative sensitivity of sulfate to emissions of NH_3. This effect is smaller in magnitude than some of the others, because the relationship between NH_3 emissions and sulfate aerosol concentrations is less direct. As total sulfate is conserved in the MARS-A aerosol equilibrium model, this effect is not due to thermodynamic interactions between ammonium and sulfate. The only species directly affected by NH_3 or ammonium concentrations are nitrate and nitric acid, via thermodynamic interactions. Therefore, the relationship between NH_3 and sulfate is dictated by the interactions between sulfate and nitrate, and, hence, NO_x. The sensitivity of nitrate to SO_x is largely negative, owing to thermodynamic competition between nitrate and sulfate for ammonium. The sensitivity of nitrate to NH_3 is entirely positive, due to the necessary presence of excess NH_3 for HNO_3 to condense. The combination of these two effects explains the overall negative relationship between sulfate and emissions of NH_3.

Within the global trends noted above, there is also much discernible local variability. For example, there are a few locations where the sensitivity of sulfate to NH_3 emissions changes abruptly from predominantly negative to locally positive. Some of these actually correspond to similarly abrupt shifts between areas that are sulfate-poor to areas that are sulfate-rich, such as the tip of South America and immediately west of the Iberian Peninsula. In other conditions or times of the day, emission of NO_x can actually lead to a decrease in nitric acid, and, hence, nitrate.

While the adjoint model accounts for nonlinearities in the relationships between emissions and aerosols, the results of the adjoint calculation are still merely tangent linear derivatives (gradients) which are likely to be valid over only a limited range of values for the parameters (emissions). We explore the robustness of the aerosol sensitivity calculations with respect to the magnitude of the emissions. Figure 3.10 shows the sensitivity of nitrate with respect to NO_x emissions calculated when the emissions are multiplied by

uniform scaling factors of 0.75 and 1.25; the relative differences between these values and base case sensitivities shown in Fig. 3.9. The sensitivities can differ substantially on a point to point basis ($>50\%$), particularly near boundaries between the positive and negative sensitivities or in areas where the sensitivities are very small. The differences are generally much less ($<20\%$) in areas with the largest sensitivities such as Europe, Eastern Asia and the Eastern United States. Despite these relative differences, the sensitivity field, viewed on the global (log) scale, remains nearly identical to the base case values. While individual sensitivities may be valid only over a limited range, the sensitivity field as a whole appears fairly robust.

Overall, the adjoint model is a promising tool for examining the dependence of aerosol concentrations on emissions. We note that the time required to calculate all of these sensitivities was less than 10 times the cost of a single forward model evaluation, while obtaining these results using the finite difference method would have required >5000 times the cost of a forward run.

3.5 Inverse modeling tests

Several inverse modeling tests are performed to assess the capabilities of the adjoint model in a data assimilation application. Using the twin experiment framework, pseudo observations, $\mathbf{c}_{\mathrm{obs}}$, are generated with the forward model using a base set of emissions parameters, $p=p_a$. An active subset of the parameters used to generate these observations is then perturbed using scaling factors, $\sigma=p/p_a$, each of which is allowed to vary independently in every grid cell for each emitted species. The inverse model uses the pseudo-observations to recover the original unperturbed values of these active parameters.

We begin by generating a week-long set of observational data using the forward model with all scaling factors set equal to unity. For these initial tests, we perturb one set of emissions by re-scaling the emissions in every cell by a factor of two, and we use observations in every grid cell once every 24 h to force the data assimilation. As there is no error in these

observations, equal weight is ascribed to each ($\mathbf{S}_{\mathrm{obs}}^{-1}$ is the identity matrix), and the error covariance of our initial (perturbed) estimate of the emissions scaling factors is infinite ($\mathbf{S}_p^{-1}$ is zero). Such conditions are unrealistic and serve only to test the adjoint model under the most ideal conditions possible.

In the first set of tests (DA1), we perturb the emission inventories of (a) surface level anthropogenic SO_x, (b) biomass burning SO_2 and (c) biofuel SO_2. We assimilate observations of sulfate for the week of 1–7 July 2001. Figure 3.11 shows the progression of the normalized (divided by the initial value) cost function at iteration i during the optimization procedure, J_i/J_1. The cost function quickly reduces by at least five orders of magnitude in each case. The correct emissions inventories are essentially entirely recovered.

In the next test (DA2), we perturb the emission inventory of NH_3 from anthropogenic sources, and assimilate observations of aerosol ammonium. This is a slightly more difficult inversion as ammonium measurements alone do not fully constrain NH_3 emissions (*Gilliland et al.*, 2006). As demonstrated in Sect. 3.3.5, ammonium is indirectly, yet appreciably, coupled to gas-phase oxidants. Utilizing observations of O_x (O_3, NO_2 and NO_3) in conjunction with ammonium observations noticeably increases the convergence rate over using either type of observations alone, see Fig. 3.12. This demonstrates, albeit in a highly idealized fashion, the potential for exploiting multi-phase measurements as constraints for aerosol modeling.

The final test (DA3) attempts to mimic a slightly more realistic scenario than the previous tests: improving estimates of global anthropogenic SO_x and NO_x emission inventories using surface measurements of sulfate, nitrate, and ammonium aerosol. In this case, the emissions inventories are perturbed regionally by 5–30% with an additional random factor of order 5%. For example, the anthropogenic SO_x and NO_x emissions in North America are perturbed by factors of $0.8+r$ and $0.85+r$, respectively, while emissions in Asia are perturbed by factors of $1.2+r$ and $1.3+r$, where r is a random number uniformly distributed between 0 and 0.05. The error covariance matrix S_p is calculated using an ascribed error

of 100% and is assumed to be diagonal. Observations are used once per day in only half of the land-based surface grid cells. The reduction of the cost function after 15 iterations is shown in Fig. 3.13. The difference between the true emission inventories for SO_x and NO_x and the estimated inventory at the first and final iterations are shown in Fig. 3.14. While there are substantial improvements in the SO_x emissions and the NO_x emissions in Europe and Asia, the NO_x emissions in North America have yet to converge. Although the cost function has reduced by nearly two orders of magnitude, the optimization procedure has clearly yet to reach a minimum. In applications of this type, the procedure is often halted according to an appropriate convergence criteria. Further iterations might be justified; however, care must be taken to avoid overly minimizing the predictive error component of the cost function at the sake of generating noisy solutions.

3.6   Summary and conclusions

The derivation of the adjoint model of GEOS-Chem has been presented in a piecewise fashion. We have implemented the first adjoint of an aerosol equilibrium thermodynamic model (MARS-A, *Binkowski and Roselle*, 2003), derived using the automatic differentiation tool TAMC (*Giering and Kaminski*, 1998), which required significant manual pre- and post-processing owing to the structure and complexity of the code. To facilitate construction of the adjoint of the GEOS-Chem gas-phase chemical mechanism, we implemented a Rosenbrock solver using the KPP numerical library (*Sandu et al.*, 2003). This has allowed for automatic generation of the adjoint of the chemical mechanism and also improved forward model performance (see Appendix A). The adjoints of wet removal, deep convection, and turbulent mixing were derived manually and with the aid of TAMC. We have used the continuous adjoint method to treat advection, wherein the same numerical algorithm is used to solve the continuous adjoint advection equation as was used for tracer advection in the forward model.

All aspects of the adjoint model have been tested both separately and together by

comparing the adjoint gradients to finite difference gradients. Each individual discrete adjoint routine showed satisfactory performance over a wide range of conditions. The adjoint gradients of the cost function evaluated using the full model are well correlated with the numerical gradients, as measured using finite difference calculations, with most $R^2 > 0.95$. The hybrid approach adopted here avoids physically unrealistic noise associated with discrete adjoints of nonlinear and discontinuous advection schemes and does not entirely preclude validation of the adjoint model as a whole via comparison to finite difference gradients. Such comparisons are understandably unrevealing when considering sparse or infrequent data; however, in both ideal test calculations with smooth adjoint forcings and realistic tests of week-long sensitivities of predictions of actual aerosol observations, the comparisons are consistent enough to ensure proper derivation of the adjoint. Nevertheless, this treatment necessitated additional inspection of model performance on a component-wise basis. While these benchmarks set the standard for further use and development of this adjoint model, future applications may require additional testing.

The adjoint model clearly demonstrates the importance and relative strengths of many complex nonlinear relationships connecting concentrations of aerosol species and their precursor emissions. Though indirect, relationships such as the dependence of sulfate aerosol concentrations on emission of NH_3 or NO_x are captured by the adjoint model, and can be determined globally in an efficient manner. The sign and magnitude of many of these sensitivities exhibit a rich array of features owing to the influence of environmental factors, such as the sulfate to ammonium ratio, cloud processing of SO_2, and variability in the NO_x and O_x levels.

We have also demonstrated the capabilities of the adjoint model in mock data assimilation applications. An adjoint model of this type allows for the possibility of exploiting multi-phase observations to constrain emissions of aerosol precursors. Here we have focused on regional variability of the emissions inventories, though the emissions can also be adjusted on a temporal basis. For real data assimilation projects, many application

specific issues inherent in this type of inverse modeling have yet to be resolved, such as specification of the error covariance matrices $\mathbf{S}_{\text{obs}}$ and $\mathbf{S}_p$. The dependance of adjoint model performance is known to depend strongly on such factors (*Chai et al.*, 2006), proper formulation of which is necessary to ensure scaling of the inventories that are physically realistic (*Stavrakou and Muller*, 2006). Real world application will also likely require conditioning of the cost function to improve convergence rate (*Meirink et al.*, 2005) and tuning of the regularization parameter (*Hakami et al.*, 2005).

Subsequent studies will focus on expanding the adjoint model to capture feedbacks such as the effect of sulfate aerosol concentrations on photolysis rates and heterogeneous reaction probabilities, seen here to occasionally be quite important. Work on the adjoint of the aerosol equilibrium model ISORROPIA (*Nenes et al.*, 1998) is also in progress. Further application of the GEOS-Chem model will focus also on the exploitation of multi-phase measurements from sources such as surface stations, aircraft, and satellites as model constraints. The adjoint of GEOS-Chem has already been used to constrain emissions of carbon monoxide from Asia using satellite (MOPITT) measurements (Kopacz et al., 2007[2]), demonstrating the potential for addressing a wide range of scientific questions with this type of inverse model.

Appendix 3.A   Implementation of a Rosenbrock solver and comparison to SMVGEARII

Solving large systems of chemical rate equations in CTMs requires the use of special numerical tools, or solvers, that are specifically designed for this purpose. Taking the adjoint of such solvers manually, or using generic automatic differentiation tools, can be an onerous task. We desire to create the adjoint of the full chemical mechanism in GEOS-Chem using the KPP software library (*Sandu et al.*, 2003), which is a set of tools specifically built for automatic differentiation of chemical mechanisms and the numerical algorithms used

to solve these systems. In order to make use of these tools, we must first implement the KPP generated numerical integration routines in the forward model. We investigate the feasibility and ramifications of replacing the current solver in GEOS-Chem, SMVGEARII (*Jacobson*, 1995), with a KPP generated Rosenbrock solver. We consider the amount of work required to make such a switch, the efficiency of the Rosenbrock solver compared to the SMVGEARII solver, and the overall effect that such a switch has on the model predictions after a week-long simulation.

After manually translating the SMVGEARII mechanism input files to KPP input files, the KPP tools easily generate a set of Fortran code that solves the given system for a variety of supported Rosenbrock type integrators in a box model setting. Minimal manual adjustment to this code was required to interface with the 3-D GEOS-Chem model and to allow support for OpenMP parallelization. Some amount of modifications to the KPP code itself will be required to fully automate this process.

Next we consider the efficiency of the Rosenbrock solver and the SMVGEARII solver in a global simulation with only chemistry. For each species, in every cell, we compare the concentrations from benchmark solutions at the end of a day-long simulation to concentrations from a reference solution for each solver. The benchmark calculations span a set of tolerance levels $\{10^{-1} \leq \text{RTOL} \leq 10^{-5}, 10^6 \text{ molecules cm}^{-3} \geq \text{ATOL} \geq 10^{-2}$ molecules cm$^{-3}\}$ while the reference solutions were computed using tight tolerances (RTOL $= 10^{-8}$, ATOL$=10^2$ molecules cm^{-3}). RTOL and ATOL are the relative and absolute error tolerance levels, respectively. Looser tolerance levels result in repeated failure to converge in numerous grid cells.

To assess the accuracy of the two methods, following *Sandu et al.* (1997) we define the significant digits of accuracy (SDA) as

$$\text{SDA} = -\log_{10}(\max_k ER_k)$$

where ER_k is a spatially modified root mean square norm of the relative error of the

benchmark solution $(\hat{c}_{k,j})$ with respect to a reference solution $(c_{k,j})$ for species k in grid cell j,

$$ER_k = \sqrt{\frac{1}{|\theta_k|} \cdot \sum_{j \in \theta_k} \left| \frac{c_{k,j} - \hat{c}_{k,j}}{c_{k,j}} \right|^2}$$

For Θ total grid cells, θ_k is the set of all locations of significant concentrations of species k, $\{0 \leq \theta \leq \Theta : c_{k,j} \geq a\}$. A threshold value of $a = 10^6$ molecules cm^{-3} is chosen to avoid inclusion of errors from locations where concentrations of a given species are less than chemically meaningful values.

We present the results in the form of a work – precision diagram, wherein the value of SDA for each test is plotted versus the average computational expense for the solver to integrate the chemical mechanism for one hour. When calculating this average, we do not consider the time required during the initial six hours of the simulation, as each solver requires a bit of "spin up" time in order to adjust internal time steps to values more appropriate than the default starting step size according to the stiffness of the local system. Such spin up time is negligible with respect to the total computational cost of any simulation longer than a few days.

Figure A1 shows the work-precision diagram for the global benchmark simulations. The Rosenbrock solver is nearly twice as efficient as the SMVGEARII solver during these tests. Based on this analysis, we choose to run the Rosenbrock solver at tolerance levels that yield an SDA of ∼1.0 as the standard setting for this work.

For practical applications, we are interested in the difference in the total model predictions, including all model processes, incurred by switching to the Rosenbrock solver. We compare the daily average concentrations after a week-long simulation, including all model processes, calculated using the new standard Rosenbrock settings versus the standard SMVGEARII settings. Figure A2 shows the values of ER_k for each species k using the Rosenbrock solver to generate the test solution and SMVGEARII for the reference solution. This figure shows that after switching to this Rosenbrock solver, the solution is changed by

less than 10% for most species. The difference is larger, between 10 and 15%, for HNO_2, HNO_4, IAP, INO_2, ISOP, N_2O_5, NO, NO_2, PP, and RIP (for full definition of species, see http://www.env.leeds.ac.uk/$\sim$mat/GEOS-CHEM/GEOS-CHEM_Chemistry.htm). Determining whether or not this is an actual improvement in the accuracy of the forward model itself would require further comparison to observations. At the very least, the switch results in an improvement in the numerical solution of the forward model equations for slightly less computational cost.

Overall, while a more detailed analysis (requiring optimization of specific species tolerance levels and the parameters that control internal step size expansion and contraction) is necessary to determine unequivocally which method is more efficient, in our experience, not only is the Rosenbrock method desirable because of its differentiability, but it also appears to improve forward model performance by providing more accurate solutions to the model's chemical mechanism than the SMVGEARII solver for less computational expense. We have reported only the results using the Rodas-3 set of Rosenbrock coefficients; however, additional tests were performed using the other available sets (Ros-2,Ros-3,Ros-4,Rodas-4), and the trends were similar. It must also be emphasized that these comparisons should not be generalized to other platforms or CTMs; the SMVGEARII algorithm is designed to perform most efficiently on vector platforms by re-ordering the grid cells every external chemistry time step, an operation which serves only to increase the cost of this method by $\sim$5% on non-vector machines such as those used in this study, and most other GEOS-Chem studies.

Appendix 3.B   Discrete adjoint derivatives with respect to reaction rate constants

We desire to calculate the gradient of the cost function with respect to NO_x emissions. In GEOS-Chem, the emission and dry deposition of many species, such as NO_x, are incorpo-

rated into reactions in the tropospheric chemical mechanism as,

$$\frac{dc_{\mathrm{NO}}}{dt} = \mathrm{ENO_x} + \ldots \tag{B-30}$$

where $\mathrm{ENO_x}$ is the $\mathrm{NO_x}$ emission rate, emitted as NO. The strong influence of $\mathrm{NO_x}$ on the overall chemistry precludes using the continuous adjoint equation of the above equation,

$$\lambda_{\mathrm{ENO_x}} = \int \lambda_{\mathrm{NO}} dt \tag{B-31}$$

Hence, we must calculate the sensitivity of the discrete chemical solver itself with respect to the reaction rate coefficients. We present a derivation of these equations here, as they have not yet been presented elsewhere, and they are necessary for accurate calculation of the desired adjoint sensitivities.

For completeness, we first present the equations for the Rosenbrock method, which advances the forward model solution (c^n) from one step to the next using the following formulas,

$$c^{n+1} = c^n + \sum_{i=1}^{s} m_i k_i, \quad \mathrm{Err}^{n+1} = \sum_{i=1}^{s} e_i k_i \tag{B-32}$$

$$T_i = t^n + \alpha_i h, \quad C_i = c^n + \sum_{j=1}^{i-1} a_{ij} k_j \tag{B-33}$$

$$A = \left[\frac{1}{h\gamma} - J^T(t^n, c^n) \right] \tag{B-34}$$

$$A \cdot k_i = f(T_i, C_i) + \sum_{j=1}^{i-1} \frac{b_{ij}}{h} k_j + h\gamma_i f_t(t^n, c^n) \tag{B-35}$$

where s is the number of stages, $\alpha_i = \sum_j \alpha_{ij}$, $\gamma_j = \sum_j \gamma_{ij}$, m_i, α_{ij}, $a_{i,j}$, b_{ij}, γ_{ij}, and e_i are method coefficients, $f(\cdot, \cdot)$ is the ODE derivative function: $c' = f(t, c)$, $f_t(\cdot, \cdot)$ is the partial time derivative: $f_t(t, c) = \partial f(t, c)/\partial t$, $J(\cdot, \cdot)$ is the Jacobian: $J(t, c) = \partial f(t, c)/\partial c$, $J_t(\cdot, \cdot)$ is the partial time derivative of the Jacobian: $J_t(t, c) = \partial J(t, c)/\partial t$, and $H(\cdot, \cdot)$ is the Hessian:

$H(t,c)=\partial^2 f(t,c)/\partial c^2$. A is the system matrix, and C_i, T_i, k_i are internal stage quantities defined by the method. The J and γ used here are not likely to be confused with the use of these notations in the cost function definitions, and allow us to maintain consistent notation with the KPP documentation, which should be consulted for further explanations and values of the method coefficients.

The equation for the adjoint of the concentrations, λ_c, is obtained by differentiating the method with respect to c_n, see Eq. (3.11).

$$A \cdot u_i = m_i \lambda_c^{n+1} + \sum_{j=i+1}^{s} \left(a_{ji} v_j + \frac{b_{ji}}{h} u_j \right) \tag{B-36}$$

$$v_i = J^T(T_i, Y_i) \cdot u_i, \quad i = s, s-1, \cdots, 1 \tag{B-37}$$

$$\lambda_c^n = \lambda_c^{n+1} + \sum_{i=1}^{s} (H(t^n, c^n) \times k_i)^T \cdot u_i \tag{B-38}$$

$$+ \ h J_t^T(t^n, c^n) \cdot \sum_{i=1}^{s} \gamma_i u_i + \sum_{i=1}^{s} v_i$$

where v_i and u_i are internal stage vectors defined by the method. For GEOS-Chem, the reaction rates are constant over the internal time steps, hence we use the reduced form of this equation for autonomous systems,

$$\lambda_c^n = \lambda_c^{n+1} + \sum_{i=1}^{s} (H(t^n, c^n) \times k_i)^T \cdot u_i \tag{B-39}$$

$$+ \ \sum_{i=1}^{s} J^T(T_i, C_i) \cdot u_i.$$

Taking the derivative of the Rosenbrock method with respect to the reaction rate parameters, and applying Eq. (3.14), gives the following equation, again for autonomous systems,

$$\lambda_p^n = \lambda_p^{n+1} + \sum_{i=1}^{s} (J_p(t^n, c^n) \times k_i)^T \cdot u_i \tag{B-40}$$

$$+ \quad \sum_{i=1}^{s} f_p^T (T_i, C_i) \cdot u_i.$$

Though Eq. (B-40) is not implemented in the KPP generated adjoint code, KPP does generate the necessary routines for calculation of f_p ($dFun\_dRcoeff$) and $(J_p(t^n, c^n) \times k_i)^T$ ($dJac\_dRcoeff$). For emissions, the function derivative is simply the identity matrix, and the Jacobian derivative is zero as the emission ODE is independent of any other species concentrations, so the discrete adjoint of the emission rates is

$$\lambda_E^n = \lambda_E^{n+1} + \sum_{i=1}^{s} I \cdot u_i. \tag{B-41}$$

Bibliography

Allen, D. J., P. Kasibhatla, A. M. Thompson, R. B. Rood, B. G. Doddridge, K. E. Pickering, R. D. Hudson, and S. J. Lin, Transport-induced interannual variability of carbon monoxide determined using a chemistry and transport model, *J. Geophys. Res.*, *101*(D22), 28,655–28,669, 1996.

Binkowski, F. S., and S. J. Roselle, Models-3 community multiscale air quality (CMAQ) model aerosol component - 1. Model description, *J. Geophys. Res.*, *108*(D6), 4183, doi: 10.1029/2001JD001409, 2003.

Byrd, R. H., P. Lu, J. Nocedal, and C. Zhu, A limited memory algorithm for bound constrained optimization, *Scientific Computing*, *16*(5), 1190–1208, 1995.

Chai, T. F., G. R. Carmichael, A. Sandu, Y. H. Tang, and D. N. Daescu, Chemical data assimilation of Transport and Chemical Evolution over the Pacific (TRACE-P) aircraft measurements, *J. Geophys. Res.*, *111*(D2), D02301, doi:10.1029/2005JD005883, 2006.

Chin, M., R. B. Rood, S. J. Lin, J. F. Muller, and A. M. Thompson, Atmospheric sulfur cycle simulated in the global model GOCART: Model description and global properties, *J. Geophys. Res.*, *105*(D20), 24,671–24,687, 2000.

Colella, P., and P. R. Woodward, The piecewise parabolic method (PPM) for gas-dynamical simulations, *J. Comput. Phys.*, *54*(1), 174–201, 1984.

Collins, W., P. Rasch, B. Eaton, B. Khattatov, J.-F. Lamarque, and C. Zender, Simulating

aerosols using a chemical transport model with assimilation of satellite aerosol retrievals: Methodology for indoex, *J. Geophys. Res.*, pp. 7313–7336, 2001.

Damian, V., A. Sandu, M. Damian, F. Potra, and G. R. Carmichael, The kinetic preprocessor KPP - a software environment for solving chemical kinetics, *Comput. Chem. Eng.*, *26*(11), 1567–1579, 2002.

Dubovik, O., T. Lapyonok, Y. J. Kaufman, M. Chin, P. Ginoux, L. A. Remer, and B. N. Holben, Inversion of global distribution of aerosol sources using MODIS and AERONET data, *Optica Pura y Aplicada*, *37*(3), 2004.

Elbern, H., and H. Schmidt, A four-dimensional variational chemistry data assimilations scheme for Eulerian chemistry transport modeling, *J. Geophys. Res.*, *104*(D15), 18,583–18,598, 1999.

Elbern, H., H. Schmidt, and A. Ebel, Variational data assimilation for tropospheric chemistry modeling, *J. Geophys. Res.*, *102*, 15,967–15,985, 1997.

Errera, Q., and D. Fonteyn, Four-dimensional variational chemical assimilation of crista stratospheric measurements, *J. Geophys. Res.*, *106*(D11), 12,253–12,265, 2001.

Errico, R. M., and T. Vukicevic, Sensitivity analysis using an adjoint of the PSU-NCAR mesoscale model, *Mon. Weather Rev.*, *120*(8), 1644–1660, 1992.

Fisher, M., and D. J. Lary, Lagrangian four-dimensional variational data assimilation of chemical species, *Q. J. R. Meteorol. Soc.*, *121*, 1681–1704, 1995.

Giering, R., and T. Kaminski, Recipes for adjoint code construction, *ACM Transactions on Mathematical Software*, *24*(4), 437, 1998.

Giering, R., T. Kaminski, R. Todling, R. Errico, R. Gelaro, and N. Winslow, Generating tangent linear and ajdoint versions of NASA / GMAO's Fortran-90 global weather forcast model, 2004.

Giles, M. B., and N. A. Pierce, An introduction to the adjoint approach to design, *Flow Turbul. Combust.*, *65*, 393–415, 2000.

Gilliland, A., and P. J. Abbitt, A sensitivity study of the discrete Kalman filter (DKF) to initial condition discrepancies, *J. Geophys. Res.*, *106*(D16), 17,939–17,952, 2001.

Gilliland, A. B., R. L. Dennis, S. J. Roselle, and T. E. Pierce, Seasonal NH_3 emission estimates for the eastern United States based on ammonium wet concentrations and an inverse modeling method, *J. Geophys. Res.*, *108*(D15), 4477, doi:10.1029/2002JD003063, 2003.

Gilliland, A. B., K. W. Appel, R. W. Pinder, and R. L. Dennis, Seasonal NH_3 emissions for continental United States: Inverse model estimation and evaluation, *Atmos. Environ.*, 2006.

Griewank, A., and A. Walther, Algorithm 799: Revolve: An implementation of checkpointing for the reverse or adjoint mode of computational differentiation, *ACM Transactions on Mathematical Software*, *26*(1), 19–45, 2000.

Hakami, A., D. K. Henze, J. H. Seinfeld, T. Chai, Y. Tang, G. R. Carmichael, and A. Sandu, Adjoint inverse modeling of black carbon during the asian pacific regional aerosol characterization experiment, *J. Geophys. Res.*, *110*(D14), D14301, doi:10.1029/2004JD005671, 2005.

Hakami, A., J. H. Seinfeld, T. F. Chai, Y. H. Tang, G. R. Carmichael, and A. Sandu, Adjoint sensitivity analysis of ozone nonattainment over the continental united states, *Environ. Sci. Technol.*, *40*(12), 3855–3864, 2006.

Henze, D. K., J. H. Seinfeld, W. Liao, A. Sandu, and G. R. Carmichael, Inverse modeling of aerosol dynamics: Condensational growth, *J. Geophys. Res.*, *109*(D14), D14201, doi:10.1029/2004JD004593, 2004.

Jacob, D. J., H. Liu, C. Mari, and B. M. Yantosca, Harvard wet deposition scheme for GMI, 2000.

Jacobson, M. Z., Computation of global photochemistry with SMVGEAR-II, *Atmos. Environ.*, *29*(18), 2541–2546, 1995.

Kahn, R. A., et al., Aerosol data sources and their roles within PARAGON, *Bulletin of the American Meteorological Society*, *85*(10), 2004.

Kim, Y. P., J. H. Seinfeld, and P. Saxena, Atmoshperic gas-aerosol equilibrium –I. Thermodynamic model, *Aerosol. Sci. Tech.*, *19*, 157–181, 1993.

Knipping, E. M., N. Kumar, B. K. Pun, C. Seigneur, S. Y. Wu, and B. A. Schichtel, Modeling regional haze during the BRAVO study using CMAQ-MADRID: 2. Source region attribution of particulate sulfate compounds, *J. Geophys. Res.*, *111*(D6), D06303, doi:10.1029/2004JD005609, 2006.

Lamb, R. G., W. H. Chen, and J. H. Seinfeld, Numerico-empirical analysis of atmospheric diffusion theories, *J. Atmos. Sci.*, *32*, 1794–1807, 1975.

Liao, H., Y. L. Yung, and J. H. Seinfeld, Effects of aerosols on tropospheric photolysis rates in clear and cloudy atmospheres, *J. Geophys. Res.*, *104*(D19), 23,697–23,707, 1999.

Lin, S. J., and R. B. Rood, Multidimensional flux-form semi-lagrangian transport schemes, *Mon. Weather Rev.*, *124*(9), 2046–2070, 1996.

Mallet, V., and B. Sportisse, 3-d chemistry-transport model Polair: numerical issues, validation and automatic-differentiation strategy, *Atmos. Chem. Phys. Discuss.*, 2004.

Mallet, V., and B. Sportisse, Uncertainty in a chemistry-transport model due to physical parameterizations and numerical approximations: An ensemble approach applied to ozone modeling, *J. Geophys. Res.*, *111*(D1), D01302, doi:10.1029/2005JD006149, 2006.

Marchuk, G., Numerical solution of the problems of the dynamics of the atmosphere and the ocean (in Russian), *Gidrometeoizdat*, 1974.

Martien, P. T., and R. A. Harley, Adjoint sensitivity analysis for a three-dimensional photochemical model: Application to Southern California, *Environ. Sci. Technol.*, *40*(13), 4200–4210, 2006.

Martien, P. T., R. A. Harley, and D. G. Cacuci, Adjoint sensitivity analysis for a three-dimensional photochemical model: implementation and method comparison, *Environ. Sci. Technol.*, *40*(8), 2663–2670, doi:10.1021/es0510257, 2006.

Martin, R. V., D. J. Jacob, R. M. Yantosca, M. Chin, and P. Ginoux, Global and regional decreases in tropospheric oxidants from photochemical effects of aerosols, *J. Geophys. Res.*, *108*(D3), 4097, doi:10.1029/2002JD002622, 2003.

Meirink, Eskes, and Goede, Sensitivity analysis of methane emissions derived from SCIA-MACHY observations through inverse modelling, *Atmos. Chem. Phys. Discuss.*, 2005.

Mendoza-Dominguez, A., and A. G. Russell, Iterative inverse modeling and direct sensitivity analysis of a photochemical air duality model, *Environ. Sci. Technol.*, *34*(23), 4974–4981, 2000.

Mendoza-Dominguez, A., and A. G. Russell, Emission strength validation using four-dimensional data assimilation: application to primary aerosol and precursors to ozone and secondary aerosol, *J. Air & Waste Manage. Assoc*, *51*, 1538–1550, 2001.

Menut, L., R. Vautard, M. Beekmann, and C. Honore, Sensitivity of photochemical pollution using the adjoint of a simplified chemistry-transport model, *J. Geophys. Res.*, *105*(D12), 15,379–15,402, 2000.

Muller, J. F., and T. Stavrakou, Inversion of CO and NO_x emissions using the adjoint of the IMAGES model, *Atmos. Chem. Phys.*, *5*, 1157–1186, 2005.

Napelenok, S. L., D. Cohan, Y. Hu, and A. G. Russel, Decoupled direct 3D sensitivity analysis for particulate matter (DDM-3DPM), *Atmos. Environ.*, *40*(32), 6112–6121, 2006.

Nenes, A., S. N. Pandis, and C. Pilinis, ISORROPIA: A new thermodynamic equilibrium model for multiphase multicomponent inorganic aerosols, *Aquatic Geochemistry*, *4*(1), 123–152, 1998.

Nester, K., and H. J. Panitz, Sensitivity analysis by the adjoint chemistry transport model DRAIS for an episode in the Berlin Ozone (BERLIOZ) experiment, *Atmos. Chem. Phys.*, *6*, 2091–2106, 2006.

Park, R. J., D. J. Jacob, M. Chin, and R. V. Martin, Sources of carbonaceous aerosols over the United States and implications for natural visibility, *J. Geophys. Res.*, *108*(D12), 4355, doi:10.1029/2002JD003190, 2003.

Park, R. J., D. Jacob, B. D. Field, R. Yantosca, and M. Chin, natural and transboundary pollution influences on sulfate-nitrate-ammonium aerosols in the United States: implications for policy, *J. Geophys. Res.*, *109*, D15204, doi:10.1029/2003JD004473, 2004.

Pudykiewicz, J. A., Application of adjoint tracer transport equations for evaluating source parameters, *Atmos. Environ.*, *32*(17), 3039–3050, 1998.

Sandu, A., J. G. Verwer, J. G. Blom, E. J. Spee, G. R. Carmichael, and F. A. Potra, Benchmarking stiff ODE solvers for atmospheric chemistry problems.2. Rosenbrock solvers, *Atmos. Environ.*, *31*(20), 3459–3472, 1997.

Sandu, A., D. N. Daescu, and G. R. Carmichael, Direct and adjoint sensitivity analysis of chemical kinetic systems with KPP: Part I - theory and software tools, *Atmos. Environ.*, *37*(36), 5083–5096, 2003.

Sandu, A., D. Daescu, G. R. Carmichael, and T. Chai, Adjoint sensitivity analysis of regional air quality models, *J. Comput. Phys.*, 2005a.

Sandu, A., W. Liao, G. R. Carmichael, D. K. Henze, and J. H. Seinfeld, Inverse modeling of aerosol dynamics using adjoints: Theoretical and numerical considerations, *Aerosol. Sci. Tech.*, *39*(8), 677–694, 2005b.

Schichtel, B. A., W. C. Malm, K. A. Gebhart, M. G. Barna, and E. M. Knipping, A hybrid source apportionment model integrating measured data and air quality model results, *J. Geophys. Res.*, *111*(D7), D07301, doi:10.1029/2005JD006238, 2006.

Schmidt, H., and D. Martin, Adjoint sensitivity of episodic ozone in the Paris area to emissions on the continental scale, *J. Geophys. Res.*, *108*(D17), 8561–8577, doi: 10.1029/2001D001583, 2003.

Stavrakou, T., and J. F. Muller, 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.

Talagrand, O., and P. Courtier, Variational assimilation of meteorological observations with the adjoint of the vorticity equations. Part I: Theory, *Q. J. R. Meteorol. Soc.*, *113*, 1311–1328, 1987.

Thuburn, J., and T. W. N. Haine, Adjoints of nonoscillatory advection schemes, *J. Comput. Phys.*, *171*(2), 616–631, 2001.

Vautard, R., M. Beekmann, and L. Menut, Applications of adjoint modelling in atmospheric chemistry: sensitivity and inverse modelling, *Environ. Modell. Softw.*, *15*, 703–709, 2000.

Vukicevic, T., and P. Hess, Analysis of troposhperic transport in the pacific basin using the adjoint technique, *J. Geophys. Res.*, *105*(D6), 7213–7230, 2000.

Vukicevic, T., M. Steyskal, and M. Hecht, Properties of advection algorithms in the context of variational data assimilation, *Mon. Weather Rev.*, *129*(5), 1221–1231, 2001.

Zhu, C., R. H. Byrd, P. Lu, and J. Nocedal, L-BFGS-B: a limited memory FORTRAN code for solving bound constrained optimization problems, *Tech. rep.*, Northwestern University, 1994.
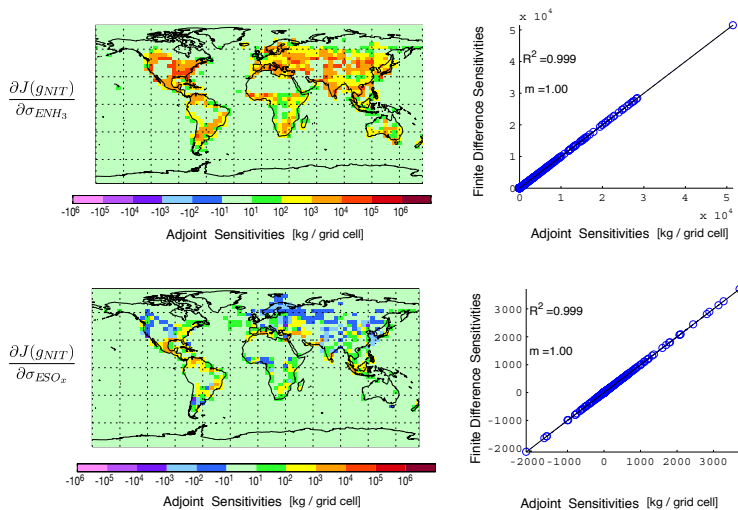
Figure 3.1: Thermodynamic adjoint validation. In the left column are the adjoint sensitivities of nitrate aerosol mass at the surface with respect to anthropogenic NH$_3$ and SO$_x$ emissions scaling factors. In the right column are the adjoint gradients compared to finite difference gradients. The cost function is evaluated once at the end of a week-long simulation that includes only aerosol thermodynamics and emissions of SO$_x$ and NH$_3$.
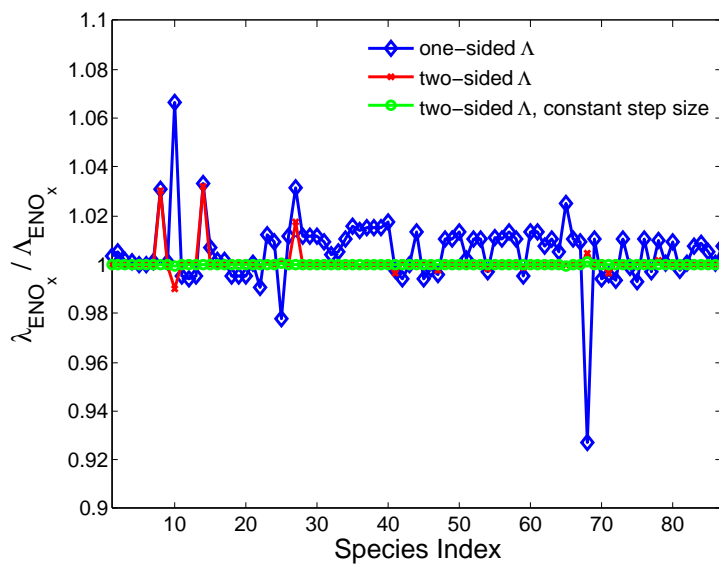
Figure 3.2: Chemistry adjoint validation. The ratios of the adjoint to finite difference sensitivities of each species with respect to NO_x emissions are calculated for a 1 h box model simulation. Results are shown for a one-sided finite difference calculation, $\delta\sigma = 0.1$ (blue ◇'s), a two-sided finite difference calculation (i.e. average of $\delta\sigma = 0.1$ and -0.1, red x's) and a two-sided finite difference calculation with a fixed internal time step of 60 s (green o's).
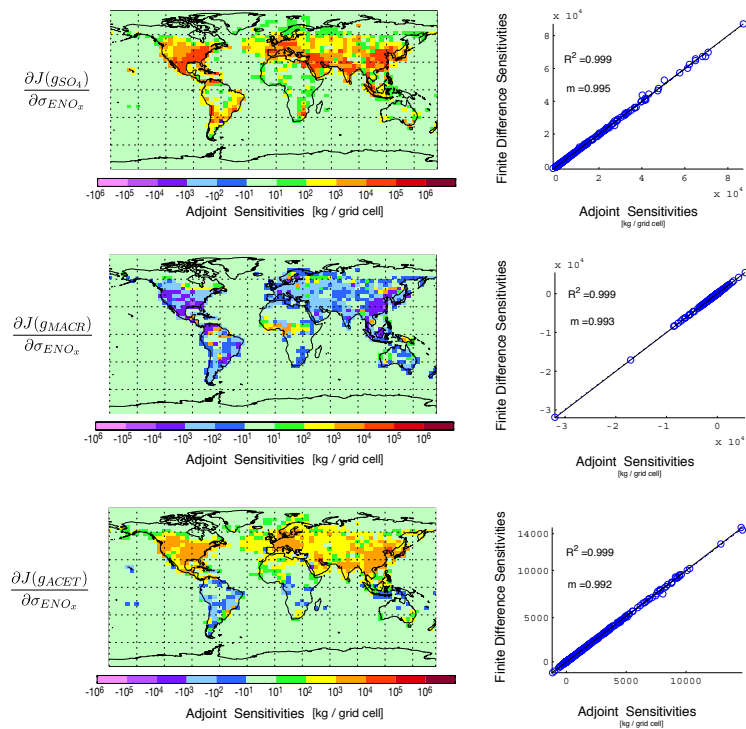
Figure 3.3: Chemistry adjoint validation. In the left column are the adjoint sensitivities of sulfate (SO$_4$), methacrolein (MACR), and acetone (ACET) at the surface with respect to surface level anthropogenic NO$_x$ emissions scaling factors. In the right column are the adjoint gradients compared to finite difference gradients. The cost function is evaluated once at the end of a week-long simulation with only chemistry and emissions$\times$0.1.
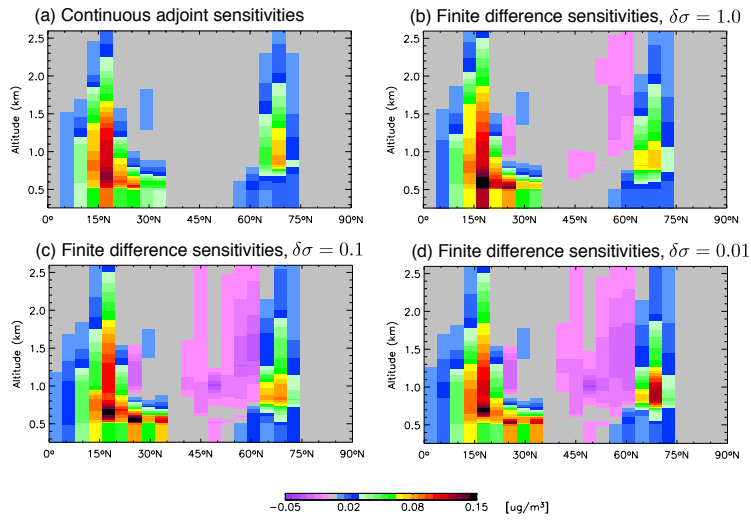
Figure 3.4: Sensitivities of aerosol concentrations with respect to concentrations in adjacent cells 6 h earlier considering only E/W advection. Sensitivities are calculated using: **(a)** continuous adjoint equation and **(b)**–**(d)** one-sided finite difference method with perturbations of $\delta\sigma$. The finite difference sensitivities contain more extreme values, including physically meaningless negative sensitivities that become more prevalent as $\delta\sigma \to 0$.
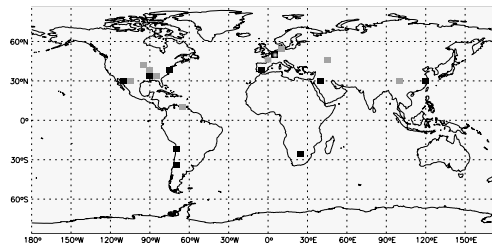
Figure 3.5: Select points for accuracy tests. Black locations used for anthropogenic emissions of SO_x and NO_x, grey points for NH_3, with one overlapping pair in Europe.
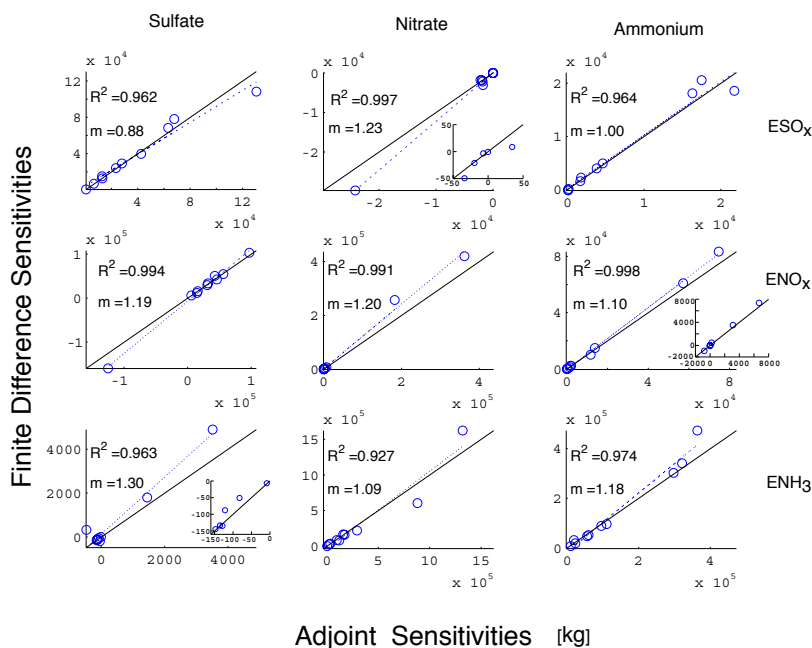
Figure 3.6: Full model performance. Comparison of sensitivities of global aerosol burdens (kg) to anthropogenic precursor emissions scaling factors calculated using the adjoint method vs. the finite difference method. A few of the plots contain insets with magnified views of a cluster of points.
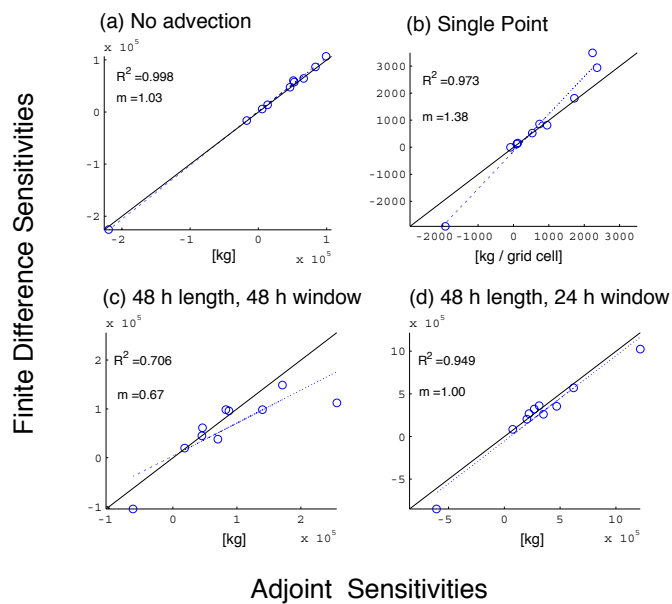
Figure 3.7: Effects of advection. Comparison of sensitivities of sulfate burdens to NO_x emissions scaling factors calculated using the adjoint method vs. the finite difference method. The base case (center left panel of Fig. 3.6) employs the standard PPM advection scheme, and the cost function is evaluated globally once at the end of a 24 h simulation. These cases differ from the base case in the following manner: **(a)** advection is turned off; **(b)** the cost function is evaluated in only a single region; **(c)** both the assimilation window and total simulation length are increased to 48 h; **(d)** the simulation length is increased to 48 h while the cost function is evaluated every 24 h.

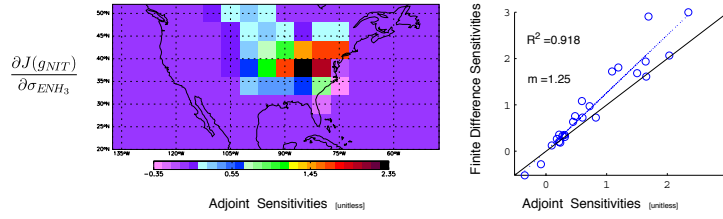$$\frac{\partial J(g_{NIT})}{\partial \sigma_{ENH_3}}$$

Figure 3.8: Sensitivities with respect to the error weighted squared difference between predicted and observed nitrate aerosol from the IMPROVE network for the first ten days of January, 2002. The cost function is evaluated only on the U. S. East Coast (72.5° W – 82.5° W). Shown are the sensitivities of the cost function with respect to natural NH_3 emissions scaling factors. On the right are the same quantities compared to finite difference sensitivities.
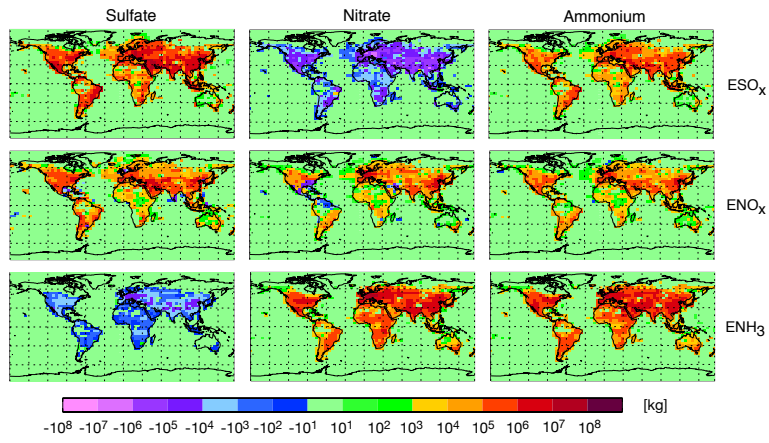
Figure 3.9: Sensitivities of global burdens of sulfate, nitrate and ammonium aerosol to anthropogenic SO_x, NO_x and NH_3 emissions scaling factors calculated using the adjoint model for a week-long simulation.
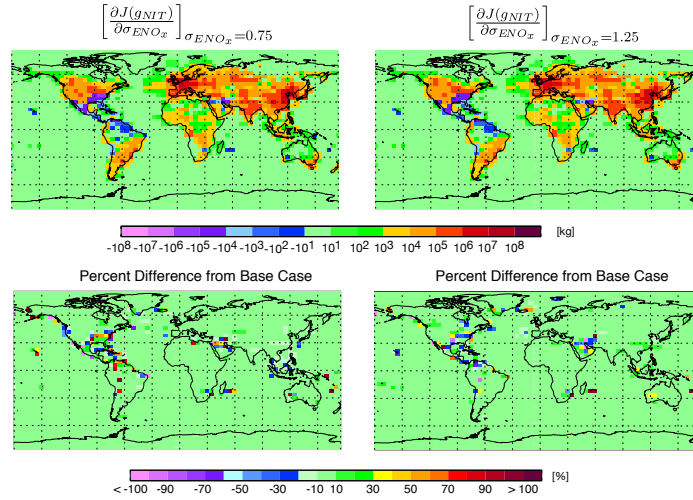
Figure 3.10: Sensitivities of nitrate aerosol to emissions of anthropogenic NO_x when the emission inventories are scaled by factors of 0.75 and 1.25, and the percent difference between these sensitivities and those calculated with the base case ($\sigma_{ENO_x}=1.0$), shown in Fig. 3.9.
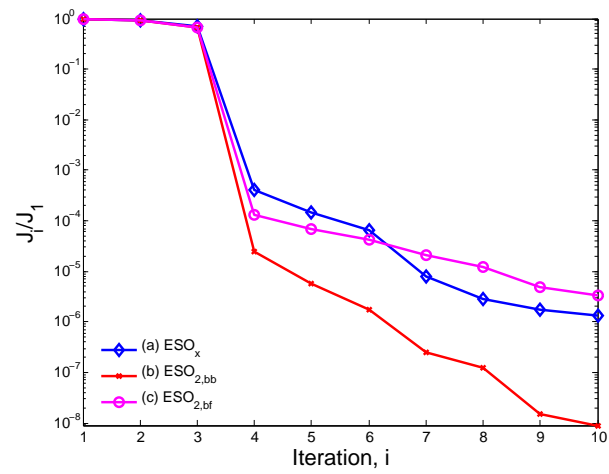
Figure 3.11: Cost function reduction for tests DA1. A uniform perturbation is applied to emission inventories of **(a)** SO_x **(b)** biomass burning SO_2 **(c)** biofuel SO_2. Complete daily measurements of sulfate aerosol are utilized for the data assimilation during a week-long simulation.
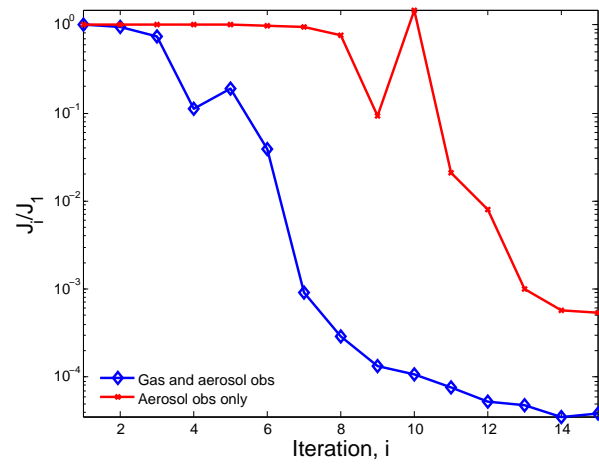
Figure 3.12: Cost function reduction for tests DA2. A uniform perturbation is applied to emission inventories of anthropogenic NH_3. Complete daily measurements of (red-crosses) ammonium aerosol and (blue-diamonds) ammonium aerosol and gas-phase O_x are utilized for the data assimilation during a week-long simulation.
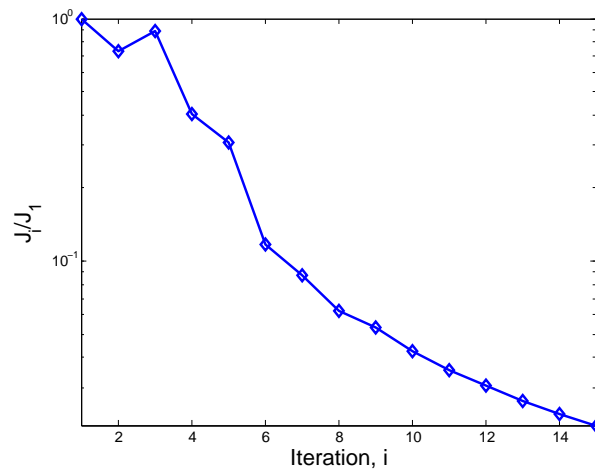
Figure 3.13: Cost function reduction for tests DA3. Emissions inventories of anthropogenic SO_x and NO_x emissions are perturbed regionally and optimized simultaneously utilizing sparse daily measurements of aerosol sulfate, ammonium, and nitrate during a week-long simulation.
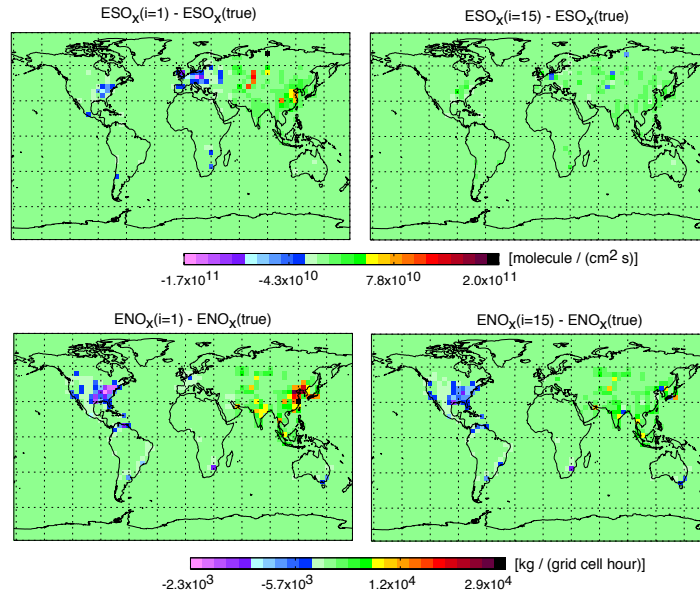
Figure 3.14: Emissions inventory estimates for test DA3. Difference between the estimated emission inventory at iteration i and the "true" inventory, which was used to generate the pseudo-observations. Results are shown for the initial estimate (left column) and after 15 iterations (right column).
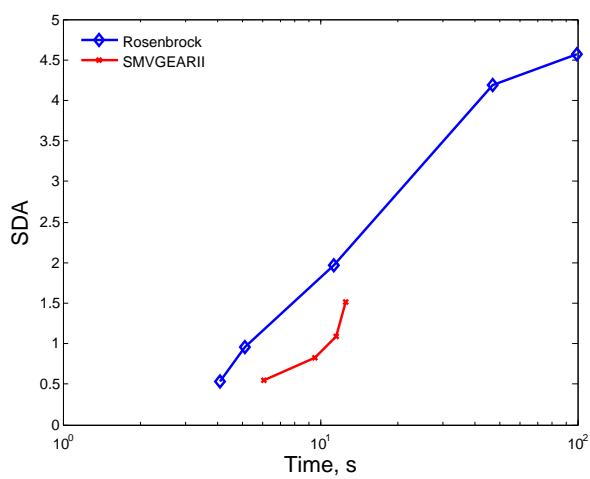
Fig. A1. Work-precision diagram for the Rosenbrock (blue circles), and SMVGEARII (red crosses) chemical solvers. Each solver is implemented in the 3-D model and run for one day using a 1 h external chemical time step. Plot shows the average time taken per external chemical time step versus the significant digits of accuracy (SDA) achieved. Tests performed using dual 1.5 GHz Itanium processors.
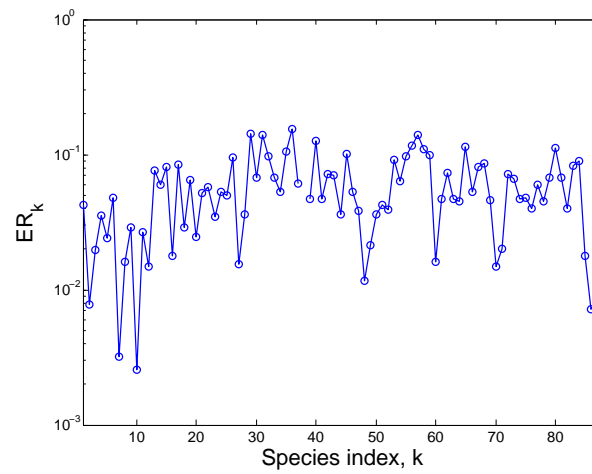
Fig. A2. Difference between the new standard GEOS-Chem simulation using the Rosenbrock solver with respect to the original GEOS-Chem solution using SMVGEARII after a week-long run. The effect of switching solvers is a ∼5–10% change in species concentrations.

Chapter 4

Source Evaluation of Secondary Inorganic Aerosol in the United States with the Adjoint of GEOS-Chem

4.1 Introduction

The persistence of airborne fine particulate matter (PM) in heavily populated areas poses a significant health hazard (*Pope*, 2000; *Pope et al.*, 2002; *WHO*, 2003). In the United States alone, it is currently estimated that 90 million people live in areas where yearly average PM concentrations exceed the National Ambient Air Quality Standards (NAAQS) (*EPA*, 2002, 2004). On average, about half of the mass of such aerosol is composed of the inorganic species sulfate (SO_4^{2-}), nitrate (NO_3^-) and ammonium (NH_4^+). Devising effective mitigation strategies for PM control requires a quantitative relationship between elevated aerosol concentrations and emissions from specific source sectors. Determining which sources to target for regulatory control is difficult, as most of this aerosol is not directly emitted; rather, it is formed in the atmosphere from gas-phase precursors via chemical and thermodynamic transformations. Precise source attribution is further complicated by transport processes that distribute these aerosols up to several thousands of kilometers from their

point of origin. Hence, formation of regulatory measures for control of PM entails having both the ability to estimate aerosol distributions and also a means of extracting source attributions from such predictions (*Marmur et al.*, 2006).

Chemical transport models are invaluable for estimating distributions of aerosol species given a set of input parameters (emissions, initial condition, etc.) and meteorological conditions, but they often suffer from substantial uncertainty in these parameters. Using inverse modeling techniques, observations can be used to improve model predictions by providing important constraints on estimates of model parameters. For example, a Kalman filter approach was used to estimate improved monthly emissions scaling factors for ammonia (NH_3) emissions over the United States using observations of ammonium wet deposition in works by *Gilliland and Abbitt* (2001) and *Gilliland et al.* (2003, 2006). *Mendoza-Dominguez and Russell* (2000, 2001) optimized domain-wide emissions scaling factors for eight species over the eastern United States using observations of gas-phase inorganic and organic species and speciated fine particles. These works provide valuable constraints on total emissions budgets. However, the inverse modeling approach used in such studies requires aggregation of emissions into large (continental scale) domains and physical approximations (neglecting transport) that may bias results or obscure important sub-domain variability.

This work presents the adjoint of a global chemical transport model (GEOS-Chem) as a tool for evaluating sources of secondary inorganic aerosol over the United States. Previous simulations over the same domain have indicated that the most difficult aspect of such predictions is estimating concentrations of NO_3^-; hence, we choose to analyze the month of January as nitrate concentrations are typically largest in the winter. The adjoint method is an efficient means of calculating model sensitivities with respect to numerous model parameters, affording optimization of these parameters on a resolution commensurate with that of the forward model itself. Sulfate and nitrate aerosol levels from surface measurements are used to constrain estimates of model control parameters by minimizing the error weighted squared difference between predictions and observations

while ensuring reasonable proximity to the prior parameter estimate. Use of a global model has the benefit of explicitly tracking transboundary influences to their sources from around the globe.

Returning to the question of aerosol source attribution, the adjoint model can be used for attainment studies (*Hakami et al.*, 2006), wherein the importance of emissions from various sectors and locations can be ranked according to their influence on aerosol concentration that are in violation of an air quality standard.

4.2 Forward model

The GEOS-Chem chemical transport model is used to estimate ambient concentrations of aerosol over the United States for the month of January, 2002. This model is driven using assimilated meteorology from the Goddard Earth Observing System (GEOS-3) of the NASA Global Modeling and Assimilation Office (GMAO). GEOS-3 data sets are downsampled to a resolution of $4° \times 5°$ to facilitate detailed simulation of tropospheric gas-phase chemistry, discussed fully in works such as *Bey et al.* (2001), *Li et al.* (2001) and *Martin et al.* (2002). The present study uses model version 6-02-05, which includes an online secondary inorganic aerosol simulation developed and described in full by *Park et al.* (2004); here we reiterate key features.

Fine mode inorganic aerosol is calculated as the mass of aerosol-phase SO_4^{2-}, NH_4^+ and NO_3^- that forms from gas-phase precursors sulfuric acid (H_2SO_4), NH_3, and nitric acid (HNO_3), as follows. H_2SO_4 is formed from oxidation of SO_2 by OH in the gas-phase, and, more importantly, by H_2O_2 and O_3 in clouds. As H_2SO_4 readily partitions into the particle phase, it is always tracked as aerosol sulfate. Thermodynamic equilibrium of aerosol NH_4^+ and NO_3^- with their gas-phase counterparts (NH_3 and HNO_3) is calculated using the MARS-A routine of *Binkowski and Roselle* (2003), which essentially allows for formation of $(NH_4)_2SO_4$ and, if excess NH_3 is available, NH_4NO_3, though formation of aerosol NO_3^- can be enhanced by cold or moist conditions. Additional important couplings

between gas and aerosol-phases treated in the model include formation of HNO_3 through heterogeneous reaction of N_2O_5 with water, where the the reaction probability is calculated as a function of aerosol type, available surface area, temperature, and relative humidity (*Evans and Jacob*, 2005).

Anthropogenic emissions of NO_x and SO_2 are taken from the Global Emission Inventory Activity (GEIA) database for the year 1985 (*Benkovitz et al.*, 1996), scaled according to fossil fuel usage for the year 1998 (*Bey et al.*, 2001). NH_3 emissions from anthropogenic (domesticated animals, fertilizers, human bodies, industry, fossil fuels) and natural (oceans, crops, soils, wild animals) sources are based on data from the 1990 GEIA inventory of *Bouwman et al.* (1997), with additional contributions owing to biomass burning and biofuel use from inventories by *Duncan et al.* (2003) and *Yevich and Logan* (2003). The total yearly source of NH_3 in the United States is scaled to match that of *Gilliland et al.* (2003), while monthly variability is calculated according to an exponential temperature scaling (*Adams et al.*, 1999). This leads to a seasonal cycle in the NH_3 emissions that is similar to, but slightly out-of-phase from, that obtained in the work by *Gilliland et al.* (2003), wherein seasonality was inferred from inverse modeling of NH_4^+.

4.3 Adjoint modeling

The GEOS-Chem model can be viewed as a numerical operator, F, acting on a vector, $\boldsymbol{c}$,

$$\boldsymbol{c}^{n+1} = F(\boldsymbol{c}^n), \tag{4.1}$$

where $\boldsymbol{c}$ is the vector of all K tracer concentrations, $\boldsymbol{c}^n=[c_1^n,\ldots,c_k^n,\ldots,c_K^n]^T$ at time step n. In practice, F comprises many individual operators representing various physical processes. For the moment we will simply let F represent a portion of the discrete forward model that advances the model vector from time step n to step $n+1$.

The adjoint of GEOS-Chem (*Henze et al.*, 2007) is used to calculate the sensitivity of a

scalar model response function, $\mathcal{J}$, with respect to the model parameters, $\boldsymbol{p}$. The response function may depend upon a specific subset of concentrations, Ω,

$$\mathcal{J} = \sum_{k,n \in \Omega} \mathcal{J}^n(c_k^n).$$

For the following derivation we will assume that the domain Ω includes all species at all times, while in practice the definition of Ω and the form of $\mathcal{J}^n$ will be application-specific.

As will become evident, it is first necessary to calculate the sensitivity of the model response with respect to the vector of species concentrations at every time step in the model,

$$\nabla_{\boldsymbol{c}^n} \mathcal{J} = \frac{\partial \mathcal{J}}{\partial \boldsymbol{c}^n}. \tag{4.2}$$

We can write the local Jacobian around any given time step as

$$\frac{\partial \boldsymbol{c}^{n+1}}{\partial \boldsymbol{c}^n} = \frac{\partial F(\boldsymbol{c}^n)}{\partial \boldsymbol{c}^n} = F_{\boldsymbol{c}}^n. \tag{4.3}$$

Using the chain rule, the right hand side of Eq. (4.2) is expanded,

$$\begin{aligned}
\nabla_{\boldsymbol{c}^n} \mathcal{J} &= (F_{\boldsymbol{c}}^n)^T (F_{\boldsymbol{c}}^{n+1})^T \cdots (F_{\boldsymbol{c}}^{N-1})^T \frac{\partial \mathcal{J}^N}{\partial \boldsymbol{c}^N} \\
&+ (F_{\boldsymbol{c}}^n)^T (F_{\boldsymbol{c}}^{n+1})^T \cdots (F_{\boldsymbol{c}}^{N-2})^T \frac{\partial \mathcal{J}^{N-1}}{\partial \boldsymbol{c}^{N-1}} + \cdots \\
&+ \frac{\partial \mathcal{J}^n}{\partial \boldsymbol{c}^n}.
\end{aligned}$$

$$\tag{4.4}$$

For the adjoint calculation, we define the adjoint variables $\boldsymbol{\lambda}_c^n = \nabla_{\boldsymbol{c}^n} \mathcal{J}$ and $\boldsymbol{\lambda}_p = \nabla_{\boldsymbol{p}} \mathcal{J}$, where the subscripts c and p indicate sensitivity with respect to $\boldsymbol{c}$ and $\boldsymbol{p}$, respectively. Initializing

$$\boldsymbol{\lambda}_c^N = \frac{\partial \mathcal{J}^N}{\partial \boldsymbol{c}^N},$$

adjoint sensitivities are found by solving the following equations iteratively from $n=N,\ldots,1$,

$$\boldsymbol{\lambda}_c^{n-1} = (F_{\mathbf{c}}^n)^T \boldsymbol{\lambda}_c^n + \frac{\partial \mathcal{J}^{n-1}}{\partial \boldsymbol{c}^{n-1}}, \tag{4.5}$$

$$\boldsymbol{\lambda}_p = (F_p^n)^T \boldsymbol{\lambda}_c^n + \frac{\partial \mathcal{J}^{n-1}}{\partial \boldsymbol{c}^{n-1}} + \boldsymbol{\lambda}_p, \tag{4.6}$$

where

$$F_p^n = \frac{\partial F^n}{\partial \boldsymbol{p}}. \tag{4.7}$$

The terms $\frac{\partial \mathcal{J}^n}{\partial \boldsymbol{c}^n}$ are referred to as the adjoint forcings. While calculation of adjoint values using this algorithm is straightforward, there are a few subtleties worth mentioning. First, evaluating sensitivities with respect to model parameters requires having first calculated sensitivities with respect to concentrations. An adjoint model used for analysis of parameters can also, by default, be used to analyze initial conditions. Secondly, while solving Eq. (4.5) iteratively along with Eq. (4.6) is not necessary, it is computationally preferable as values of $\boldsymbol{\lambda}_c^n$ need not be stored for more than a single step. Finally, it is not necessarily recommended to always derive the adjoint model directly from the numerical operator, as opposed to starting from the continuous forward model governing equations (*Sirkes and Tziperman*, 1997; *Giles and Pierce*, 2000; *Vukicevic et al.*, 2001). A full description of the derivation and validation of the adjoint model of GEOS-Chem is given in *Henze et al.* (2007).

4.4 Inverse modeling

In a Bayesian sense, inverse modeling is the process by which measurements are used to reduce the set of all possible models by rejecting those that do not likely represent the observations while also being consistent with prior information (*Tarantola*, 2006). A range

of models is typically constructed using control parameters,

$$\boldsymbol{\sigma} = [\sigma^1, \sigma^2, \ldots, \sigma^M]^T,$$

which are used to adjust elements of the model parameters, $\boldsymbol{p}$, when applied as exponential scaling factors,

$$p = p_a e^{\sigma},$$

where $\boldsymbol{p}_a$ is the prior parameter estimate. The inverse problem seeks $\boldsymbol{\sigma}$ that minimizes the cost function, J, given by

$$J = \frac{1}{2} \sum_{\boldsymbol{c} \in \Omega} (H\boldsymbol{c} - \boldsymbol{c}_{obs})^T \mathbf{S}_{obs}^{-1} (H\boldsymbol{c} - \boldsymbol{c}_{obs}) + \frac{1}{2} \gamma_r (\boldsymbol{\sigma} - \boldsymbol{\sigma}_a)^T \mathbf{S}_{\sigma}^{-1} (\boldsymbol{\sigma} - \boldsymbol{\sigma}_a), \qquad (4.8)$$

where $\boldsymbol{c}$ is the vector of species concentrations mapped to the observation space by H, $\boldsymbol{c}_{\text{obs}}$ is the vector of species observations, $\mathbf{S}_{\text{obs}}$ is the observation error covariance matrix, $\boldsymbol{\sigma}_a$ is the prior estimate of the parameter scaling factors (equal to 0), $\mathbf{S}_{\sigma}$ is the error covariance estimate of the parameter scaling factors, γ_r is a regularization parameter, and Ω is the domain (in time and space) over which observations and model predictions are available. This form of a cost function is rigorously optimal from the perspective of Bayesian analysis if the operators F and H are linear and the distributions of $\boldsymbol{c}_{obs}$ and p_a are Gaussian. For our purposes, we adopt this form for J in an ad hoc fashion, as the operators are nonlinear, and introduce γ_r to compensate for lack of precise statistical knowledge of the distributions of $\boldsymbol{c}_{obs}$ and p_a.

The adjoint model is used to calculate the gradient of the cost function with respect to the parameter scaling factors, $\nabla_{\boldsymbol{\sigma}} J$. Using the variational approach, these gradients are supplied to an optimization routine (the quasi-Newton L-BFGS-B optimization routine (*Byrd et al.*, 1995; *Zhu et al.*, 1994)) and the minimum of the cost function is sought iteratively. At each iteration, improved estimates of the model parameters are implemented

and the forward model solution is recalculated.

4.4.1 Measurements

We compare predictions of sulfate and nitrate aerosol levels to observations from the Inter-agency Monitoring of Protected Visual Environments (IMPROVE) network (*Malm et al.*, 1994) during the month of January, 2002. Mass concentrations of sulfate and nitrate are determined from analysis of fine aerosol (aerodynamic diameter less than 2.5 μm) collected on teflon and nylon filters, respectively, sampled over a 24 h period every third day. Figure 4.1 compares the predicted and observed monthly average aerosol sulfate and nitrate, where the observations from individual IMPROVE sites have been averaged to lie on the GEOS-Chem model grid. While estimates of sulfate concentrations are reasonable, the nitrate simulation shows significant discrepancy with observations, similar to previous studies (*Park et al.*, 2004, 2006; *Liao et al.*, 2007). Comparisons for nitrate aerosol are potentially biased, particularly in the southwestern United States, where a portion of measured nitrate may come from uptake on dust particles, which is a source of nitrate aerosol not considered in the model.

The observation error covariance, $\mathbf{S}_{obs}$, includes contributions from instrumental and representational sources of error. Instrument error is generally small, within a few percent. The representational error, on the other hand, is likely significant owing to the low resolution of the model. To estimate the magnitude of such error, two model simulations are performed, one at the base case of $4° \times 5°$ resolution and one at $2° \times 2.5°$ resolution, each with the same initial conditions and emissions inventories. Samples of the estimated daily average nitrate aerosol concentrations at the surface are shown in Fig. 4.2. The difference between the aerosol concentration in a given grid cell from the coarse simulation, panel (a), and the average concentration over the same $4° \times 5°$ domain from the fine simulation, panel (b), is used to estimate the representational error, shown in panel (c). Although still only a crude approximation of the representational error, it is likely better than assuming

a single value for all observations, as representational errors depend upon the variability in the aerosol distribution. For example, consider the two grid cells at 38° N where concentrations of nitrate peak (yellow) in the coarse simulation. The cell on the right actually encompasses a region of considerable variability, at least as estimated by the $2° \times 2.5°$ model calculation, hence the representational error assigned to observations in this location are substantial (43%) for this day. In contrast, the cell on the left encompasses a much more consistent distribution, hence the representational error there is smaller (23%).

4.4.2 Model parameters

The set of variable parameters includes scaling factors for emissions of SO_x, NO_x and NH_3 from several specific sectors listed in Table 4.1. Also considered are scaling factors for the initial concentrations of each tracer (initial conditions) and for several kinetic parameters, such as the heterogeneous reaction probability for formation of HNO_3 from N_2O_5, which is an important (*Dentener and Crutzen*, 1993), yet still highly uncertain (*Brown et al.*, 2006), mechanism for loss of NO_x.

After a single evaluation of $\nabla_{\boldsymbol{p}} J$ using the adjoint model, it is clear which variable parameters are the most influential. Figure 4.3 shows the sensitivity of the cost function with respect to anthropogenic emissions of SO_x, NO_x and NH_3, and natural emissions of NH_3. These sectors have the largest sensitivities of all those considered. It is worth noting that the sensitivities are not necessarily largest where sources or prediction error is largest.

Sensitivities with respect to sources outside North America are generally several orders of magnitude smaller than those shown in Fig. 4.3. The only sensitivity with an appreciable magnitude on another continent is that of anthropogenic surface SO_x emissions (see panel (a) of Fig. 4.4), which is widely positive, suggesting that import to the United States of sulfate aerosol in the model is possibly too high. This could come from overestimates of exterior SO_x emissions, or from over-efficient transport / production of sulfate from SO_2. Figure 4.4 shows sensitivities with respect to initial conditions, displaying just the

values at 900 hPa. Values peak in the 900-700 hPa range (initial concentrations closer to the surface are less influential owing to quick depositional losses), but are still one to two orders of magnitude smaller than emissions sensitivities over the course of the simulation, as the average aerosol lifetime is much shorter than one month. Figure 4.5 shows the column integrated sensitivity with respect to the heterogeneous reaction probability for formation of HNO_3. The distribution is similar to that for the sensitivity with respect to NO_x emissions, both of which are positive in regions where nitrate aerosol was over-estimated (Midwest) and negative in regions where nitrate was under-estimated (Southwest). However, these sensitivities are again much lower than those with respect to emissions, even if they are integrated over the horizontal model domain.

4.4.3 Optimization

The adjoint sensitivities are used as gradients to minimize the cost function. To simplify the optimization process, we do not allow the scaling factors for initial conditions or kinetic parameters to be variable, as, assuming all are equally uncertain, they were found to be much less critical than emissions parameters. Each emissions inventory is ascribed a standard 100% uncorrelated error. The significance of the prior information is thus more of a smoothness constraint than a rigorous estimate of prior uncertainty (*Rodgers*, 2000). Figure 4.6 shows the difference between predictions and observations of aerosol nitrate and sulfate after only eight iterations, by which point the cost function has been reduced by 40%. Note that for displaying the discrepancy between predictions and observations, Figs. 4.1 and 4.6 show the difference between the monthly average concentrations; however, it is the differences between the 24 h concentrations on every third day (the frequency of the observations) that are actually used as forcing for the adjoint model. Estimates of sulfate aerosol still lie within a reasonable range (± 1 μg) of the observations. Estimates of nitrate aerosol have been substantially improved, particularly in the central Midwest.

4.4.4 Analysis of posterior emissions

The real interest lies the changes in the model parameters that minimize the cost function. The prior and posterior emissions estimates are shown in Fig. 4.7, again displaying only emissions from sectors with significant differences from the prior estimates. Surface emissions of anthropogenic NO_x also increased in the Southwest and southern part of California, while SO_x emission were generally reduced rather uniformly. The largest adjustments were made to the NH_3 inventories, with large reduction in the central Midwest, and an increase in anthropogenic NH_3 in the Southwest. The total magnitude of the posterior natural and anthropogenic NH_3 emissions in the United States is 0.10 Tg / month, reduced by 12% from the prior inventory. Beginning with a similar prior estimate (*Park et al.*, 2004), *Gilliland et al.* (2006) found a 40% reduction in NH_3 emissions, including northern Mexico and southern Canada. The scaling from the present study is likely less severe owing to the flexibility of the adjoint calculation to redistribute NH_3 emissions rather than simply scale the entire domain magnitude up or down. Similar comparisons between aggregated vs resolved inversions were found in *Kopacz et al.* (submitted).

A conclusion of *Gilliland et al.* (2006) was that observations of wet NH_4^+ (i.e., dissolved NH_3 and aerosol NH_4^+) were required to constrain NH_3 emissions unless the sulfate and nitrate budgets were verified. Here we have in essence taken the opposite approach, having based the inversion on observations of sulfate and nitrate; hence it is reasonable to inquire how the estimates of NH_4^+ using the posterior emissions compare with observations. We compare measurements of NH_4^+ from the CASTNet network (*Baumgardner et al.*, 2002) using both the prior emissions inventory and the posterior inventory obtained using the optimized scaling factors from Section 4.4.3. Figure 4.8 shows the comparison. Estimates of NH_4^+ are largely improved throughout the Midwest. Posterior estimates of NH_4^+ now also capture the observed increase in the Southwest that was entirely missing from the prior estimate, though the model now overestimates the magnitude of this maximum. Nevertheless, that both NH_4^+ and NO_3^- concentrations were being originally underpredicted

in this area implies that the source of error is related to both, which would not be explained by biases in the observations owing to coarse mode NO_3^- from uptake on dust.

4.5 Attainment

The model response can be defined as a metric of non-attainment for ambient aerosol concentrations,

$$J_a = \frac{1}{2} \sum_{i,j \in \text{U.S.}} \theta(a_{i,j}) a_{i,j}^2, \tag{4.9}$$

where

$$a_{i,j} = \left(\sum_{\hat{k}} 24 h c_{i,j,\hat{k}} \right) - \gamma_a, \quad \hat{k} = \{SO_4^{2-}, NO_3^-, NH_4^+\},$$

and

$$\theta(a) = \begin{cases} 0 & a \leq 0 \\ 1 & a > 0 \end{cases}$$

Use of the L_2 norm emphasizes peak concentrations, which are of most concern for air quality. The air quality threshold is γ_a, here taken to have a value of 10 μg m^{-3}. Although this threshold is lower than the actual yearly NAAQS standard of 15 μg m^{-3}, here we have not included carbonaceous aerosol in the set of active species, $\hat{k}$.

4.5.1 Regional variability

First we consider the month of July, 2001. The contribution of ammonium aerosol to non-attainment, as defined by Eq. (4.9), is significant. Figure 4.9, panel (a), shows the average ammonium concentrations in locations where $\theta(a) = 1$. Figure 4.9, panel (b), shows the emissions of NH$_3$, plotted as percentages of the total emissions of NH$_3$ throughout the United States.

The sensitivities of J_a with respect to the various emissions inventories listed in Table

4.1 are calculated with a single evaluation of the forward and adjoint models. Shown in Fig. 4.9 are the sensitivities of J_a with respect to emissions of NH_3, also reported as a normalized percentage,

$$\left(\frac{\partial J_a}{\partial p_{ENH_3}} \right) \left(\frac{p_{ENH_3}}{J_a} \right) \times 100\%.$$

These sensitivities give the linear approximation of the percent reduction in non-attainment achieved per percent reduction in emissions. For example, a non-attainment sensitivity of 10% in a particular location would imply that reducing emissions by 25% in that location would cause a 2.5% reduction in J_a. When $\gamma_a > 0$, these sensitivities can exceed 100%.

The combination of the three plots in Fig. (4.9) provides a clear and quantitative way of analyzing air quality attainment (*Hakami et al.*, 2006). The distribution of non-attainment, panel (a), shows locations that will benefit from implementation of emissions regulations that enforce air quality attainment. The distribution of the emissions, panel (b), shows the areas most heavily burdened by any simple emissions abatement strategy, while the adjoint sensitivities in panel (c) show locations where reducing emissions would actually be the most effective towards achieving air quality attainment. While a bulk of the NH_3 is emitted in the northern midwest, it is the sources of NH_3 that are collocated with sources of SO_x further east that ultimately contribute most substantially to the non-attainment regions. The disparity between these three plots concisely depicts the challenge in designing regulation measures to control long-lived pollutants.

4.5.2 Seasonal variability

Next we consider the seasonal dependence of the non-attainment sensitivities. Figure 4.10 shows the normalized sensitivities of J_a, evaluated in July, with respect to emissions of anthropogenic NH_3 and SO_x. This comparison indicates that the most effective way to reduce non-attainment during this month is reduction of anthropogenic stack SO_x emissions. During the summer, reducing NH_3 will reduce NH_4^+, and, hence, NO_3^-. However, most NH_3 is in the form of $(NH_4)_2SO_4$, so simply reducing the available sulfate is more effective.

Conversely, plots in the second row show sensitivities of J_a with respect to the same emissions sectors, this time evaluated for the month of April. During the spring, the efficacy of reducing emissions of NH_3 vs SO_x has now entirely reversed, with NH_3 emissions being the most effective target for abatement. In January, using the posterior emissions from Section 4.4.3, there are only a few instances of non-attainment; reducing NH_3 emissions during this month is hence very efficient.

The nonlinear relationship between sulfate and total PM mass has been noted previously (*West et al.*, 1999; *Vayenas et al.*, 2005) to reduce effectiveness of SO_x control in colder seasons. Reducing SO_x, and hence sulfate, is rendered ineffective owing to rapid replacement of SO_4^{2-} by NO_3^-, formation of the latter being favored by colder temperatures. In the present work, we find this effect to be so extreme that during the winter, the non-attainment sensitivity of SO_x emissions has a sensitivity near the non-attainment region that is actually negative. If removal of sulfate aerosol in the presence of fixed total ammonia and nitric acid concentrations causes an entire mole of $(NH_4)_2SO_4$ to be replaced by two moles of $(NH_4)NO_3$, then the total PM mass would be enhanced by decreases in SO_x emissions.

4.5.3   Long-range influences

It is interesting to note that despite the obvious influence of transport in the differences between locations of non-attainment and locations of non-attainment sensitivities, nearly all emissions sensitivities still lie within the contiguous United States. For example, panel (a) in Fig. 4.11 shows the non-attainment sensitivities with respect to anthropogenic surface emissions of SO_x throughout the Northern Hemisphere, where J_a is defined as in Eq. (4.9). As noted, nearly all values are located within the Eastern United States. However, if we consider the following cost function,

$$J_{a,\infty} = \sum_{i,j \in \text{U.S.}} \theta(a_{i,j}) a_{i,j}, \tag{4.10}$$

where $\gamma_a = 0$, then the sensitivities with respect to $J_{a,\infty}$ simply show which emissions are influencing aerosol concentrations in the United States, at any level. These values (panel (b)) shown significantly more distributed sensitivities. Even some influence from Eastern China is apparent. The conclusion is that while transcontinental sources of inorganic fine aerosol do not (yet) appreciably influence aerosol concentrations in the United States from an air quality standpoint, they are of importance for determining background levels, which is of concern for visibility-related regulations (*Park et al.*, 2004, 2006).

Finally, it should be reiterated that conclusions drawn from analysis of the adjoint sensitivities, which are inherently linear, are valid only over ranges of modest changes ($\sim$25%) in the emissions inventory (*Henze et al.*, 2007, see examples therein).

4.6   Conclusions

The adjoint of the chemical transport model GEOS-Chem (*Henze et al.*, 2007) has been applied to evaluate sources of secondary inorganic aerosol throughout the United States. In an effort to analyze the model's abilities to estimate nitrate aerosol (previously noted to be relatively poor (*Park et al.*, 2004, 2006; *Liao et al.*, 2007)), comparisons are made to measurements of sulfate (SO_4^{2-}) and nitrate (NO_3^-) aerosol from the IMPROVE network of monitoring stations (*Malm et al.*, 1994) during the month of January, 2002. Significant discrepancies initially exist for estimates of NO_3^-. The adjoint model is used to select variable model parameters that most significantly (to a linear approximation) influence this discrepancy. Parameters initially considered include scaling factors for emissions of SO_x, NO_x, and NH_3 from several sectors, initial conditions of all tracers, and a few uncertain heterogeneous reaction probabilities. Not surprisingly, anthropogenic emissions of NH_3 were found to be most influential, followed by natural emissions of NH_3, anthropogenic stack emission of SO_x, and surface emissions of NO_x.

The adjoint model has been used in an inverse modeling framework to constrain the estimates of emissions inventories. The posterior emissions inventories show modest changes

to budgets of NO_x and SO_x, with significant changes to NH_3 emissions, which are largely reduced in the East and Midwest, and increased in the Southwest. Despite this simple summary, there is considerable variability in the spatial distributions of the differences between the prior and posterior emissions. The locations of most significant differences are not coincident with areas of largest initial discrepancies, underlining the importance of transport and subsequent chemical and thermodynamic transformations that govern formation of secondary inorganic aerosol.

Observations of NH_4^+ from CASTNet stations are used as an independent assessment of the posterior emissions. The observations showed discrepancies with predictions using the prior emissions that were generally similar to that for nitrate aerosol, which would imply that such differences are not simply a bias in one set of measurements owing to artifacts such as HNO_3 from dust or evaporation / contamination of filter samples. Using the posterior inventory improves model performance, largely in the Midwest, with some overcorrection in the Southwest. The magnitude of the total adjustment to the NH_3 inventory is similar to that found in *Gilliland et al.* (2006), though considerable regional variability is found in the present study. Overall, inverse modeling constraints from measurements of aerosol SO_4^{2-} and NO_3^- are shown to be a promising approach towards constraining estimates of NH_3 emissions, which have been recognized as highly uncertain and difficult to assess directly (*Pinder et al.*, 2006; *Nowak et al.*, 2006). In the future, this approach will be applied to further constrain emissions inventories during other seasons, and to analyze additional sources of model uncertainty.

The adjoint model provides a convenient means of assessing which emissions are predicted to most influence air quality non-attainment of fine particulate matter. The disparity between locations of peak emissions, regions of non-attainment and locations of the non-attainment sensitivities, highlights the importance of transport in the formation of these types of aerosols from their gas-phase precursors, and the complications that thus arise when devising local control strategies for air quality attainment. In July, the inventory

with the largest non-attainment reducing efficacy is emission of SO_x from anthropogenic (stack) sources. Conversely, in April, SO_x controls are rendered ineffective owing to displacement by NO_3^-, favored in cooler seasons. This leaves NH_3 as the most effective means of particulate control in the spring. This finding is in qualitative agreement with the work of *Takahama et al.* (2004), which pointed out the importance of NH_3 control for air quality in Pittsburgh, and the findings of *Pinder et al.* (2007), who reached similar conclusions based upon economic considerations of the cost to reduce aerosol concentrations in select locations on the East Coast. Long-range influence is found to be minimal for estimates of air quality in January, April and July. However, there is some influence in background concentrations. As noted in previous works by *Park et al.* (2004, 2006), such influence, while small, could have important consequences for attainment of regional haze goals. The benefit of the adjoint model is that estimates of such efficacies are easily calculated over a range of conditions for minimal computational cost, each time providing detailed evaluations throughout the model parameter space.

Bibliography

Adams, P. J., J. H. Seinfeld, and D. M. Koch, Global concentrations of tropospheric sulfate, nitrate, and ammonium aerosol simulated in a general circulation model, *J. Geophys. Res.*, *104*(D11), 13,791–13,823, 1999.

Baumgardner, R. E., T. F. Lavery, C. M. Rogers, and S. S. Isil, Estimates of the atmospheric deposition of sulfur and nitrogen species: Clean air status and trends network, 1990-2000, *Environ. Sci. Technol.*, *36*(12), 2614–2629, 2002.

Benkovitz, C. M., M. T. Scholtz, J. Pacyna, L. Tarrason, J. Dignon, E. C. Voldner, P. A. Spiro, J. A. Logan, and T. E. Graedel, Global gridded inventories of anthropogenic emissions of sulfur and nitrogen, *J. Geophys. Res.*, *101*(D22), 29,239–29,253, 1996.

Bey, I., et al., Global modeling of tropospheric chemistry with assimilated meteorology: Model description and evaluation, *J. Geophys. Res.*, *106*(D19), 23,073–23,095, 2001.

Binkowski, F. S., and S. J. Roselle, Models-3 community multiscale air quality (CMAQ) model aerosol component - 1. Model description, *J. Geophys. Res.*, *108*(D6), 4183, doi: 10.1029/2001JD001409, 2003.

Bouwman, A. F., D. S. Lee, W. A. H. Asman, F. J. Dentener, K. W. VanderHoek, and J. G. J. Olivier, A global high-resolution emission inventory for ammonia, *Global Biogeochem. Cy.*, *11*(4), 561–587, 1997.

Brown, S. S., et al., Variability in nocturnal nitrogen oxide processing and its role in regional air quality, *Science*, *311*(5757), 67–70, 2006.

Byrd, R. H., P. Lu, J. Nocedal, and C. Zhu, A limited memory algorithm for bound constrained optimization, *Scientific Computing*, *16*(5), 1190–1208, 1995.

Dentener, F. J., and P. J. Crutzen, Reaction of N_2O_5 on tropospheric aerosols - impact on the global distributions of NO_x, O_3, and OH, *J. Geophys. Res.*, *98*(D4), 7149–7163, 1993.

Duncan, B. N., R. V. Martin, A. C. Staudt, R. Yevich, and J. A. Logan, Interannual and seasonal variability of biomass burning emissions constrained by satellite observations, *J. Geophys. Res.*, *108*(D2), 2003.

Environmental Protection Agency (EPA), Latest findings on national air quality: 2001 status and trends, *Tech. rep.*, 2002.

Environmental Protection Agency (EPA), TTN NAAQS - PM2.5 - NAAQS Implementation PM2.5 Designations - Technical Information, 2004.

Evans, M. J., and D. J. Jacob, Impact of new laboratory studies of N_2O_5 hydrolysis on global model budgets of tropospheric nitrogen oxides, ozone, and OH, *Geophys. Res. Lett.*, *32*(9), 2005.

Giles, M. B., and N. A. Pierce, An introduction to the adjoint approach to design, *Flow Turbul. Combust.*, *65*, 393–415, 2000.

Gilliland, A., and P. J. Abbitt, A sensitivity study of the discrete Kalman filter (DKF) to initial condition discrepancies, *J. Geophys. Res.*, *106*(D16), 17,939–17,952, 2001.

Gilliland, A. B., R. L. Dennis, S. J. Roselle, and T. E. Pierce, Seasonal NH_3 emission estimates for the eastern United States based on ammonium wet concentrations and an

inverse modeling method, *J. Geophys. Res.*, *108*(D15), 4477, doi:10.1029/2002JD003063, 2003.

Gilliland, A. B., K. W. Appel, R. W. Pinder, and R. L. Dennis, Seasonal NH$_3$ emissions for continental United States: Inverse model estimation and evaluation, *Atmos. Environ.*, *40*, 4986–4998, 2006.

Hakami, A., J. H. Seinfeld, T. F. Chai, Y. H. Tang, G. R. Carmichael, and A. Sandu, Adjoint sensitivity analysis of ozone nonattainment over the continental united states, *Environ. Sci. Technol.*, *40*(12), 3855–3864, 2006.

Henze, D. K., A. Hakami, and J. H. Seinfeld, Development of the adjoint of GEOS-Chem, *Atmos. Chem. Phys.*, *7*, 2413–2433, 2007.

Kopacz, M., D. Jacob, D. K. Henze, C. L. Heald, D. G. Streets, and Q. Zhang, A comparison of analytical and adjoint Bayesian inversion methods for constraining Asian sources of CO using satellite (MOPITT) measurements of CO columns, submitted.

Li, Q. B., et al., A tropospheric ozone maximum over the middle east, *Geophys. Res. Lett.*, *28*(17), 3235–3238, 2001.

Liao, H., D. K. Henze, J. H. Seinfeld, S. Wu, and L. J. Mickley, Biogenic secondary organic aerosol over the United States: Comparison of climatological simulations with observations, *J. Geophys. Res.*, *112*, D06201, doi:10.1029/2006JD007813, 2007.

Malm, W. C., J. F. Sisler, D. Huffman, R. A. Eldred, and T. A. Cahill, Spatial and seasonal trends in particle concentration and optical extinction in the United States, *J. Geophys. Res.*, *99*(D1), 1347–1370, 1994.

Marmur, A., S. K. Park, J. A. Mulholland, P. E. Tolbert, and A. G. Russell, Source apportionment of PM$_{2.5}$ in the southeastern United States using receptor and emissions-

based models: Conceptual differences and implications for time-series health studies, *Atmos. Environ.*, *40*(14), 2533–2551, 2006.

Martin, R. V., et al., Interpretation of toms observations of tropical tropospheric ozone with a global model and in situ observations, *J. Geophys. Res.*, *107*(D18), 4351, 2002.

Mendoza-Dominguez, A., and A. G. Russell, Iterative inverse modeling and direct sensitivity analysis of a photochemical air duality model, *Environ. Sci. Technol.*, *34*(23), 4974–4981, 2000.

Mendoza-Dominguez, A., and A. G. Russell, Emission strength validation using four-dimensional data assimilation: application to primary aerosol and precursors to ozone and secondary aerosol, *J. Air & Waste Manage. Assoc*, *51*, 1538–1550, 2001.

Nowak, J. B., et al., Analysis of urban gas phase ammonia measurements from the 2002 Atlanta Aerosol Nucleation and Real-Time Characterization Experiment (ANARChE), *J. Geophys. Res.*, *111*(D17), D17308, doi:10.1029/2006JD007113, 2006.

Park, R. J., D. Jacob, B. D. Field, R. Yantosca, and M. Chin, Natural and transboundary pollution influences on sulfate-nitrate-ammonium aerosols in the United States: implications for policy, *J. Geophys. Res.*, *109*, D15204, doi:10.1029/2003JD004473, 2004.

Park, R. J., D. J. Jacob, N. Kumar, and R. M. Yantosca, Regional visibility statistics in the United States: Natural and transboundary pollution influences, and implications for the Regional Haze Rule, *Atmos. Environ.*, *40*(28), 5405–5423, 2006.

Pinder, R. W., P. J. Adams, S. N. Pandis, and A. B. Gilliland, Temporally resolved ammonia emission inventories: Current estimates, evaluation tools, and measurement needs, *J. Geophys. Res.*, *111*(D16), 2006.

Pinder, R. W., P. J. Adams, and S. N. Pandis, Ammonia emission controls as a cost-effective

strategy for reducing atmospheric particulate matter in the eastern United States, *Environ. Sci. Technol.*, *41*(2), 380–386, 2007.

Pope, C. A., Review: Epidemiological basis for particulate air pollution health standards, *Aerosol. Sci. Tech.*, *32*(1), 4–14, 2000.

Pope, C. A., R. T. Burnett, M. J. Thun, E. E. Calle, D. Krewski, K. Ito, and G. D. Thurston, Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution, *J. A. M. A.*, *287*(9), 1132–1141, 2002.

Rodgers, C. D., *Inverse Methods for Atmospheric Sounding*, *Series on Atmospheric, Oceanic and Planetary Physics*, vol. 2, World Scientific, Singapore, 2000.

Sandu, A., W. Liao, G. R. Carmichael, D. K. Henze, and J. H. Seinfeld, Inverse modeling of aerosol dynamics using adjoints: Theoretical and numerical considerations, *Aerosol. Sci. Tech.*, *39*(8), 677–694, 2005.

Sirkes, Z., and E. Tziperman, Finite difference of adjoint or adjoint of finite difference?, *Mon. Weather Rev.*, *125*(12), 3373–3378, 1997.

Takahama, S., A. E. Wittig, D. V. Vayenas, C. I. Davidson, and S. N. Pandis, Modeling the diurnal variation of nitrate during the Pittsburgh Air Quality Study, *J. Geophys. Res.*, *109*(D16), D16S06, doi:10.1029/2003JD004149, 2004.

Tarantola, A., Popper, Bayes and the inverse problem, *Nat. Physics*, *2*(8), 492–494, 2006.

World Health Organization (WHO), Health aspects of air pollution with particulate matter, ozone, and nitrogen dioxide, Rep. EUR/03/5042688, Bonn, 2003.

Vayenas, D. V., S. Takahama, C. I. Davidson, and S. N. Pandis, Simulation of the thermodynamics and removal processes in the sulfate-ammonia-nitric acid system during winter: Implications for $PM_{2.5}$ control strategies, *J. Geophys. Res.*, *110*(D7), D07S14, doi:10.1029/2004JD005038., 2005.

Vukicevic, T., M. Steyskal, and M. Hecht, Properties of advection algorithms in the context of variational data assimilation, *Mon. Weather Rev.*, *129*(5), 1221–1231, 2001.

West, J. J., A. S. Ansari, and S. N. Pandis, Marginal $PM_{2.5}$: Nonlinear aerosol mass response to sulfate reductions in the Eastern United States, *J. Air Waste Manage. Assoc.*, *49*(12), 1415–1424, 1999.

Yevich, R., and J. A. Logan, An assessment of biofuel use and burning of agricultural waste in the developing world, *Global Biogeochem. Cy.*, *17*(4), 2003.

Zhu, C., R. H. Byrd, P. Lu, and J. Nocedal, L-BFGS-B: a limited memory FORTRAN code for solving bound constrained optimization problems, *Tech. rep.*, Northwestern University, 1994.

Table 4.1: Emissions inventories treated as variable parameters.

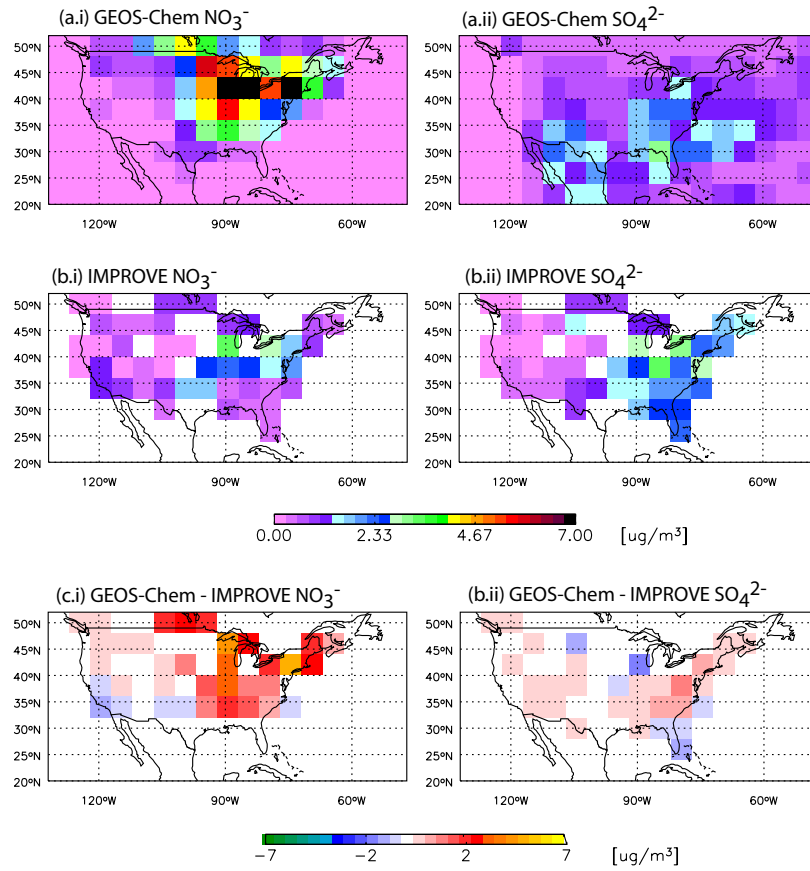| Emitted species	Source sectors considered
SO_x	surface anthropogenic, stack anthropogenic, ships, biomass burning, biofuel
NH_3	anthropogenic, natural, biomass burning, biofuel
NO_x	surface anthropogenic, stack anthropogenic, lightning

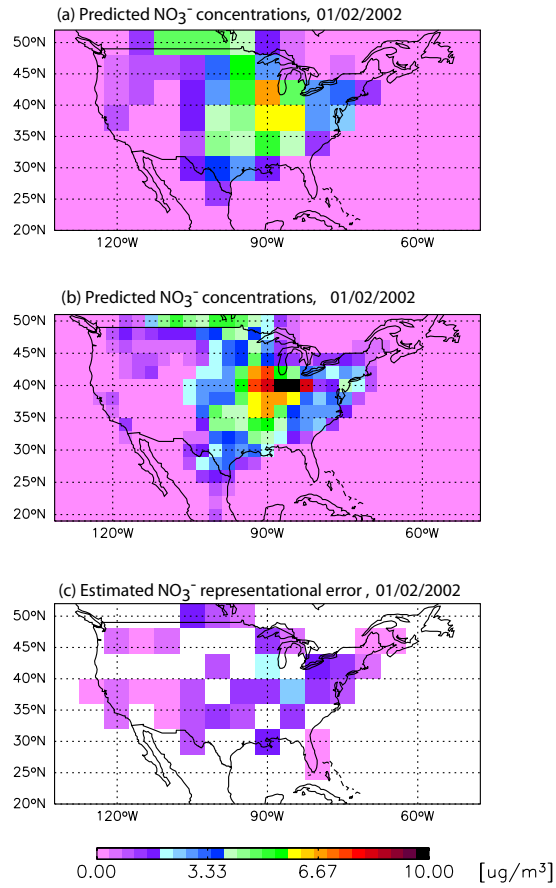Figure 4.1: Predicted (GEOS-Chem) vs observed (IMPROVE) NO_3^- and SO_4^{2-} for the month of January, 2002.

Figure 4.2: Representational error estimate for aerosol nitrate (NO_3^-). Panels (a) and (b) show model-predicted average nitrate concentrations at the surface for January 2, 2002, calculated at $4° \times 5°$ and $2° \times 2.5°$ resolutions, respectively. These are used to estimate the representational error for this day at the IMPROVE sites, panel (c).
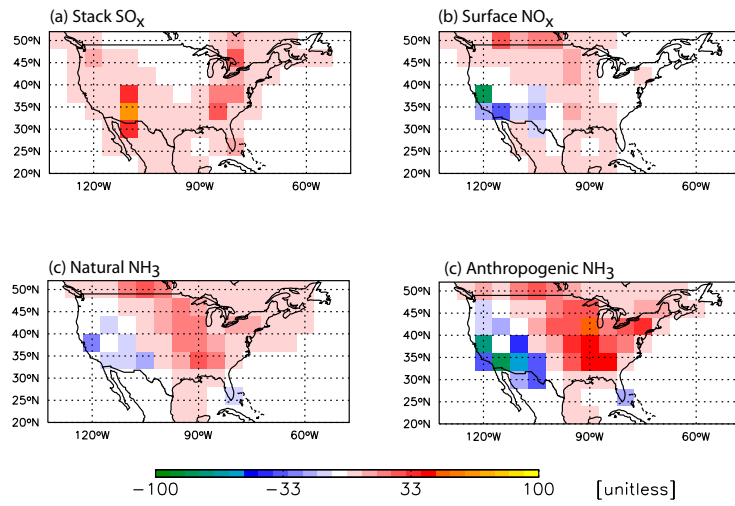
Figure 4.3: Sensitivity of the cost function with respect to emissions scaling factors for the following inventories: (a) stack SO_x, (b) surface NO_x, (c) natural NH_3, and (d) anthropogenic NH_3.
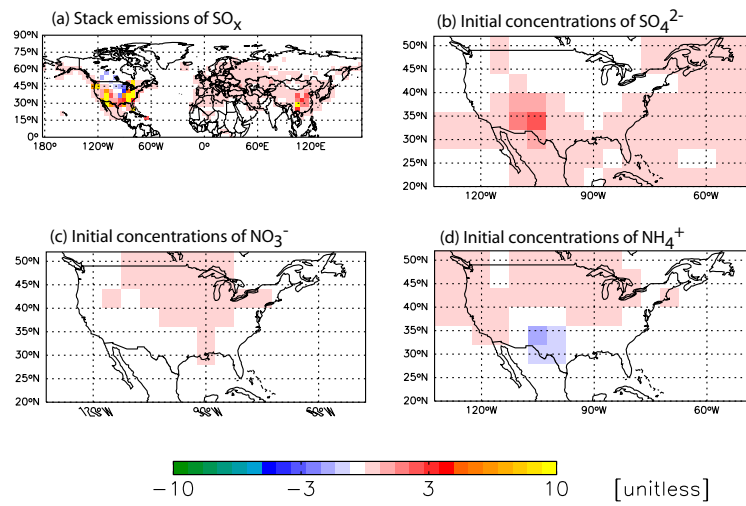
Figure 4.4: Sensitivity of the cost function with respect to (a) stack SO_x (which saturates on this plot scale over the United States, see Fig. 4.3 for detail), (b) initial concentrations of SO_4^{2-}, (c) initial concentrations of NO_3^-, and (d) initial concentrations of NH_4^+.
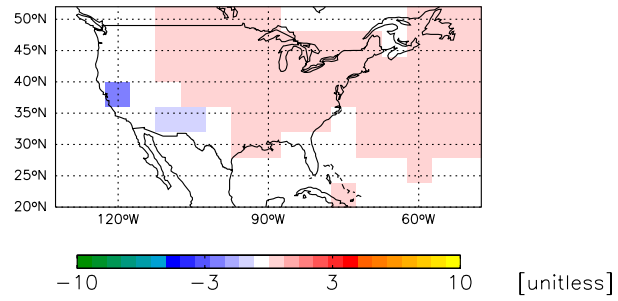
Figure 4.5: Sensitivity of the cost function with respect to the reaction probability for heterogeneous N_2O_5 hydrolysis.
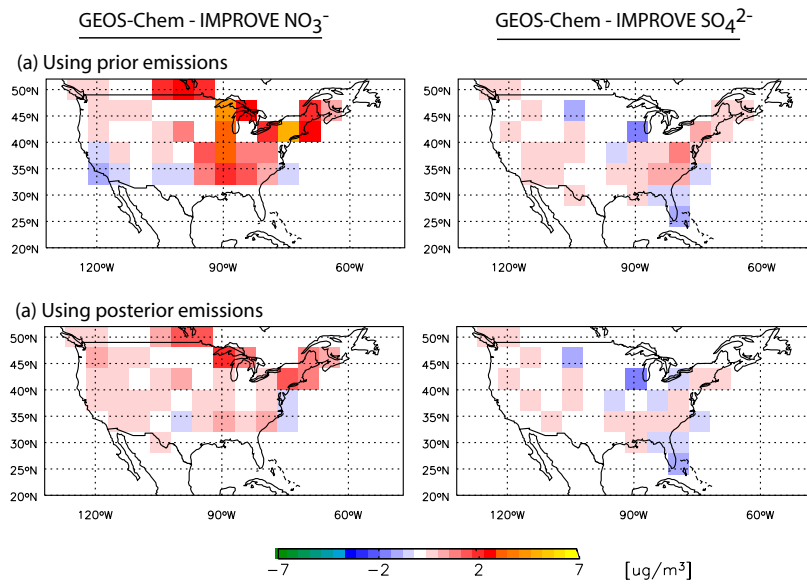
Figure 4.6: Difference between predicted (GEOS-Chem) and observed (IMPROVE) NO$_3^-$ and SO$_4^{2-}$ for the month of January, 2002. Predictions are calculated using (a) prior and (b) posterior emissions inventories (see Fig. 4.7).
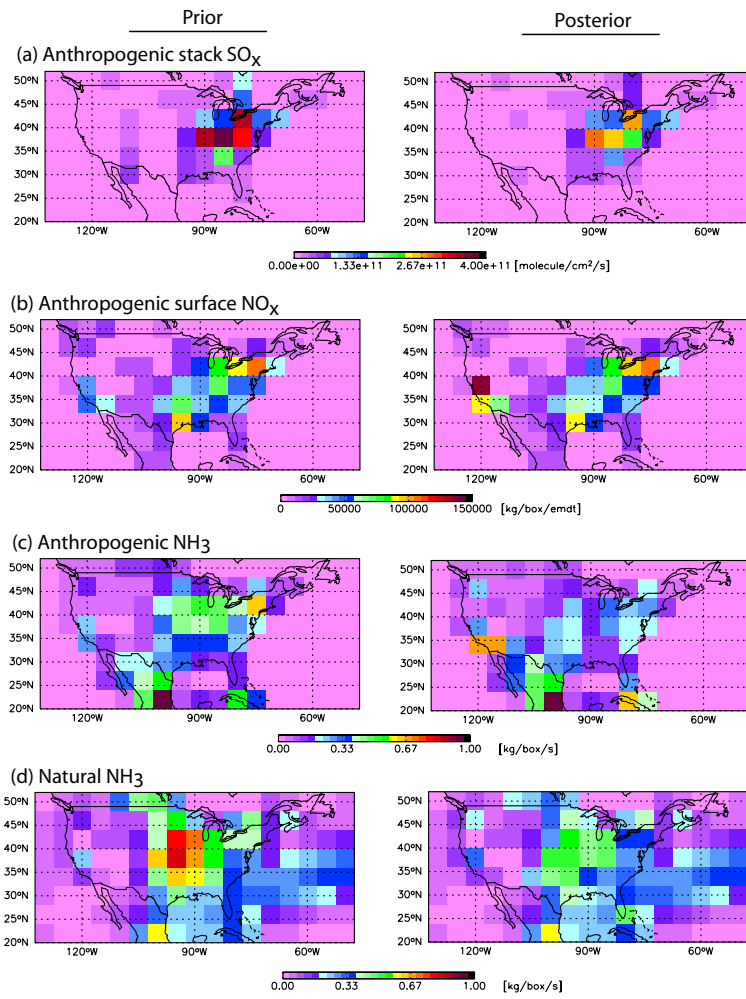
Figure 4.7: Prior and posterior emissions inventories.
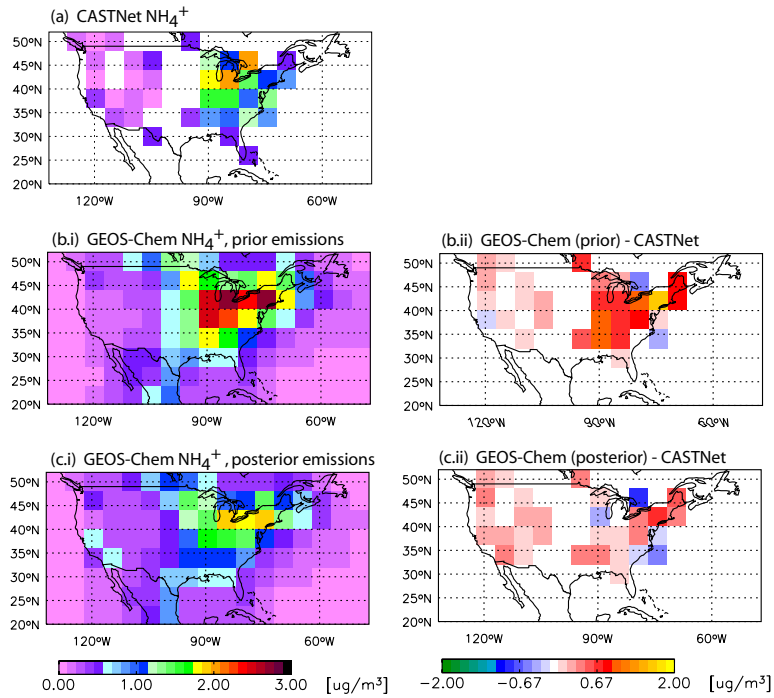
Figure 4.8: Predicted (GEOS-Chem) vs observed (CASTNet) average NH_4^+ concentrations for January, 2002. Panels (b.i) and (b.ii) use the prior emissions inventories, while panels (c.i) and (c.ii) use the optimized posterior emissions inventories.
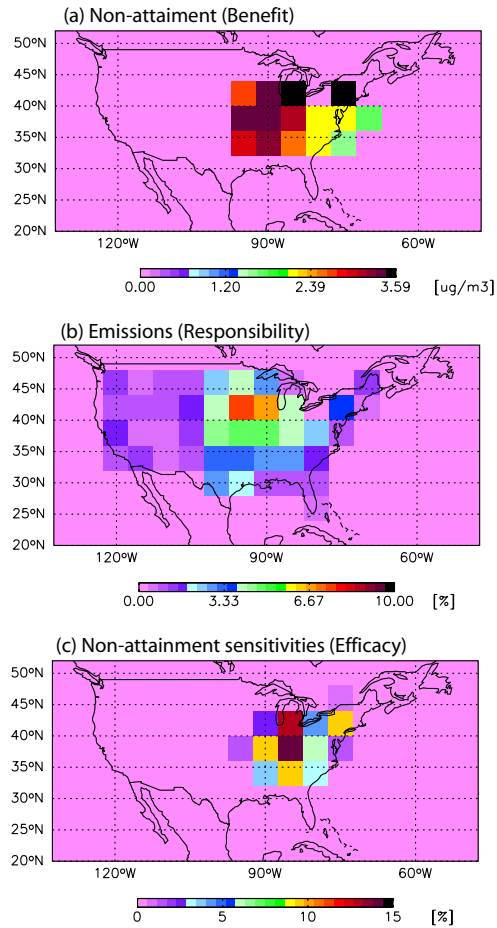
Figure 4.9: Non-attainment sensitivities: regional variability. Panel (a) shows the average contribution (μg m^{-3}) of NH$_4^+$ to non-attainment (J_a) during the month of July 2001. Panel (b) shows the normalized emissions (%) of anthropogenic NH$_3$ during this month, the non-attainment sensitivities (%) of which are shown in panel (c).
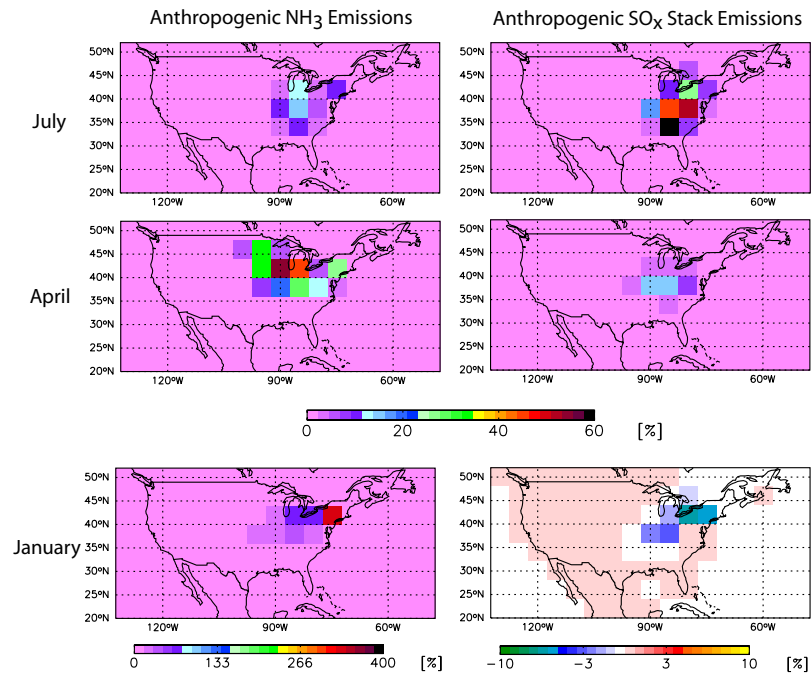
Figure 4.10: Non-attainment sensitivities: seasonal variability. Shown are the normalized non-attainment sensitivities (%) with respect to anthropogenic emissions of NH_3 (left column) and anthropogenic stack emissions of SO_x (right). The efficacy of reducing SO_x emissions dominates in July, while NH_3 controls are more efficient during April and January. Evaluation for January uses the posterior emissions estimate from Section 4.4.3.
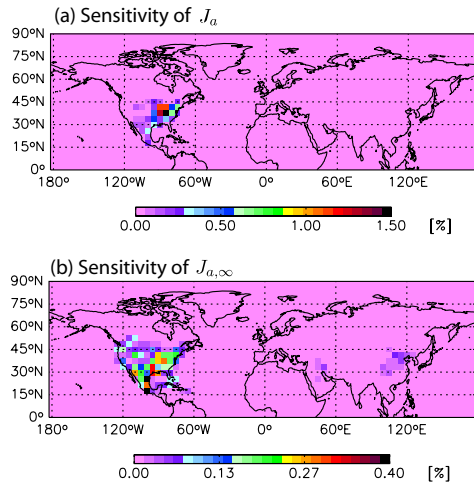
Figure 4.11: Non-attainment sensitivities: long-range influences. Shown are the normalized non-attainment sensitivities (%) with respect to anthropogenic surface emissions of SO_x given two different forms of the non-attainment cost function J_a.

Chapter 5

# Global Secondary Organic Aerosol Formation from Isoprene Oxidation[1]

5.1 Introduction

Secondary organic aerosol (SOA), formed when oxidized products of volatile hydrocarbons condense, often comprises a substantial portion of the organic mass fraction of atmospheric aerosols. The prevalence of organic carbon aerosol on a global scale makes identifying significant sources of SOA an important task, as carbonaceous aerosol is known to strongly influence air quality and climate change. Model predictions of organic carbon aerosol concentrations have exhibited a low bias not present in coincident predictions of black (elemental) carbon, with this bias being attributed to under-prediction of SOA (*Heald et al.*, 2005; *Tsigaridis and Kanakidou*, 2003).

Isoprene (C_5H_8) is the second most abundant hydrocarbon emitted into the Earth's atmosphere after methane ($\sim$500 Tg yr^{-1} (*Guenther et al.*, 1995)). Although it has long been assumed that all its products remain in the gas phase, if isoprene were to yield even a small amount of aerosol, this would have a profound effect on global sources of organic

aerosol. Biogenic volatile organic compounds other than isoprene, such as terpenes and sesquiterpenes, are presently believed to be the largest source of SOA mass on a global scale, with model estimates of the magnitudes of these sources ranging from 12 - 70 Tg yr^{-1} (*Kanakidou et al.*, 2005). Recent laboratory chamber studies of isoprene photooxidation show that SOA yields are 1-2% at high NO$_x$ levels (*Kroll et al.*, 2005) and $\sim$3% at low NO$_x$ levels (*Kroll et al.*, 2006). Furthermore, organic aerosol collected in forested areas is strongly indicative of an isoprene precursor (*Claeys et al.*, 2004a,b; *Ion et al.*, 2005; *Kourtchev et al.*, 2005; *Matsunaga et al.*, 2003). The impact of such a potentially large source of carbonaceous aerosol necessitates careful investigation of the fate of isoprene oxidation products on a global scale.

5.2   SOA modeling

Claeys et al. (2004b) estimated SOA production from isoprene to be 2 Tg yr^{-1} by simply multiplying an estimate of global isoprene emissions by an observed yield of condensed polyols from isoprene; subsequent recognition of additional SOA production pathways increases this estimate (*Claeys et al.*, 2004a). Cloud processing of isoprene oxidation products alone has been calculated to contribute 1.6 Tg yr^{-1} of SOA (*Lim et al.*, 2005). *Matsunaga et al.* (2005) estimated a source of SOA from isoprene in the range of 10-120 Tg yr^{-1}; however, this study neglects the effects of temperature and background organic particulate matter concentrations on gas - particle partitioning, factors known to strongly influence SOA formation.

Recent availability of data from laboratory chamber studies of isoprene oxidation (*Kroll et al.*, 2005, 2006) allow us to now assess the global SOA forming potential of isoprene in a more fundamental manner. Several factors influence SOA formation, such as the ambient NO$_x$ concentration, RO$_2$ concentration, temperature, and heterogeneous reactions (*Limbeck et al.*, 2003; *Czoschke et al.*, 2003; *Edney et al.*, 2005; *Kroll et al.*, 2006). Until the mechanisms that govern these types of behavior are precisely known one must use

empirical parameterizations based on actual laboratory data (*Kanakidou et al.*, 2005).

For inclusion of SOA in global models, the framework of the two-product model (*Odum et al.*, 1996; *Seinfeld and Pankow*, 2003) provides a method for predicting the formation of SOA based upon empirical parameters determined from laboratory chamber studies even when the exact chemical nature of the aerosol products, or even the intermediate gas-phase oxidation products, are not known (*Griffin et al.*, 1999b). The model describes the oxidation of a parent hydrocarbon to produce two representative gas-phase products with stoichiometric coefficients α_1 and α_2. Subsequent partitioning of these products into the aerosol phase is governed by the availability of pre-existing organic aerosol and by their equilibrium partitioning coefficients, K_1, K_2, taking into account the temperature dependence of the partitioning coefficients using the Clausius-Clapyeron equation. At the moment condensation onto other (non-organic) aerosol species is not considered, though this would afford increased SOA formation from all species (*Tsigaridis and Kanakidou*, 2003).

We simulate global SOA formation using the chemical transport model GEOS-Chem (version 7.2.4 with a horizontal resolution of $4° \times 5°$ and 30 layers up to 0.01 hPa, GEOS-3 meteorological fields (*Park et al.*, 2004)), previously implemented with a gas-particle partitioning model of SOA formation from terpenes (*Chung and Seinfeld*, 2002; *Heald et al.*, 2005), updated here to include formation of SOA from oxidation of isoprene using parameters shown in Table 5.1. The α's and K's were derived from the final amount of SOA formed in chamber studies of isoprene oxidation by OH (*Kroll et al.*, 2006) using the same method as *Griffin et al.* (1999a). We assume reaction with OH is the only pathway for formation of SOA from isoprene. Though reaction with O_3 or NO_3 may also lead to SOA formation, the magnitudes of these sources are assumed to be minor, because an order of magnitude more isoprene reacts with OH than with O_3 or NO_3 on a global scale (*Calvert et al.*, 2000). We assign a molecular weight of 130 for the oxidation products from isoprene, which is that of tetrol, a compound prevalent in SOA that originates from isoprene (*Claeys*

et al., 2004b).

An issue with empirical partitioning models is that the conditions of the chamber studies from which the yield parameters are derived may not be representative of atmospheric conditions. The main concern has been that NO_x levels in these experiments tend to be larger than those in the troposphere. The experiments used to derive the yield parameters for isoprene given here were carried out under low NO_x concentrations (<1 ppb) and at cooler temperatures more relevant to tropospheric conditions (*Kroll et al.*, 2006). Still, a single set of yield parameters may not fully represent SOA formation throughout the entire range of conditions present in the atmosphere — further laboratory and modeling studies are required to explicitly specify the dependence the SOA yield parameters on the chemical environment.

Implementation of this model on global scales requires knowledge of thermophysical parameters that are not easily determined experimentally. The enthalpy of vaporization of SOA, $\triangle H_v$, is critical for extrapolating the equilibrium gas-particle partition coefficients to colder temperatures (*Tsigaridis and Kanakidou*, 2003). The value of $\triangle H_v$ depends upon the nature of the SOA and how it was formed (*Offenberg et al.*, 2006), though there is not yet enough experimental data available to justify the use of more than a single value of $\triangle H_v$ for all SOA. The base case value of $\triangle H_v$ used here, 42 kJ mol^{-1} (*Chung and Seinfeld*, 2002), originally considered a lower estimate in comparison to values from similar studies which ranged as high as 156 kJ mol^{-1}, is perhaps in fact quite reasonable, as recent experimental studies of the temperature dependence of SOA formed from α-pinene have placed $\triangle H_v$ closer to the lower estimates (*Offenberg et al.*, 2006; *Stanier et al.*, 2006). The sensitivity of SOA predictions to the aqueous solubility of the oxidation product species, governed by an estimated average Henry's law constant of the oxidation products, H, has also been mentioned by *Tsigaridis and Kanakidou* (2003), though the consequences of variations in H on global SOA predictions have not yet been explored. Loss of these products by wet removal depends strongly on H. Given that polyols resulting from isoprene oxidation are

more soluble than many of the previously identified species in SOA, which were taken to have an average Henry's law constant of 10^5 M atm^{-1} (*Sander*, 1999), we consider the effect of increasing the Henry's law constant of the oxidation products to 10^6 M atm^{-1}, and for comparison, decreasing it to 10^4 M atm^{-1}.

5.3   Results and conclusions

Model predictions of global yearly average SOA concentrations for the year March 2001 - February 2002 are shown in Figure 5.1. We select this time period because it encompasses the ACE-Asia campaign, for which the observed amount of organic carbon aerosol in the free troposphere exceeds predictions by the base case model by a factor of 10-100 (*Heald et al.*, 2005). Panels (a) and (c) show the total concentrations of SOA generated by the existing (base case) biogenic VOCs (terpenes and OVOCs) at the surface and at 5.2 km, respectively. Panels (b) and (d) show total SOA concentrations when isoprene is included as an additional source of SOA. The difference between these two simulations is striking, most notably in the magnitude of the increases in the free troposphere, where typically more than 70% of the SOA is from isoprene. SOA concentrations increase by a factor of 1.5 to 3 in regions of relatively high SOA concentrations, and they increase by more than a factor of 10 in remote marine regions where SOA concentrations are small (<0.01 μg m^{-3}), such as the Indian and South Central Pacific oceans.

The yearly average total SOA burden (B_T) and the net yearly SOA production (P_T) are given in Table 5.2, where the total production is also broken down into contributions from isoprene (P_I) and from the original set of VOCs (P_O). The amount of SOA produced directly from isoprene is 6.2 Tg yr^{-1}, almost as large as the original SOA source in the base calculation, 8.7 Tg yr^{-1}. The presence of this much additional organic substrate enhances SOA formation from other sources by 17%. The total SOA burden more than doubles, and the lifetime of the SOA from isoprene (13.5 d) is twice that of the base case SOA (6.7 d). Results from a one month simulation with model resolution of $2° \times 2.5°$ were equivalent.

Two factors give rise to the distinct distributions and lifetimes of the SOA formed from isoprene compared to the base case set of VOCs. Emissions of isoprene are generally much greater. As a result, isoprene is not completely oxidized near its sources, and substantial amounts of isoprene can be lofted to much greater altitudes. Also, gas-particle partitioning of the isoprene oxidation products is shifted less toward the particle phase than that of the products of the base case VOCs; hence, the lifetime of the isoprene oxidation products is also greater. The combined effect of these factors increases SOA precursor concentrations in the free troposphere where partitioning to the aerosol phase is enhanced owing to lower temperatures, leading to formation of SOA in regions where there was little in the base case. Although this increase alone is not enough to account for the discrepancy between predicted and observed tropospheric organic carbon aerosol in the region studied during the ACE-Asia campaign (*Heald et al.*, 2005), it does significantly impact our global picture of organic carbon aerosol distributions.

The total amount of isoprene predicted to be oxidized by OH is 209 Tg yr^{-1}; the global isoprene SOA "yield" is 2.9%, which is essentially the same as those from the low-NO_x chamber experiments ($\sim$3%). We find that simply calculating the formation of SOA from isoprene from a direct calculation (wherein SOA is formed, irreversibly, as a constant percentage of the amount of isoprene that reacts) leads to lower SOA burdens than the two-product model, in contrast to previous studies comparing these methods (*Lack et al.*, 2004) and *Tsigaridis and Kanakidou* (2003). The reason for this discrepancy is, as noted earlier, a signifcant portion of the SOA from isoprene is formed from the semivolatile oxidation products that only condense substantially at lower temperatures, an effect that may not be as critical for modeling SOA from sources with greater yields.

We examine SOA levels predicted by the base case model (without isoprene as a source of SOA) when using a reasonably larger value of $\triangle H_v = 50$ kJ mol^{-1} or when $H = 10^4$ M atm^{-1}. Use of this value of $\triangle H_v$ leads to a modest increase in the global SOA burden of 0.08 Tg, and average SOA concentrations in the troposphere increase by a factor of

2 to 3. Decreasing H increases the burden by almost 40%. While these are substantial consequences, the overall magnitude of these effects is still small compared to increases of SOA concentrations from isoprene, as shown in Figure 5.2. When isoprene is included as a source of SOA, increasing H to 10^6 M atm^{-1} has little overall effect, as the oxidation products are effectively completely soluble beyond $H \geq 10^5$ M atm^{-1}.

Including isoprene as a source of SOA causes substantial increases in predicted SOA concentrations, particularly in the free troposphere and remote marine environments. A detailed comparison with measured organic carbon aerosol is now in order. This source of SOA may help explain observations of organic carbon aerosol, noted previously to be under-predicted by this (and others) model in these regions (*Heald et al.*, 2005; *Tsigaridis and Kanakidou*, 2003), particularly considering recent revisions in estimates of isoprene emissions (*Guenther et al.*, 2006). This study highlights the need for further research into the chemical fate of the oxidation products of isoprene (*Kroll et al.*, 2006) and the importance of developing SOA models that can explicitly represent condensation of oxidation products normally considered too volatile to contribute to organic aerosol formation (*Donahue et al.*, 2006). These results may have implications for climate change given the magnitude of the predicted top of the atmosphere radiative forcing of organic carbon in year 2100 climate (*Liao et al.*, 2006).

Bibliography

Calvert, J., R. Atkinson, J. Kerr, S. Madronich, G. K. Moortgat, T. Wallington, and G. Yarwood (2000), *The Mechanisms of Atmospheric Oxidation of the Alkenes*, Oxford University Press, New York.

Chung, S. H., and J. H. Seinfeld (2002), Global distribution and climate forcing of carbonaceous aerosols, *J. Geophys. Res.*, *107*(D19), 4407, doi:10.1029/2001JD001397.

Claeys, M., W. Wang, A. C. Ion, I. Kourtchev, A. Gelencser, and W. Maenhaut (2004a), Formation of secondary organic aerosols from isoprene and its gas-phase oxidation products through reaction with hydrogen peroxide, *Atmos. Environ.*, *38*(25), 4093–4098.

Claeys, M., et al. (2004b), Formation of secondary organic aerosols through photooxidation of isoprene, *Science*, *303*(5661), 1173–1176.

Czoschke, N. M., M. Jang, and R. M. Kamens (2003), Effect of acidic seed on biogenic secondary organic aerosol growth, *Atmos. Environ.*, *37*(30), 4287–4299.

Donahue, N. M., A. L. Robinson, C. O. Stanier, and S. N. Pandis (2006), The coupled dilution, partitioning, and chemical aging of semivolatile organics, *Environ. Sci. Technol.*, *in press*.

Edney, E. O., T. E. Kleindienst, M. Jaoui, M. Lewandowski, J. H. Offenberg, W. Wang, and M. Claeys (2005), Formation of 2-methyl tetrols and 2-methylglyceric acid in secondary organic aerosol from laboratory irradiated isoprene/NO_x/SO_2/air mixtures and

their detection in ambient $PM_{2.5}$ samples collected in the eastern United States, *Atmos. Environ.*, *39*(29), 5281–5289.

Griffin, R. J., D. R. Cocker, R. C. Flagan, and J. H. Seinfeld (1999a), Organic aerosol formation from the oxidation of biogenic hydrocarbons, *J. Geophys. Res.*, *104*(D3), 3555–3567.

Griffin, R. J., D. R. Cocker, J. H. Seinfeld, and D. Dabdub (1999b), Estimate of global atmospheric organic aerosol from oxidation of biogenic hydrocarbons, *Geophys. Res. Lett.*, *26*(17), 2721–2724.

Guenther, A., T. Karl, P. Harley, C. Wiedinmyer, P. I. Palmer, and C. Geron (2006), Estimates of global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and Aerosols from Nature), *Atmos. Chem. Phys. Discuss.*, *6*, 107–173.

Guenther, A., et al. (1995), A global-model of natural volatile organic-compound emissions, *J. Geophys. Res.*, *100*(D5), 8873–8892.

Heald, C. L., D. J. Jacob, R. J. Park, L. M. Russell, B. J. Huebert, J. H. Seinfeld, H. Liao, and R. J. Weber (2005), A large organic aerosol source in the free troposphere missing from current models, *Geophys. Res. Lett.*, *32*, L18809, doi:10.1029/2005GL023831.

Ion, A. C., R. Vermeylen, I. Kourtchev, J. Cafmeyer, X. Chi, A. Gelencser, W. Maenhaut, and M. Claeys (2005), Polar organic compounds in rural PM2.5 aerosols from k-puszta, Hungary, during a 2003 summer field campaign: Sources and diurnal variations, *Atmos. Chem. Phys.*, *5*, 1805–1814.

Kanakidou, M., et al. (2005), Organic aerosol and global climate modelling: a review, *Atmos. Chem. Phys.*, *5*, 1053–1123.

Kourtchev, I., T. Ruuskanen, W. Maenhaut, M. Kulmala, and M. Claeys (2005), Observa-

tion of 2-methyltetrols and related photo-oxidation products of isoprene in boreal forest aerosols from Hyytiala, Finland, *Atmos. Chem. Phys.*, *5*, 2761–2770.

Kroll, J. H., N. L. Ng, S. M. Murphy, R. C. Flagan, and J. H. Seinfeld (2005), Secondary organic aerosol formation from isoprene photooxidation under high-NO_x conditions, *Geophys. Res. Lett.*, *32*, L18808, doi:10.1029/2005GL023637.

Kroll, J. H., N. L. Ng, S. M. Murphy, R. C. Flagan, and J. H. Seinfeld (2006), Secondary organic aerosol formation from isoprene photooxidation, *Environ. Sci. Technol.*, *40*(6), 1869–1877, doi:10.1021/es0524301.

Lack, D. A., X. X. Tie, N. D. Bofinger, A. N. Wiegand, and S. Madronich (2004), Seasonal variability of secondary organic aerosol: A global modeling study, *J. Geophys. Res.*, *109*, D03203, doi:10.1029/2003JD003418.

Liao, H., W.-T. Chen, and J. H. Seinfeld (2006), Role of climate change in global predictions of future tropospheric ozone and aerosols, *J. Geophys. Res.*, *in press*, doi: 10.1029/2005JD006852.

Lim, H. J., A. G. Carlton, and B. J. Turpin (2005), Isoprene forms secondary organic aerosol through cloud processing: Model simulations, *Environ. Sci. Technol.*, *39*(12), 4441–4446.

Limbeck, A., M. Kulmala, and H. Puxbaum (2003), Secondary organic aerosol formation in the atmosphere via heterogeneous reaction of gaseous isoprene on acidic particles, *Geophys. Res. Lett.*, *30*(19), doi:10.1029/2003GL017738.

Matsunaga, S., M. Mochida, and K. Kawamura (2003), Growth of organic aerosols by biogenic semi-volatile carbonyls in the forestal atmosphere, *Atmos. Environ.*, *37*(15), 2045–2050.

Matsunaga, S., C. Wiedinmyer, A. B. Guenther, J. J. Orlando, T. Karl, D. W. Toohey, J. P. Greenberg, and Y. Kajii (2005), Isoprene oxidation products are a significant atmospheric aerosol component, *Atmos. Chem. Phys. Discuss.*, *5*, 11,143–11,156.

Odum, J. R., T. Hoffmann, F. Bowman, D. Collins, R. C. Flagan, and J. H. Seinfeld (1996), Gas/particle partitioning and secondary organic aerosol yields, *Environ. Sci. Technol.*, *30*(8), 2580–2585.

Offenberg, J. H., T. E. Kleindienst, M. Jaoui, M. Lewandowski, and E. O. Edney (2006), Thermal properties of secondary organic aerosols, *Geophys. Res. Lett.*, *33*, L03816, doi: 10.1029/2005GL024623.

Park, R. J., D. Jacob, B. D. Field, R. Yantosca, and M. Chin (2004), natural and transboundary pollution influences on sulfate-nitrate-ammonium aerosols in the United States: implications for policy, *J. Geophys. Res.*, *109*, D15204, doi: 10.1029/2003JD004473.

Sander, R. (1999), Compilation of Henry's law constants for inorganic and organic species of potential importance in environmental chemistry (version 3), *http://www.mpch-mainz.mpg.de/~sander/res/henry.html*.

Seinfeld, J. H., and J. F. Pankow (2003), Organic atmospheric particulate material, *Annu. Rev. Phys. Chem.*, *54*(1), 121–140.

Stanier, C. O., N. Donahue, and S. N. Pandis (2006), Atmospheric secondary organic aerosol yields: Model parameter estimation from smog chamber results, *submitted*.

Tsigaridis, K., and M. Kanakidou (2003), Global modelling of secondary organic aerosol in the troposphere: a sensitivity analysis, *Atmos. Chem. Phys.*, *3*, 1849–1869.
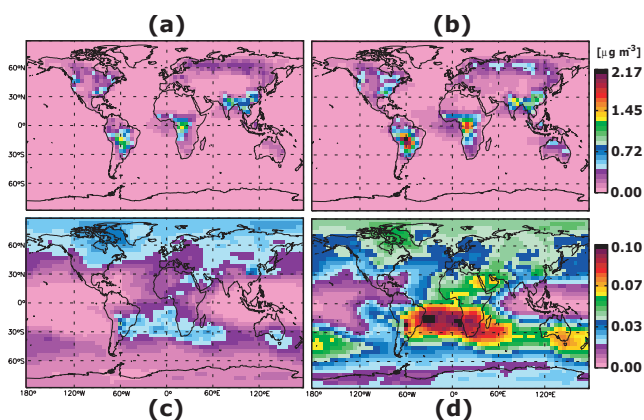
Figure 5.1: Yearly average total SOA concentrations during the year March 2001 - Feb 2002 (a) at the surface without isoprene source (b) at the surface with isoprene source (c) at 5.2 km without isoprene source (d) at 5.2 km with isoprene source.
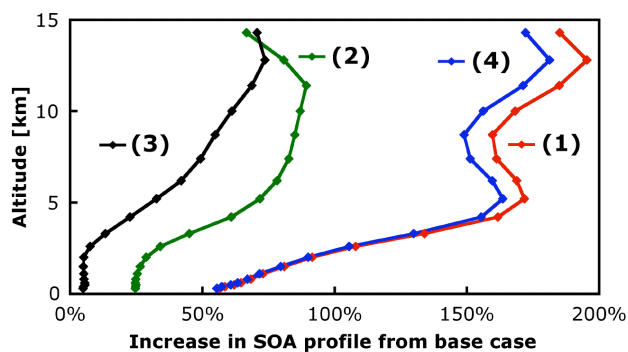
Figure 5.2: Relative increases in global average SOA concentration profiles with respect to the base case model. Line (1) shows the increase in SOA resulting from including isoprene as a source. The two lines at the left show increases in SOA formed from non-isoprene sources due to (2) increasing $\triangle H_v$ to 50 kJ mol^{-1} (3) decreasing H to 10^4 M atm^{-1}. Line (4) shows increases in SOA from including isoprene as a source but increasing H to 10^6 M atm^{-1}.

Table 5.1: Stoichiometric coefficients, α_i, and equilibrium partitioning coefficients, K_i, for SOA formation from low NO_x chamber experiments of reaction of isoprene with OH (*Kroll et al.*, 2006).

Product	α_i	$K_i[\text{m}^3\mu\text{g}^{-1}]$ [b]
1	0.232	0.00862
2	0.0288	1.62

[a]Reference temperature is 295 K

Table 5.2: Summary of yearly SOA production rates and average burdens as a function of SOA sources, SOA enthalpy of vaporization ($\triangle H_v$), and the Henry's law constant of the oxidation products (H).

Source[c]	$\triangle H_v$ [kJ/mol]	H [M/atm]	P_O [Tg/yr]	P_I [Tg/yr]	P_T [Tg/yr]	B_T [Tg]
O	42	10^5	8.7	-	8.7	0.16
$O+I$	42	10^5	10.2	6.2	16.4	0.39
		Sensitivity Calculations				
O	50	10^5	11.1	-	11.1	0.24
O	42	10^4	9.7	-	9.7	0.22
$O+I$	42	10^6	10.1	6.1	16.2	0.38

[b]I = isoprene, O = other biogenic VOC's

Chapter 6

Modeling High vs Low-Yield Pathways of Secondary Organic Aerosol Formation from Aromatic Hydrocarbons

6.1 Introduction

Organic aerosols play an important role in global climate (*Kanakidou et al.*, 2005). A significant fraction of organic aerosol material results from the gas-phase oxidation of volatile hydrocarbons to yield semivolatile products that condense into the particulate phase; this is referred to as secondary organic aerosol (SOA). At present, measured levels of organic aerosol tend to exceed those predicted by global chemical transport models (*Heald et al.*, 2005, 2006). While it is predicted that, on the global scale, secondary organic aerosol from biogenic sources substantially exceeds that from anthropogenic sources (*Tsigaridis et al.*, 2006), data from recent field studies suggest that SOA from anthropogenic hydrocarbons might be more significant than previously thought (*de Gouw et al.*, 2005; *Volkamer et al.*, 2006).

Among anthropogenic hydrocarbons, aromatic compounds are generally considered to be the most important SOA precursors. It has recently been established that the SOA

yields (aerosol yield is defined as the ratio of the mass of organic aerosol produced to the mass of parent hydrocarbon reacted) from aromatics, as well as those from a variety of other hydrocarbons, depend critically on the prevailing NO_x level (*Hurley et al.*, 2001; *Martin-Reviejo and Wirtz*, 2005; *Song et al.*, 2005; *Johnson et al.*, 2004, 2005; *Ng et al.*, 2007). This discovery led to a re-evaluation of aromatic SOA yields (*Ng et al.*, 2007) from the historical yields of *Odum et al.* (1996, 1997) that were measured under high-NO_x conditions. In particular, aromatic yields under the low-NO_x conditions typical of most of the global atmosphere significantly exceed those under high-NO_x conditions typical of urban cores (and of past laboratory chamber experiments). However, given that sources of aromatics are likely co-located with sources of NO_x, the significance of this finding on the global SOA burden is not readily apparent.

The recent data on SOA yields from aromatic hydrocarbons prompt a re-evaluation of the contribution of aromatic SOA to the global SOA production rate and burden. In the current study, the GEOS-Chem global chemical transport model is updated to include abbreviated aromatic oxidation chemistry and SOA formation from aromatics.

6.2   Summary of SOA yields from aromatic hydrocarbons

The SOA-forming potentials of *m*-xylene, toluene, and benzene have been measured in a series of laboratory chamber experiments (*Ng et al.*, 2007). Atmospheric reaction of these aromatics with the hydroxyl radical (OH) initiates a complex series of gas-phase reactions, the mechanisms of which are not fully understood (*Calvert et al.*, 2002). As noted above, the crucial factor governing the nature of the gas-phase chemistry and subsequent aerosol formation is the NO_x level. Experiments were conducted under both low- and high-NO_x conditions; these correspond to NO_x levels of about 1 parts-per-billion by volume (ppb) and several hundred ppb, respectively.

SOA yields of the three aromatics studied are highly dependent on the prevailing NO_x level. Under high-NO_x conditions, measured yields are in general agreement with those

reported by *Odum et al.* (1996, 1997) for toluene and *m*-xylene, in the range of 5 to 10%, with a strong dependence on the amount of organic aerosol present. The first studies of SOA yields from benzene reported yields in the range of 10% for both high and low-NO_x conditions (*Martin-Reviejo and Wirtz*, 2005); *Ng et al.* (2007) found considerably higher yields for benzene, nearly 30% under high-NO_x conditions, party owing to kinetic enhancement. Under low-NO_x conditions, all three compounds exhibit constant yields, in the range of 30%.

The mechanisms for atmospheric oxidation of both biogenic and anthropogenic hydrocarbons are complex and not fully understood; it does appear, however, that the low- and high-NO_x behavior of SOA formation hinges on the competitive reactions of the peroxy radicals (ARO_2) that result from initial attack of OH followed by O_2 addition (*Johnson et al.*, 2004, 2005; *Presto et al.*, 2005; *Kroll et al.*, 2006; *Ng et al.*, 2007).

$$AROM + OH \xrightarrow{k_{OH}} ARO_2 \tag{R1}$$

The peroxy radicals react either with the hydroperoxyl radical (HO_2) or NO, depending on the relative concentrations of HO_2 and NO. Under low-NO_x conditions, reaction with HO_2 is favored, and the resulting products, including hydroperoxides, are generally less volatile than those that result from the NO reaction path. This competition can be represented as follows:

$$ARO_2 + HO_2 \xrightarrow{k_H} \alpha_H SOG_H \tag{R2}$$

$$ARO_2 + NO \xrightarrow{k_N} \alpha_1 SOG_1 + \alpha_2 SOG_2 \tag{R3}$$
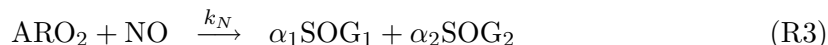
where SOG designates gas-phase semivolatile products, and the α's are mass-based stoichiometric coefficients. The fact that, for all three aromatics, the SOA yield is constant under low-NO_x conditions implies that the semivolatile products are essentially nonvolatile; thus, the RO_2 + HO_2 pathway can be represented as leading to a single nonvolatile prod-

uct, $SOG_H = SOA_H$. Since the high-NO_x pathway exhibits yields that depend on the total amount of absorbing organic aerosol, we use the customary two-product model for SOA formation, originally formulated by *Odum et al.* (1996, 1997); SOG_1 and SOG_2 represent these products, which have associated gas-particle partitioning equilibrium constants, K_1 and K_2. Parameters describing yields under both sets of NO_x levels are given in Table 1. Equilibrium constants and stoichiometric coefficients are based on laboratory studies of *Ng et al.* (2007), where the latter are adjusted to reflect formation of SOG species directly from the peroxy radical rather than the parent hydrocarbon. An implicit assumption, supported by kinetics simulations (*Ng et al.*, 2007), is that reaction (R2) and reaction (R3) are the primary sinks of the peroxy radicals under low and high-NO_x conditions, respectively.

6.3   Aromatic SOA formation in GEOS-Chem

In the current study, the GEOS-Chem global chemical transport model (version 7-04-11 with a horizontal resolution of $2° \times 2.5°$ and 30 layers up to 0.01 hPa, GEOS-4 meteorological fields) is used to simulate one year of present day conditions (2004). This model includes detailed simulation of gas-phase tropospheric chemistry (e.g., *Bey et al.*, 2001; *Hudman*, 2007) in addition to external mixtures of several aerosol components (*Park et al.*, 2004, 2006). Previous versions have been implemented with a gas-particle partitioning model of SOA formation from terpenes, alcohols, sesquiterpenes (*Chung and Seinfeld*, 2002; *Heald et al.*, 2005) and isoprene (*Henze and Seinfeld*, 2006). The addition of abbreviated aromatic oxidation chemistry and SOA formation from aromatics to the chemical reactions and SOA module in GEOS-Chem is described in the following sections.

6.3.1   Aromatic global emissions

Global emissions of benzene, toluene and xylene are taken from the Emission Database for Global Atmospheric Research (EDGAR V2.0) (*Olivier et al.*, 1996, 1999) for 1990 and scaled to the year 1998 using liquid fossil fuel usage from the Global Emission Inventory

Activity (GEIA) project (*Benkovitz et al.*, 1996) following *Bey et al.* (2001). Total emissions are 7.3, 7.3 and 4.7 Tg / year for benzene, toulene and xylene, respectively. The yearly average emission fluxes of each aromatic species, and the combined total, are shown in Fig. 6.1. Primary sources are road transport, solvent use and biomass burning, the latter being a pronounced source of benzene in India.

6.3.2 Implementation of aromatic SOA formation

Gas-phase oxidation of each parent aromatic hydrocarbon (R1) and subsequent reaction of the peroxy radical product with HO_2 (R2) and NO (R3) is calculated online as an additional part of the tropospheric chemical reaction mechanism. By explicitly treating the competition between these two pathways, this implementation allows natural transitions between low and high-yield environments as governed by temperature and concentrations of ARO_2, HO_2 and NO, avoiding static delineation of these regimes based upon VOC / NO_x ratios (*Song et al.*, 2005; *Tsigaridis et al.*, 2006). Parameters for calculating the reaction rate constants for these steps are listed in Table 6.2. Kintetic parameters for reaction of the aromatic species with OH (the rate limiting step in peroxy radical formation) are from *Calvert et al.* (2002). Rate constants for peroxy radical reactions, (R2) and (R3), are from *Atkinson et al.* (1997), assuming similar temperature dependence as peroxy radical reactions with isoprene, as most reactions of hydrocarbons with NO and HO_2 have similar kinetics (*Lightfoot et al.*, 1992; *Eberhard and Howard*, 1997). All forms of xylene are assumed to behave as *m*-xylene for both gas and aerosol processes.

Calculation of SOA formation follows the approach outlined in *Chung and Seinfeld* (2002). SOA concentrations are governed by the gas-particle equilibrium relation,

$$[SOA_{i,j}] = \frac{[SOG_{i,j}]}{K_{i,j}M_0}$$

where M_0 is the total mass of organic aerosol from both primary and secondary organic

aerosol sources.[1] The ratio of organic to carbon mass is assumed to be 2.1. Though this differs from previous studies that used a ratio of 1.4 (*Henze and Seinfeld*, 2006; *Liao et al.*, 2007; *Heald et al.*, 2006; *Zhang et al.*, submitted; *van Donkelaar et al.*, 2007), 2.1 is recommended for non-urban aerosols (*Turpin and Lim*, 2001). The sensitivity of the calculation with respect to this assumption is explored in Section 6.5. The Clausius-Clapeyron equation is used to extrapolate equilibrium constants to tropospheric temperatures. The enthalpy of vaporization for SOA, a major source of uncertainty in such calculations (*Tsigaridis and Kanakidou*, 2003; *Henze and Seinfeld*, 2006), is assumed to be 42 kJ mol^{-1} for all species, in the range of limited available experimental data (*Offenberg et al.*, 2006; *Stanier et al.*, 2006).

6.4 Simulation of global aromatic SOA

The predicted contributions of benzene, toluene, and xylene to global aromatic SOA via high- and low-yield pathways (the opposite of the NO$_x$ levels) and the percentage of the aromatics that react via each pathway are given in Table 6.3. Thus, 49% of globally emitted benzene, for example, reacts via reaction R2, the high-yield path, whereas only 26% of xylene follows this route. Unexpectedly, the less reactive aromatics, as measured by their OH reaction rate constants, actually produce more SOA globally. The explanation is that lower reactivity affords the parent hydrocarbon more time to be transported to regions of lower [NO$_x$]/[HO$_2$] ratios, where ultimately, once reacted, the SOA yield is larger. Proportionately more of a reactive molecule like xylene is consumed in regions in which the NO$_x$ levels are more characteristic of the source emission areas. On a global average, 40% of the total aromatics proceed through the high yield pathway. As a result, 75% of global aromatic SOA is produced via the low-NO$_x$ pathway.

Figure 6.2 shows the predicted annual average, surface-level concentration of aromatic

---

[1]The standard model, as well as some other studies (*Tsigaridis and Kanakidou*, 2003), considers condensation directly on sulfate aerosol, which is not included in these simulations.

SOA by season. The concentrations generally reflect the distribution of areas with substantial anthropogenic emissions. Peak SOA concentrations form from benzene in India during DJF, while concentrations in the eastern parts of the United States and Europe are low owing to elevated $[NO_x]/[HO_2]$ ratios during these months. During the summer, SOA formation in the latter areas is facilitated by lower NO_x concentrations and is not inhibited by increased temperatures as formation of SOA via the low-NO_x pathway is treated as irreversible.

6.5   Anthropogenic vs biogenic SOA

Table 6.4 presents the predicted production rates of SOA from the three aromatics (Tg/yr) and their global burdens (Tg), as compared with the values for total biogenics, about 50% of which is attributable to isoprene. Even though aromatic SOA is appreciable, the global SOA burden continues to be dominated by biogenic sources, as predicted based on the best current SOA yields from laboratory chamber studies.

While the bulk of the total modeled SOA is biogenic in origin, there are regions of the global distribution where concentrations of SOA from aromatics are predicted to be equal or larger. Figure 6.3 shows the natural log of the ratio of anthropogenic (aromatic) to biogenic SOA concentrations,

$$\mathcal{R} = \ln \left(\frac{[\text{SOA}_\text{anth}]}{[\text{SOA}_\text{biogenic}]} \right)$$

excluding locations where the total SOA is less than 1% of the maximum. On a yearly average, the only region predicted to have a substantial anthropogenic fraction of SOA is a small area of outflow from eastern Asia. However, during winter, much of the outflow regions in the northern hemisphere are dominated by anthropogenic SOA. For example, Figure 6.4 shows the vertical distribution of SOA types at 75° W. Nevertheless, concentrations are still an order of magnitude smaller than the total (primary plus secondary)

organic aerosol.

It is interesting to note that, as a result of revision of biogenic emissions inventories, the predicted burden of SOA from biogenic sources is more than double that of previous works (*Henze and Seinfeld*, 2006). Isoprene and terpene emissions are calculated using MEGAN (*Guenther et al.*, 2006), which gives a total yearly emission of 462 Tg. Emissions of other SOA precursors are calculated the same as previously (see *Chung and Seinfeld*, 2002; *Heald et al.*, 2005). Isoprene emissions are about 20% higher than previous global modeling studies of SOA from isoprene with this model (*Henze and Seinfeld*, 2006; *Heald et al.*, 2006; *Liao et al.*, 2007), which had used a modified (*Bey et al.*, 2001) emissions from *Guenther et al.* (1995). Given the limited data used to construct the older inventory, and that the newer inventory shows improved agreement with in-situ data as well as consistency with top-down constraints from satellite measurements of formaldehyde (*Palmer et al.*, 2003; *Shim et al.*, 2005), it is likely the latter inventory represents substantial improvements. Even though emissions inventories for non-isoprene biogenic SOA precursors (terpenes, alcohols, sesquiterpenes) are the same as in previous works, the resulting SOA from these sources is also more than double. The consequence of assuming that the total organic / carbon mass ratio is 2.1, rather than 1.4, causes a 10-20% increase in SOA concentrations from all biogenic sources.

6.6   Conclusions

Simulation of formation of SOA from the aromatic species benzene, toluene and xylene is added to the global chemical transport model GEOS-Chem. A simple mechanism is presented that accounts for the competition between low and high-NO_x pathways on SOA formation in a continuous fashion. Depending upon the immediate chemical environment, secondary peroxy radicals from photooxidation of aromatics by OH react with either NO or HO_2. Formation of SOA from reaction with NO leads to reversible formation of SOA following the two-product model of *Odum et al.* (1996, 1997) using empirically determined

yield and partitioning coefficients from the high-NO_x studies of *Ng et al.* (2007). Aromatic peroxy radicals that react with HO_2 are treated as forming SOA irreversibly, following the low-NO_x results of *Ng et al.* (2007).

The dependence of SOA formation on the NO_x environment for aromatic leads to some surprising predictions, with potential implications for other sources of SOA noted to depend strongly on the NO_x environment (*Song et al.*, 2005; *Kroll et al.*, 2006). Previously assumed to generate negligible amount of SOA owing to its low reactivity with OH, benzene is estimated in this work to be the most important aromatic species with regards to formation of SOA. Its low initial reactivity allows benzene to be transported away from source regions, where $[NO_x]/[HO_2]$ ratios are high, to more remote regions, where this ratio is lower and, hence, the ultimate yield of SOA is higher. In total, while only 40% percent of the aromatic species react via the low-NO_x pathway, nearly 75% of the aromatic SOA is formed via this mechanism. Predicted SOA concentrations from aromatics in the eastern United States and Europe are actually largest during the summer, when NO_x concentrations are lower. Influence of NO_x variability on SOA formation is particularly interesting as current models would appear to underestimate seasonal NO_x cycles in these areas as indicated by observations from GOME (*van Noije et al.*, 2006).

Even though the predicted burden and production rate of aromatic SOA is twice that of previous estimates (*Tsigaridis and Kanakidou*, 2003), the contribution of these sources to global SOA is small relative to biogenic sources. Biogenic sources are estimated to comprise 90% of the global SOA, about half of which comes from isoprene (*Henze and Seinfeld*, 2006). However, owing to differences in spatial distributions of sources and seasons of peak production, there are regions in which aromatic SOA is predicted to contribute substantially, and even dominate, the local SOA concentrations, such as outflow regions from North America and South East Asia during the wintertime. These estimates highlight the importance of additional studies of the NO_x dependence of SOA formation from other species, and the role that such dependance plays on SOA yields of long lived hydrocarbons

in general.

Bibliography

Atkinson, R., D. L. Baulch, R. A. Cox, R. F. Hampson, J. A. Kerr, M. J. Rossi, and J. Troe, Evaluated kinetic and photochemical data for atmospheric chemistry: Supplement VI - IUPAC subcommittee on gas kinetic data evaluation for atmospheric chemistry, *J. Phys. Chem. Ref. Data*, *26*(6), 1329–1499, 1997.

Benkovitz, C. M., M. T. Scholtz, J. Pacyna, L. Tarrason, J. Dignon, E. C. Voldner, P. A. Spiro, J. A. Logan, and T. E. Graedel, Global gridded inventories of anthropogenic emissions of sulfur and nitrogen, *J. Geophys. Res.*, *101*(D22), 29,239–29,253, 1996.

Bey, I., et al., Global modeling of tropospheric chemistry with assimilated meteorology: Model description and evaluation, *J. Geophys. Res.*, *106*(D19), 23,073–23,095, 2001.

Calvert, J., R. Atkinson, K. H. Becker, R. M. Kamens, J. H. Seinfeld, T. J. Wallington, and G. Yarwood, *The mechanisms of atmospheric oxidation of aromatic hyrdocarbons*, Oxford University Press, New York, 2002.

Chung, S. H., and J. H. Seinfeld, Global distribution and climate forcing of carbonaceous aerosols, *J. Geophys. Res.*, *107*(D19), 4407, doi:10.1029/2001JD001397, 2002.

de Gouw, J. A., et al., Budget of organic carbon in a polluted atmosphere: Results from the New England Air Quality Study in 2002, *J. Geophys. Res.*, *110*(D16), D16305, doi: 10.1029/2004JD005623, 2005.

Eberhard, J., and C. J. Howard, Rate coefficients for the reactions of some C-3 to C-5 hydrocarbon peroxy radicals with NO, *J. Phys. Chem. A*, *101*(18), 3360–3366, 1997.

Guenther, A., T. Karl, P. Harley, C. Wiedinmyer, P. I. Palmer, and C. Geron, Estimates of global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and Aerosols from Nature), *Atmos. Chem. Phys.*, *6*, 107–173, 2006.

Guenther, A., et al., A global-model of natural volatile organic-compound emissions, *J. Geophys. Res.*, *100*(D5), 8873–8892, 1995.

Heald, C. L., D. J. Jacob, R. J. Park, L. M. Russell, B. J. Huebert, J. H. Seinfeld, H. Liao, and R. J. Weber, A large organic aerosol source in the free troposphere missing from current models, *Geophys. Res. Lett.*, *32*, L18809, doi:10.1029/2005GL023831, 2005.

Heald, C. L., et al., Concentrations and sources of organic carbon aerosols in the free troposphere over North America, *J. Geophys. Res.*, *111*(D23), 2006.

Henze, D. K., and J. H. Seinfeld, Global secondary organic aerosol from isoprene oxidation, *Geophys. Res. Lett.*, *33*(9), L09812, doi:10.1029/2006GL025976, 2006.

Hudman, R. C., 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.

Hurley, M. D., O. Sokolov, T. J. Wallington, H. Takekawa, M. Karasawa, B. Klotz, I. Barnes, and K. H. Becker, Organic aerosol formation during the atmospheric degradation of toluene, *Environ. Sci. Technol.*, *35*(7), 1358–1366, 2001.

Johnson, D., P. Cassanelli, and R. A. Cox, Isomerization of simple alkoxyl radicals: New temperature-dependent rate data and structure activity relationship, *J. Phys. Chem. A*, *108*(4), 519–523, 2004.

Johnson, D., M. E. Jenkin, K. Wirtz, and M. Martin-Reviejo, Simulating the formation of secondary organic aerosol from the photooxidation of aromatic hydrocarbons, *Environmental Chemistry*, *2*(1), 35–48, 2005.

Kanakidou, M., et al., Organic aerosol and global climate modelling: a review, *Atmos. Chem. Phys.*, *5*, 1053–1123, 2005.

Kroll, J. H., N. L. Ng, S. M. Murphy, R. C. Flagan, and J. H. Seinfeld, Secondary organic aerosol formation from isoprene photooxidation, *Environ. Sci. Technol.*, *40*(6), 1869–1877, doi:10.1021/es0524301, 2006.

Liao, H., D. K. Henze, J. H. Seinfeld, S. Wu, and L. J. Mickley, Biogenic secondary organic aerosol over the United States: Comparison of climatological simulations with observations, *J. Geophys. Res.*, *112*, D06201, doi:10.1029/2006JD007813, 2007.

Lightfoot, P. D., R. A. Cox, J. N. Crowley, M. Destriau, G. D. Hayman, M. E. Jenkin, G. K. Moortgat, and F. Zabel, Organic peroxy-radicals - kinetics, spectroscopy and tropospheric chemistry, *Atmos. Environ.*, *26*(10), 1805–1961, 1992.

Martin-Reviejo, M., and K. Wirtz, Is benzene a precursor for secondary organic aerosol?, *Environ. Sci. Technol.*, *39*(4), 1045–1054, doi:10.1021/es049802a, 2005.

Ng, N. L., J. H. Kroll, A. W. H. Chan, P. S. Chhabra, R. C. Flagan, and J. H. Seinfeld, Secondary organic aerosol formation from m-xylene, toluene, and benzene, *Atmos. Chem. Phys. Discuss.*, 2007.

Odum, J. R., T. Hoffmann, F. Bowman, D. Collins, R. C. Flagan, and J. H. Seinfeld, Gas/particle partitioning and secondary organic aerosol yields, *Environ. Sci. Technol.*, *30*(8), 2580–2585, 1996.

Odum, J. R., T. P. W. Jungkamp, R. J. Griffin, H. J. L. Forstner, R. C. Flagan, and J. H.

Seinfeld, Aromatics, reformulated gasoline, and atmospheric organic aerosol formation, *Environ. Sci. Technol.*, *31*(7), 1890–1897, 1997.

Offenberg, J. H., T. E. Kleindienst, M. Jaoui, M. Lewandowski, and E. O. Edney, Thermal properties of secondary organic aerosols, *Geophys. Res. Lett.*, *33*, L03816, doi: 10.1029/2005GL024623, 2006.

Olivier, J. G. J., A. F. Bouwman, C. W. M. Van der Maas, J. J. M. Berdowski, C. Veldt, J. P. J. Bloos, A. J. H. Visschedijk, P. Y. J. Zandveld, and J. L. Haverlag, Description of EDGAR Version 2.0: A set of global emission inventories of greenhous gases and ozone-depleting substances for all anthropogenic and most natural sources on a per country basis and on $1° \times 1°$ grid, *Tech. rep.*, 1996.

Olivier, J. G. J., A. F. Bouwman, J. J. M. Berdowski, C. Veldt, J. P. J. Bloos, A. J. H. Visschedijk, C. W. M. Van der Maas, and P. Y. J. Zandveld, Sectoral emission inventories of greenhouse gases for 1990 on a per country basis as well as on 1x1 degree, *Environmental Science & Policy*, *2*, 241–264, 1999.

Palmer, P. I., D. J. Jacob, A. M. Fiore, R. V. Martin, K. Chance, and T. P. Kurosu, Mapping isoprene emissions over North America using formaldehyde column observations from space, *J. Geophys. Res.*, *108*(D6), 2003.

Park, R. J., D. Jacob, B. D. Field, R. Yantosca, and M. Chin, Natural and transboundary pollution influences on sulfate-nitrate-ammonium aerosols in the United States: implications for policy, *J. Geophys. Res.*, *109*, D15204, doi:10.1029/2003JD004473, 2004.

Park, R. J., D. J. Jacob, N. Kumar, and R. M. Yantosca, Regional visibility statistics in the United States: Natural and transboundary pollution influences, and implications for the Regional Haze Rule, *Atmos. Environ.*, *40*(28), 5405–5423, 2006.

Presto, A. A., K. E. H. Hartz, and N. M. Donahue, Secondary organic aerosol production

from terpene ozonolysis. 2. effect of NOx concentration, *Environ. Sci. Technol.*, *39*(18), 7046–7054, 2005.

Shim, C., Y. H. Wang, Y. Choi, P. I. Palmer, D. S. Abbot, and K. Chance, Constraining global isoprene emissions with Global Ozone Monitoring Experiment (GOME) formaldehyde column measurements, *J. Geophys. Res.*, *110*(D24), 2005.

Song, C., K. S. Na, and D. R. Cocker, Impact of the hydrocarbon to NO_x ratio on secondary organic aerosol formation, *Environ. Sci. Technol.*, *39*(9), 3143–3149, doi: 10.1021/es0493244, 2005.

Stanier, C. O., N. Donahue, and S. N. Pandis, Atmospheric secondary organic aerosol yields: Model parameter estimation from smog chamber results, *submitted*, 2006.

Tsigaridis, K., and M. Kanakidou, Global modelling of secondary organic aerosol in the troposphere: a sensitivity analysis, *Atmos. Chem. Phys.*, *3*, 1849–1869, 2003.

Tsigaridis, K., M. Krol, F. J. Dentener, Y. Balkanski, J. Lathiere, S. Metzger, D. A. Hauglustaine, and M. Kanakidou, Change in global aerosol composition since preindustrial times, *Atmos. Chem. Phys.*, *6*, 5143–5162, 2006.

Turpin, B. J., and H. J. Lim, Species contributions to $PM_{2.5}$ mass concentrations: Revisiting common assumptions for estimating organic mass, *Aerosol. Sci. Tech.*, *35*(1), 602–610, 2001.

van Donkelaar, A., R. V. Martin, R. J. Park, C. L. Heald, T.-M. Fu, H. Liao, and A. Guenther, Model evidence for a significant source of secondary organic aerosol from isoprene, *Atmos. Environ.*, 2007.

van Noije, T. P. C., et al., Multi-model ensemble simulations of tropospheric NO_2 compared with GOME retrievals for the year 2000, *Atmos. Chem. Phys.*, *6*, 2943–2979, 2006.

Volkamer, R., J. L. Jimenez, F. San Martini, K. Dzepina, Q. Zhang, D. Salcedo, L. T. Molina, D. R. Worsnop, and M. J. Molina, Secondary organic aerosol formation from anthropogenic air pollution: Rapid and higher than expected, *Geophys. Res. Lett.*, *33*(17), L17811, doi:10.1029/2006GL02689, 2006.

Zhang, Y., J.-P. Huang, D. K. Henze, and J. H. Seinfeld, The role of isoprene in secondary organic aerosol formation on a regional scale, submitted.

Table 6.1: Stoichiometric coefficients, $\alpha_{i,j}$, and equilibrium partitioning coefficients, $K_{i,j}$, for SOA formation derived from high and low-NO_x chamber experiments of reaction of aromatics with OH (*Ng et al.*, 2007). The reference temperature for the $K_{i,j}$'s is 295 K.

Parent aromatic i	$\alpha_{i,H}$	$\alpha_{i,1}$	$\alpha_{i,2}$	$K_{i,1}$ $[m^3\ \mu\ g^{-1}]$	$K_{i,2}$ $[m^3\ \mu\ g^{-1}]$
benzene	0.2272	0.0442	0.5454	3.3150	0.0090
toluene	0.2349	0.0378	0.0737	0.4300	0.0470
m-xylene	0.2052	0.0212	0.0615	0.7610	0.0290

Table 6.2: Reaction rate constants, $k = Ae^{B/T}$

Reaction[a]	k_{298} $[cm^3\ s^{-1}molec^{-1}]$	A $[cm^3\ s^{-1}molec^{-1}]$	B $[K]$
$k_{OH,B}$	1.22×10^{-12}	2.33×10^{-12}	-193
$k_{OH,T}$	3.67×10^{-12}	1.81×10^{-12}	338
$k_{OH,X}$	2.31×10^{-11}	2.31×10^{-11}	0
k_H	1.5×10^{-11}	1.4×10^{-12}	700
k_N	8.5×10^{-12}	2.6×10^{-12}	350

[a]B = benzene, T = toluene, X = xylene. Constants k_H and k_N assumed equal for each parent aromatic.

Table 6.3: Percentages of peroxy radical (ARO_2) from each parent aromatic that react via the high-yield (R2) vs low-yield (R3) pathways. In parenthesis are the eventual contributions from each pathway to the total SOA from all aromatic species.

Parent Aromatic	Pathway (R2) RO_2 fate (% of SOA_{arom})	Pathway (R3) RO_2 fate (% of SOA_{arom})
benzene	49% (36%)	51% (16%)
toluene	40% (29%)	60% (5%)
xylene	26% (10%)	74% (3%)

Table 6.4: Global SOA budgets from aromatics and biogenics

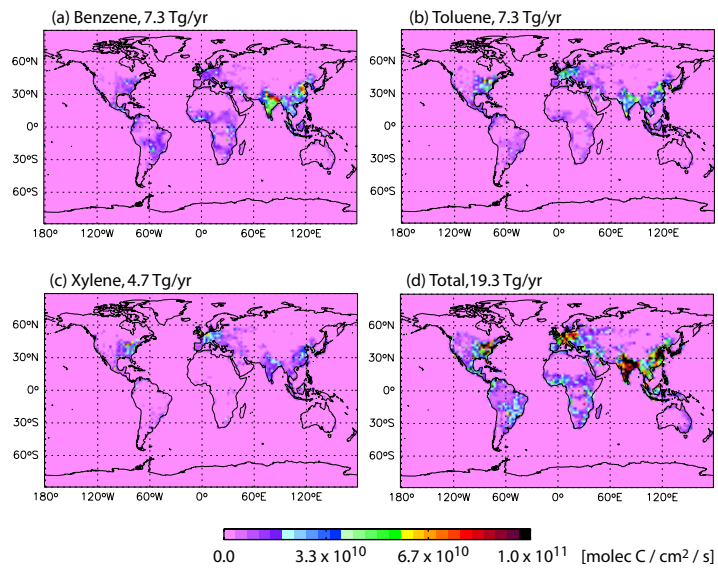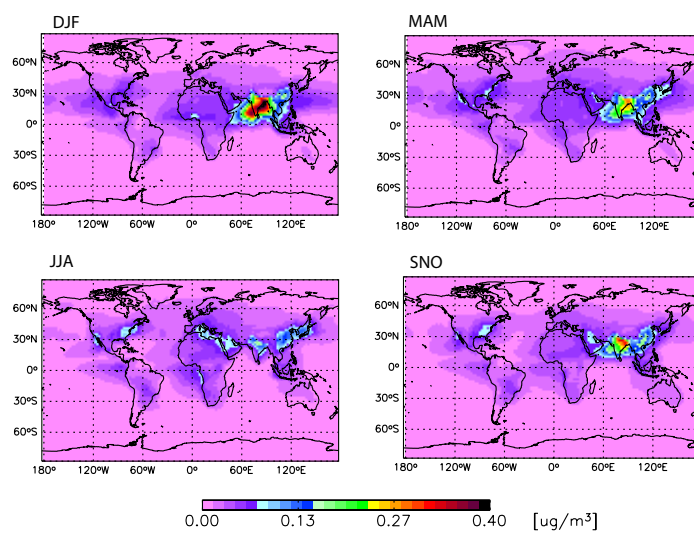| Hydrocarbon	SOA Production (Tg/yr)	Burden (Tg)
aromatics	3.7	0.10
biogenics	30.3	0.81
isoprene	14.1	0.44
other	16.2	0.37
total	34.0	0.91

Figure 6.1: Emissions of aromatic compounds.

Figure 6.2: Seasonal distributions of the total surface level SOA from benzene, toluene and xylene.
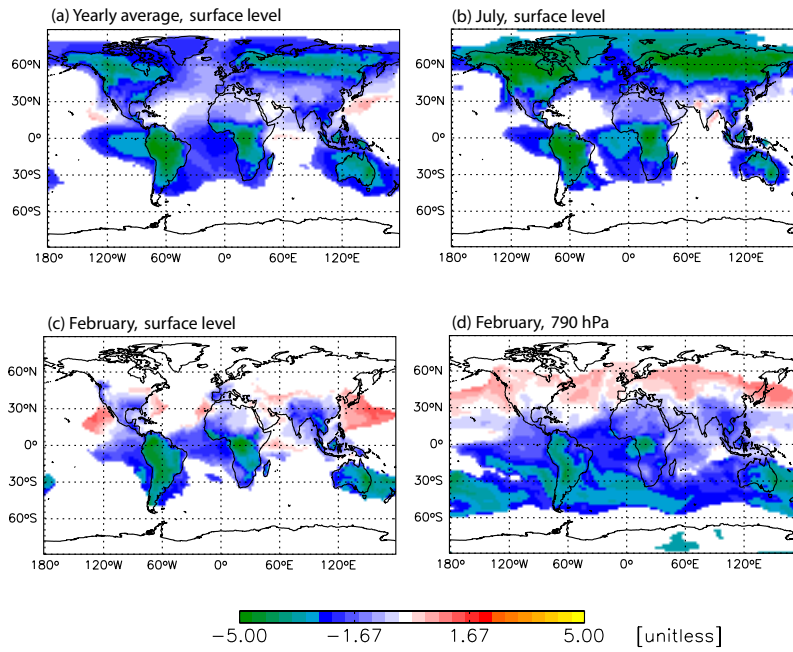
Figure 6.3: Natural log of the ratio of concentrations of anthropogenic to biogenic SOA. White areas are either a transition between anthropogenic and biogenic dominated regions or areas where total concentrations were smaller than 1% of the maximum.
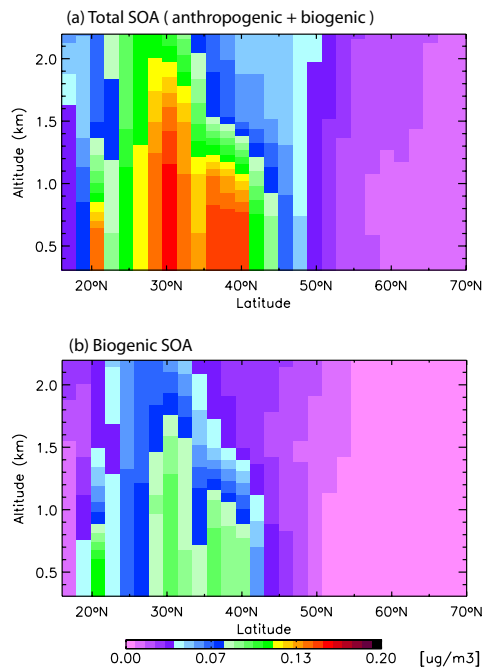
Figure 6.4: SOA export from the eastern United States during February from: (a) anthropogenic and biogenic sources, (b) from biogenic sources only.

Chapter 7

Conclusions

Inclusion of isoprene as a source of secondary organic aerosol (SOA) in a global model has been shown to increase the global burden of SOA from all sources by more than a factor of two. The isoprene source substantially increases SOA concentrations in the free troposphere, because isoprene, and, more importantly, isoprene's oxidation products, have much greater concentrations at higher altitudes than other biogenic SOA precursors, highlighting the importance of semi-volatile organics for SOA formation. Despite considerable uncertainty in model parameters, these results were demonstrated to be robust with respect to increases in partitioning of non-isoprene oxidation products at higher altitudes and increased wet removal of isoprene oxidation products. This additional source of SOA enhances production of SOA from other parent hydrocarbons by 17%, and leads to an overall distribution of SOA that differs enough from previous predictions to warrant reevaluation of the radiative effects of organic carbon aerosol.

Formation of SOA from anthropogenic sources in global models has also been assessed, prompted by recent re-evaluation of the SOA yields from benzene, toluene and m-xylene in low vs high-NO_x environments. A simple mechanism has been presented for utilizing standard empirical yield parameters from these experiments to describe the competition between the low-NO_x (high-yield) and high-NO_x (low yield) pathways in global models of SOA formation. Of all the aromatics considered, an important finding was that benzene,

the least reactive of the aromatics with respect to initial reaction with OH, was predicted to have the greatest overall SOA yield, as its long lifetime afforded it the chance to transport away from anthropogenic source regions to areas where the $[NO_x]/[HO_2]$ ratio was smaller, favoring the high-yield pathways. Consideration of the high-yield, irreversible pathway also lead to higher predicted concentrations of aromatic SOA during the summer in the eastern United States and Europe than during the winter owing to seasonal cycling of NO_x levels. Though not a significant source on global scales, SOA formation from aromatic species was estimated to comprise a substantial portion of SOA concentrations in select regions and seasons. Overall, enhancements of SOA formation owing to isopene and aromatic compounds still generally falls short of observed levels in a variety of environments. The global modeling work presented here have highlighted the importance of further laboratory and modeling studies of the NO_x dependent behavior of SOA yields from additional hydrocarbons.

The feasibility of inverse modeling a multicomponent, size resolved aerosol evolving by condensation / evaporation has been investigated. The adjoint method was applied to the multicomponent aerosol dynamic equation in a box model (zero-dimensional) framework. Both continuous and discrete formulations of the model (the forward equation) and the adjoint were considered. A test example was studied in which the initial aerosol size-composition distribution and the pure component vapor concentrations (i.e., vapor pressures) were estimated based upon measurements of all species, or a subset of the species, and the entire size distribution, or a portion of the size distribution. It was found that the inverse model using the adjoint method could successfully retrieve initial size distributions and pure component vapor concentrations even when only a subset of the species or a portion of the size distribution was observed. The ability to resolve these parameters, of course, depends upon the form of the initial estimates, the nature of the observations and the length of the assimilation period.

We have presented the adjoint of the global chemical transport model GEOS-Chem,

focusing on the chemical and thermodynamic relationships between sulfate – ammonium – nitrate aerosols and their gas-phase precursors. The adjoint model was constructed from a combination of manually and automatically derived discrete adjoint algorithms and numerical solutions to continuous adjoint equations. Explicit inclusion of the processes that govern secondary formation of inorganic aerosol has been shown to afford efficient calculation of model sensitivities such as the dependence of sulfate and nitrate aerosol concentrations on emissions of SO_x, NO_x, and NH_3. The accuracy of the adjoint model has been extensively verified by comparing adjoint to finite difference sensitivities, which were shown to agree within acceptable tolerances. We have explored the robustness of these results, noting how discontinuities in the advection routine hinder, but do not entirely preclude, the use of such comparisons for validation of the adjoint model.

The potential for inverse modeling using the adjoint of GEOS-Chem was assessed first in a data assimilation framework using simulated observations, demonstrating the feasibility of exploiting gas- and aerosol-phase measurements for optimizing emission inventories of aerosol precursors. Next, data from the IMPROVE network of aerosol sulfate and nitrate measurements was used to evaluate sources of SO_x, NO_x and NH_3 during January, 2002. Optimized emissions inventories were found to substantially redistribute emissions of NH_3, with reductions in the Midwest and increases in the Southwest. Improved comparison with observations of aerosol NH_4^+ provided an independent measure of confidence in the posterior emissions estimates. Sensitivities with respect to PM non-attainment metrics were calculated for January, April, and July. The adjoint sensitivities clearly demonstrating how reduction in SO_x would be most effective during the summer, while NH_3 controls would be the most effective in April and January. As of yet, intercontinental influences were found to minimally affect peak aerosol concentrations, but were not as negligible in terms of their perturbation on background levels of aerosol concentrations.

Looking toward the future, recent deployment of remote sensing instruments affords unprecedented opportunity for furthering our understanding of the chemical state of the

troposphere. Instruments aboard several satellites have begun to provide detailed measurements of CO, CH_4 and O_3 throughout the troposphere, in addition to column measurements of species such as SO_2 and NO_2. Further estimates of tropospheric composition are also being provided by remote sensing measurements of aerosol optical properties. What constraints do combinations of such measurements place on the lifetimes and production pathways of important gas and aerosol phase species? How does assimilating data from one satellite affect the agreement with another, and how can we devise additional measurement strategies that make maximal use of such combinations of information? Explicit treatment of gas-phase chemistry, aerosol thermodynamics and secondary aerosol formation in a global transport model renders the combination of forward and inverse model analysis presented in this work a powerful means for starting to address such questions.