Targeted Observations for Atmospheric Chemistry and Transport Models

Adrian Sandu

Department of Computer Science, Virginia Polytechnic Institute and State
University, Blacksburg, VA 24061
`sandu@cs.vt.edu`

Abstract. The aim of this paper is to address computational aspects of the targeted observations problem for atmospheric chemical transport models. The fundamental question being asked is where to place the observations such that, after data assimilation, the uncertainty in the resulting state is minimized. Our approach is based on reducing the system along the subspace defined by the dominant singular vectors, and computing the locations of maximal influence on the verification area. Numerical results presented for a simulation of atmospheric pollution in East Asia in March 2001 show that the optimal location of observations depends on the pattern of the flow but is different for different chemical species. Targeted observations have been previously considered in the context of numerical weather prediction. This work is, to the best of our knowledge, the first effort to study targeted observations in the context of chemical transport modeling. The distinguishing feature of these models is the presence of stiff chemical interactions.

Keywords: Chemical transport models, data assimilation, adjoint models, singular vectors, targeted observations.

1 Introduction

Our ability to anticipate and manage changes in atmospheric pollutant concentrations relies on an accurate representation of the chemical state of the atmosphere. As our fundamental understanding of atmospheric chemistry advances, novel computational tools are needed to integrate observational data and models together to provide the best, physically consistent estimate of the evolving chemical state of the atmosphere. Such an analysis state better defines the spatial and temporal fields of key chemical components in relation to their sources and sinks. This information is critical in designing cost-effective emission control strategies for improved air quality, for the interpretation of observational data such as those obtained during intensive field campaigns, and to the execution of air-quality forecasting.

Data assimilation is the process of integrating observational data and model predictions in order to provide an optimal analysis of the state of the system (here, the atmosphere) and/or optimal estimates of important model parameters (e.g., the strength of the anthropogenic emissions that drive the pollutant

concentrations in the atmosphere). In a variational approach data assimilation is formulated as an optimization problem where the model parameters are estimated such that the mismatch between model predictions and observations is minimized. In atmospheric applications data assimilation is typically used to find the optimal initial state of the system; the resulting parameter estimation problems have millions of control variables.

The objective of this work is to develop techniques for optimal placement of observations in chemical transport modeling. Adaptive observations placed in well-chosen locations can reduce the initial condition uncertainties and decrease forecast errors. A number of methods were proposed to “target observations”, i.e. to select areas where additional observations are expected to improve considerably the skill of a given forecast. Our proposed approach uses singular vectors to identify the most sensitive regions of the atmospheric flow and to optimally configure the observational network. The observations are placed in locations that have a maximal perturbation energy impact on the verification area at the verification time.

Singular vectors (SVs) are the directions of fastest error growth over a finite time interval [10, 14]. Buizza and Montani [1] showed that SVs can identify the most sensitive regions of the atmosphere for targeted observations. Majudmar et al.[12] compare the SV approach for observation targeting to the ensemble transform Kalman filter. Dăescu and Navon [5] discuss the adaptive observation problem in the context of 4D-Var data assimilation. Estimation of the optimal placement of adaptive observations is also discussed in [7, 11]. Leutbecher [9] derives optimal locations of observations by minimizing the variance of the assimilated field; a computationally tractable problem is obtained by projecting the covariance on the subspace of the dominant singular vectors.

The paper is organized as follows. In Section 2 we introduce the chemical transport models and the concept of singular vectors as the directions of maximal energy growth. Section 3 discusses a maximum energy impact criterion for placing the observations. Numerical results from a simulation of air pollution in East Asia are shown in Section 4. Section 5 summarizes the main findings of this work.

2 Background

2.1 3D Chemical-Transport Models

Chemical transport models solve the mass-balance equations for concentrations of trace species in order to determine the fate of pollutants in the atmosphere [16]. Let c_i be the mole-fraction concentration of chemical species i , Q_i be the rate of surface emissions, E_i be the rate of elevated emissions and f_i be the rate of chemical transformations. Further, u is the wind field vector, K the turbulent diffusivity tensor, and ρ is the air density. The evolution of c_i is described by the following equations

$$\frac{\partial c_i}{\partial t} = -u \cdot \nabla c_i + \frac{1}{\rho} \nabla \cdot (\rho K \nabla c_i) + \frac{1}{\rho} f_i(\rho c) + E_i, \quad t_0 \leq t \leq t_v,$$

$$\begin{aligned}
c_i(t_0, x) &= c_i^0(x) , \\
c_i(t, x) &= c_i^{\text{in}}(t, x) \quad \text{for } x \in \Gamma^{\text{in}} , \quad K \frac{\partial c_i}{\partial n} = 0 \quad \text{for } x \in \Gamma^{\text{out}} , \\
K \frac{\partial c_i}{\partial n} &= V_i^{\text{dep}} c_i - Q_i \quad \text{for } x \in \Gamma^{\text{ground}} , \quad \text{for all } 1 \leq i \leq N_{\text{spec}} .
\end{aligned} \tag{1}$$

Here Γ denotes the domain boundary (composed of the inflow, outflow, and ground level parts) and x is the spatial location within the domain. We will use $\mathcal{M}$ to denote the solution operator of the model (1). The state is propagated forward in time from the “initial” time t_0 to the “verification” time t_v (i.e., the final time of interest)

$$c(t_v) = \mathcal{M}_{t_0 \rightarrow t_v} (c(t_0)) . \tag{2}$$

Perturbations (small errors) evolve according to the tangent linear model (TLM)

$$\delta c(t_v) = \mathbf{M}_{t_0 \rightarrow t_v} \delta c(t_0) , \tag{3}$$

and adjoint variables according to the adjoint model

$$\lambda(t_0) = \mathbf{M}_{t_v \rightarrow t_0}^* \lambda(t_v) . \tag{4}$$

Here M and M^* denote the solution operators of the two linearized models. A detailed description of chemical transport models, and the corresponding tangent linear and adjoint models, is given in [17].

Our main interest is to minimize the forecast uncertainty over a well defined area (the “verification domain” $\Omega_v \subset \Omega$) at a well defined time (the “verification time” t_v). We define a spatial restriction operator G from the entire model domain to the verification domain:

$$G : \Omega \subset \mathbb{R}^n \longrightarrow \Omega_v \subset \mathbb{R}^{n_v} , \quad n_v \ll n . \tag{5}$$

2.2 Singular Vectors

Singular vectors (SVs) determine the most rapidly growing perturbations in the atmosphere. The magnitude of the perturbation at the initial time t_0 is measured in the L^2 norm defined by a symmetric positive definite matrix E

$$\| \delta c(t_0) \|_E^2 = \langle \delta c(t_0) , E \delta c(t_0) \rangle . \tag{6}$$

Similarly, the perturbation magnitude at the verification time t_v is measured in a norm defined by a positive definite matrix F

$$\| \delta c(t_v) \|_F^2 = \langle \delta c(t_v) , F \delta c(t_v) \rangle . \tag{7}$$

We call the norms (6) and (7) squared the “perturbation energies”. The ratio between perturbation energies at t_v (over the verification domain) and at t_0 (over the entire domain) offers a measure of error growth:

$$\sigma^2 = \frac{\| G \delta c(t_v) \|_F^2}{\| \delta c(t_0) \|_E^2} = \frac{\langle \delta c(t_0) , \mathbf{M}_{t_v \rightarrow t_0}^* G^* F G \mathbf{M}_{t_0 \rightarrow t_v} \delta c(t_0) \rangle}{\langle \delta c(t_0) , E \delta c(t_0) \rangle} \tag{8}$$

In (8) we use the fact that perturbations evolve in time according to the dynamics of the tangent linear model (3).

SVs are defined as the directions of maximal error growth, i.e. the vectors $s_k(t_0)$ that maximize the ratio σ^2 in equation (8). These directions are the solutions of the generalized eigenvalue problem

$$\mathbf{M}_{t_v \rightarrow t_0}^* \mathbf{G}^* \mathbf{F} \mathbf{G} \mathbf{M}_{t_0 \rightarrow t_v} s_k(t_0) = \sigma_k^2 \mathbf{E} s_k(t_0) . \quad (9)$$

The left side of (9) involves one integration with the tangent linear model followed by one integration with the adjoint model. The eigenvalue problem (9) can be solved efficiently using the software package ARPACK [8].

Using the square root of the the symmetric positive definite matrix $\mathbf{E}$ the generalized eigenvalue problem (9) can be reduced to a simple eigenvalue problem

$$\mathbf{E}^{-\frac{1}{2}} \mathbf{M}_{t_v \rightarrow t_0}^* \mathbf{G}^* \mathbf{F} \mathbf{G} \mathbf{M}_{t_0 \rightarrow t_v} \mathbf{E}^{-\frac{1}{2}} v_k = \sigma_k^2 v_k(t_0) , \quad v_k = \mathbf{E}^{\frac{1}{2}} s_k(t_0) . \quad (10)$$

Furthermore, $v_k(t_0)$ are the left singular vectors in the singular value decomposition

$$\mathbf{F}^{\frac{1}{2}} \mathbf{G} \mathbf{M}_{t_0 \rightarrow t_v} \mathbf{E}^{-\frac{1}{2}} = \mathbf{U} \cdot \mathbf{\Sigma} \cdot \mathbf{V}^T \quad \text{where} \quad \mathbf{\Sigma} = \text{diag}_k \{ \sigma_k \} , \quad \sigma_k u_k = \mathbf{F}^{\frac{1}{2}} \mathbf{G} s_k(t_v) . \quad (11)$$

The SVs s_k are $\mathbf{E}$ -orthogonal at t_0 and $\mathbf{F}$ -orthogonal at t_v

$$\langle s_k(t_0), \mathbf{E} s_j(t_0) \rangle = 0 \quad \text{and} \quad \langle \mathbf{G} s_k(t_v), \mathbf{F} \mathbf{G} s_j(t_v) \rangle = 0 \quad \text{for } j \neq k . \quad (12)$$

The equations (11) and (12) justify the name of “singular vectors”. The singular value decomposition of the linear operator $M_{t_0 \rightarrow t_v}$, with the $\mathbf{E}$ scalar product at t_0 and the $\mathbf{F}$ scalar product at t_v , has the left singular vectors $s_k(t_0)$ and the right singular vectors $s_k(t_v)$. The singular values σ_k are the error amplification factors along each direction s_k .

The computation of singular vectors in the presence of stiff chemistry is discussed in [17], where computational challenges are reported related to the loss of symmetry due to the stiff nature of equations.

2.3 Perturbation Norms

In numerical weather prediction models variables have different physical meanings (wind velocity, temperature, air density, etc). The energy norms correspond to physical total energy, potential enstrophy, etc. Such norms provide a unified measure for the magnitude of perturbations in variables of different types.

In chemical transport models variables are concentrations of chemical species. Since all variables have the same physical meaning, and similar units, we expect that simple L^2 norms in (6) and (7) will provide a reasonable measure of the “magnitude of the perturbation”. Since concentrations of different species vary by many orders of magnitude we expect that the perturbations of the more abundant species (e.g., CO) will dominate the total perturbation norms. To have a balanced account for the influence of all species it is of interest to consider

the directions of maximal *relative error* growth. For practical reasons [17] it is advantageous to approximate the relative errors by the absolute errors δc_{ijk}^s scaled by “typical” concentration values w_{ijk}^s at each time instant. Therefore the choice of matrices in the norms (6) and (7) is

$$W(t) = \text{diag}_{i,j,k,s} \{w_{i,j,k}^s(t)\} \quad , \quad E = W(t_0)^{-2} \quad , \quad F = W(t_v)^{-2} \quad .$$

One reason for this approximation is that the “typical” concentrations $w_{i,j,k}^s$ can be chosen to be bounded away from zero. More importantly, having the weights independent of the system state c keeps the maximization problem equivalent to a generalized eigenvalue problem.

3 Targeted Chemical Observations

We now determine those locations where perturbations have the largest energy impact over the verification area. For this, consider an initial perturbation vector δ_k equal to zero everywhere, except for one component at a given location where its value is 1. The index k spans all variables in the system, and a particular value of k identifies a single chemical component and a single location.

A certain species at a certain location is perturbed (or, equivalently, is observed, and therefore the perturbation is reduced). This vector can be written in terms of the singular vectors

$$\delta_{i,j,k}^s = \sum_m \alpha_m s_m(t_0) \quad ,$$

where the expansion coefficients can be obtained by the orthogonality relations of the singular vectors

$$\alpha_m = \langle \delta_{i,j,k}^s, E s_m(t_0) \rangle = (E s_m(t_0))_{i,j,k}^s$$

The vector of perturbations at the final time is

$$\delta c(t_v) = \sum_m \alpha_m \mathbf{M}_{t_0 \rightarrow t_v} s_m(t_0) = \sum_m \alpha_m \sigma_m s_m(t_v) \quad .$$

Using the orthogonality of the singular vectors at the final time in the F-norm we have that the total perturbation energy is

$$\mathcal{E}_{i,j,k}^s = \langle G \delta c(t_v), F G \delta c(t_v) \rangle = \sum_m \sigma_m^2 \alpha_m^2 = \sum_m \sigma_m^2 \left((E s_m(t_0))_{i,j,k}^s \right)^2$$

A vector which has each component equal to the energy impact of the corresponding delta initial perturbation is therefore:

$$\mathcal{E} = \sum_m \sigma_m^2 (E s_m(t_0))^2 = \sum_m \zeta_m^2 \quad . \quad (13)$$

The squares of the vectors are considered in an element by element sense. Clearly this sum can be well approximated by the first several terms which correspond to the dominant singular values.

The observations should be located at those points where the energetic impact over the verification area is the largest. These points are easily identified as they are the largest entries of the E vector.

4 Numerical Results

The numerical tests use the state-of-the-art regional atmospheric chemical transport model STEM [2]. The simulation covers a region of $7200 \text{ km} \times 4800 \text{ km}$ in East Asia and the simulated conditions correspond to March 2001. More details about the forward model simulation conditions and comparison with observations are available in [2].

The computational grid has $N_x \times N_y \times N_z$ nodes with $N_x=30$, $N_y=20$, $N_z=18$, and a horizontal resolution of $240 \text{ km} \times 240 \text{ km}$. The chemical mechanism is SAPRC-99 [3] which considers the gas-phase atmospheric reactions of volatile organic and nitrogen oxides in urban and regional settings. The adjoint of the comprehensive model STEM is discussed in detailed in [16]. Both the forward and adjoint chemical models are implemented using KPP [4, 6, 15]. The forward and adjoint models are parallelized using PAQMSG [13]. ARPACK [8] was used to solve the symmetric generalized eigenvalue problems and compute the singular vectors.

We are interested in minimizing the uncertainty in the prediction of ground level ozone concentrations above Korea at 0 GMT March 4, 2001 (shaded area in Figure 1). Thus the “verification area” is Korea, and the “verification time” is

Fig. 1. Optimal placement of chemical observations using the maximum energy impact criterion. Observations of O_3 , NO_2 , $HCHO$, and CO are shown. The verification is ground level ozone over Korea (shaded area) at 0 GMT March 4, 2001. The observations are taken at 6h (circle), 12h (diamond), 18h (downward triangle), 24h (square), 36h (pentagon), and 48h (upward triangle) before verification time.

0 GMT March 4 2001. Most of the uncertainty in the verification region at the final time is determined by the uncertainty along the dominant singular vectors at the initial time. The uncertainty (error) growth rates along each direction are given by the corresponding singular values.

In order to improve predictions within the verification region observations are needed in areas of maximal perturbation impact, determined using the dominant singular vectors. The optimal location of chemical observations at 6h, 12h, 18h, 24h, and 48h before the verification time is illustrated in Figure 1. As expected the optimal location of observations changes in time and drifts away from the verification area for longer intervals. Due to the different roles played by different chemical species in ozone formation, the optimal location of O_3 measurements is different than the optimal location of NO_2 or $HCHO$ observations. For example O_3 can be formed in the presence of NO_2 emissions and then transported over long distances. In contrast, the $HCHO$ lifetime is short and it can produce O_3 only locally.

5 Conclusions

The integration of observations and model predictions through data assimilation is essential for improved forecast capabilities in numerical weather prediction and in air pollution simulations. Atmospheric data assimilation is a parameter estimation problem with millions of degrees of freedom: the optimal analysis state of the atmosphere is found by minimizing the mismatch between model predictions and observations.

This paper develops a computationally tractable approach to target atmospheric chemical observations. The fundamental question being asked is where to place the observations such that, after data assimilation, the uncertainty in the resulting state is minimized. Our approach is based on reducing the system along the subspace defined by the dominant singular vectors, and placing the observations are placed in those locations that have a maximal perturbation energy impact on the verification area at the verification time. Numerical results presented for a simulation of atmospheric pollution in East Asia in March 2001 show that the optimal location of observations depends on the pattern of the flow but is different for different chemical species.

Targeted observations have been previously considered in the context of numerical weather prediction. This work is, to the best of our knowledge, the first effort to target chemical observations in the context of reactive atmospheric flow models.

Acknowledgments. This work was supported by the National Science Foundation through the awards NSF ITR AP&IM 0205198, NSF CAREER ACI-0413872, and NSF CCF-0515170, by the National Oceanic and Atmospheric Administration (NOAA) and by the Texas Environmental Research Consortium (TERC). We would like to thank Virginia Tech's Laboratory for Advanced Scientific Computing (LASCA) for the use of the Anantham cluster.

References

1. R. Buizza and A. Montani. Targeting observations using singular vectors. *Journal of the Atmospheric Sciences*, 56:2965–2985, 1999.
2. G.R. Carmichael et al. Regional-scale chemical transport modeling in support of the analysis of observations obtained during the TRACE-P experiment. *Journal of Geophysical Research*, 108:10649–10671, 2003.
3. W.P.L. Carter. Implementation of the SAPRC-99 chemical mechanism into the Models-3 framework. Technical report, United States Environmental Protection Agency, January 2000.
4. D.N. Daescu, A. Sandu, and G.R. Carmichael. Direct and adjoint sensitivity analysis of chemical kinetic systems with KPP: I – numerical validation and applications. *Atmospheric Environment*, 37:5097–5114, 2003.
5. D.N. Daescu and I.M. Navon. Adaptive observations in the context of 4d-Var data assimilation. *Meteorology and Atmospheric Physics*, 84(4):205–226, 2004.
6. V. Damian, A. Sandu, M. Damian, F. Potra, and G.R. Carmichael. The kinetic pre-processor KPP - a software environment for solving chemical kinetics. *Computers and Chemical Engineering*, 26:1567–1579, 2002.
7. R. Gelaro, R. Buizza, T.N. Palmer, and E. Klinker. Sensitivity analysis of forecast errors and the construction of optimal perturbations using singular vectors. *Journal of the Atmospheric Sciences*, 55:1012–1037, 1998.
8. R. Lehoucq, K. Maschhoff, D. Sorensen, and C. Yang. ARPACK software home page. <http://www.caam.rice.edu/software/ARPACK>.
9. M. Leutbecher. A reduced rank estimate of forecast error variance changes due to intermittent modifications of the observing network. *Journal of the Atmospheric Sciences*, 60:729–742, 2003.
10. E.N. Lorenz. A study of the predictability of a 28 variable atmospheric model. *Tellus*, 17:321–333, 1965.
11. E.N. Lorenz and K.A. Emanuel. Optimal sites for supplementary observations: simulation with a small model. *Journal of the Atmospheric Sciences*, 55:399–414, 1998.
12. S.J. Majumdar, C.H. Bishop, R. Buizza, and R. Gelaro. A comparison of ensemble transform Kalman filter targeting guidance with ECMWF and NRL total-energy singular vector guidance. *Quarterly Journal of the Royal Meteorological Society*, 128:2527–2549, 2002.
13. P. Miehe, A. Sandu, G.R. Carmichael, Y. Tang, and D. Daescu. A communication library for the parallelization of air quality models on structured grids. *Atmospheric Environment*, 36:3917–3930, 2002.
14. F. Molteni and T.N. Palmer. Predictability and finite-time instability of the Northern winter circulation. *Quarterly Journal of the Royal Meteorological Society*, 119:269–298, 1993.
15. A. Sandu, D. Daescu, and G.R. Carmichael. Direct and adjoint sensitivity analysis of chemical kinetic systems with KPP: I – Theory and software tools. *Atmospheric Environment*, 37:5083–5096, 2003.
16. A. Sandu, D. Daescu, G.R. Carmichael, and T. Chai. Adjoint sensitivity analysis of regional air quality models. *Journal of Computational Physics*, 204:222–252, 2005.
17. W. Liao, A. Sandu, G.R. Carmichael, and T. Chai. Total Energy Singular Vector Analysis for Atmospheric Chemical Transport Models. *Monthly Weather Review*, accepted, 2005.