Singular Vector Analysis
for Atmospheric Chemical Transport Models

Wenyuan Liao and Adrian Sandu†
Department of Computer Science
Virginia Polytechnic Institute and State University
Blacksburg, VA 24061
†A. Sandu is the corresponding author (sandu@cs.vt.edu)

Gregory R. Carmichael and Tianfeng Chai
Center for Global and Regional Environmental Research
The University of Iowa, Iowa City, IA 52240

Submitted to Monthly Weather Review on October 26, 2005
Abstract

The singular vectors of a chemical transport model are the directions of maximum perturbation growth over a finite time interval. They have proved useful for the estimation of error growth, the initialization of ensemble forecasts, and the optimal placement of adaptive observations.

The aim of this paper is to address computational aspects of singular vector analysis for atmospheric chemical transport models. The distinguishing feature of these models is the presence of stiff chemical interactions. We discuss a projection approach to preserve the symmetry of the tangent linear–adjoint operator for stiff systems, and extend it to 3D chemical transport simulations. Numerical results are presented for a simulation of atmospheric pollution in East Asia in March 2001. The singular values and the structure of the singular vectors depend on the length of the simulation interval, the meteorological data, the location of the optimization region and the selection of optimization species, the choice of error norms, and the size of the optimization region.

Keywords: Chemical transport models, singular vectors, adjoint models, sensitivity analysis, data assimilation, targeted observations.
1 Introduction

The singular vectors of a chemical transport model are the directions of fastest perturbation growth over a finite time interval. Singular vector analysis was introduced in meteorology by Lorenz (1965), who computed the largest error growth rates in an idealized model of the atmosphere. The adjoint technique was used by Molteni and Palmer (1993) and Mureau et al. (1993) to compute singular vectors for meteorological models. Singular vector analysis of general circulation models with millions of variables is now possible (see e.g., (Buizza and Palmer, 1995; Li, Navon, and Hussaini, 2005)).

Applications of singular vector analysis in numerical weather prediction include: (1) normal mode analysis of atmospheric flow instability, estimation of error growth, and the assessment of atmospheric predictability (Borges and Hartmann, 1992; Ehrendorfer and Tribbia, 1997; Molteni and Palmer, 1993; Oortwijn, 1998); (2) initialization of ensemble forecasts (Molteni et al., 1996; Mureau et al., 1993) and (3) estimation of the optimal placement of adaptive observations (Buizza and Montani, 1999; Gelaro et al., 1998; Lorenz and Emanuel, 1998; Palmer et al., 1998).

Numerous studies have shown that the structure of singular vectors in atmospheric general circulation models is determined by: (1) the atmospheric episode under consideration (Farrel, 1988; Barkmeijer et al., 2001); (2) model physics (e.g., the treatment of boundary layer processes) (Barkmeijer et al., 2001; Buizza and Montani, 1999; Buizza and Palmer, 1995; Ehrendorfer et al., 1999; Mahfouf, 1999); (3) model resolution (Buizza and Montani, 1999; Buizza and Palmer, 1995; Ehrendorfer et al., 1999; Mahfouf, 1999); and (4) the particular choice of error norms (Kuang, 2004).
The objective of this work is to study the singular vectors for atmospheric chemical transport models. The distinguishing feature of these models is the presence of chemical interactions between tracer species, which leads to a very stiff, nonlinear system of partial differential equations. This poses nontrivial challenges in the computation of singular vectors, and allows interesting interpretations stemming from the complex interactions between emission sources, chemical transformations, transport, and deposition processes.

While in the study of atmospheric dynamics the dominant singular vectors are associated with unstable modes, in the study of chemical transport systems the dominant singular vectors are useful to describe the uncertainty in a limited subdomain (e.g., where the model prediction needs to be improved). In this paper we illustrate the use of singular vectors to describe uncertainties in the initial conditions. Other very important sources of uncertainty in air quality models that need to be quantified in real applications are the emissions, the meteorological fields, the deposition velocities, and the top and lateral boundary conditions for regional models.

The paper is organized as follows. In Section 2 we introduce the chemical transport model singular vectors and discuss two of their possible applications in the context of air pollution modeling. Computational aspects of chemical singular vectors are discussed in Section 3. An introduction to chemical transport modeling and the formulation of the tangent linear and adjoint models is presented in Section 4. Several possible perturbation norms are discussed in Section 5. Numerical results from a simulation of air pollution in East Asia are shown in Section 6. Section 7 summarizes the main findings of this work.
2 Singular Vectors and Chemical Transport Models

An atmospheric chemical transport model (CTM) propagates the model state forward in time from the “initial” state \( x(t_0) \) to the “final” state \( x(t_F) \). With \( \mathcal{M} \) denoting the model solution operator we have

\[
x(t_F) = \mathcal{M}_{t_0 \rightarrow t_F} (x(t_0)) .
\] (1)

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

\[
\delta x(t_F) = \mathcal{M}_{t_0 \rightarrow t_F} \delta x(t_0) ,
\] (2)

and adjoint variables according to the adjoint model

\[
\lambda(t_0) = \mathcal{M}^*_{t_F \rightarrow t_0} \lambda(t_F) .
\] (3)

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, will be given in Section 4. The error covariance matrix propagates from \( P(t_0) \) to \( P(t_F) \) according to

\[
P(t_F) = \mathcal{M}_{t_0 \rightarrow t_F} P(t_0) M^*_{t_F \rightarrow t_0} + Q .
\] (4)

The additional term \( Q \) represents the covariance of the model errors.

2.1 Singular Vectors

Singular vectors 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 \( A \)

\[
\| \delta x(t_0) \|^2_A = \langle \delta x(t_0) , A \delta x(t_0) \rangle .
\] (5)
Similarly, the perturbation magnitude at the final time $t_F$ is measured in a semi-norm defined by a semi-positive definite matrix $B$

$$\| \delta x(t_F) \|_B^2 = \langle \delta x(t_F), B \delta x(t_F) \rangle .$$

(6)

We call the norms (5) and (6) squared the “perturbation magnitudes”. The ratio between perturbation magnitudes at $t_F$ and $t_0$ offers a measure of error growth:

$$\sigma^2 = \frac{\| \delta x(t_F) \|_B^2}{\| \delta x(t_0) \|_A^2} = \frac{\langle \delta x(t_F), B \delta x(t_F) \rangle}{\langle \delta x(t_0), A \delta x(t_0) \rangle}$$

(7)

$$= \frac{\langle M_{t_0 \rightarrow t_F} \delta x(t_0), B M_{t_0 \rightarrow t_F} \delta x(t_0) \rangle}{\langle \delta x(t_0), A \delta x(t_0) \rangle}$$

$$= \frac{\langle \delta x(t_0), M_{t_F \rightarrow t_0}^* B M_{t_0 \rightarrow t_F} \delta x(t_0) \rangle}{\langle \delta x(t_0), A \delta x(t_0) \rangle}$$

In (7) we use the fact that perturbations evolve in time according to the dynamics of the tangent linear model (2).

Singular vectors are defined as the directions of maximal error growth, i.e. the vectors $s_k(t_0)$ that maximize the ratio $\sigma^2$ in equation (7). These directions are the solutions of the generalized eigenvalue problem

$$M_{t_F \rightarrow t_0}^* B M_{t_0 \rightarrow t_F} s_k(t_0) = \sigma_k^2 A s_k(t_0) .$$

(8)

The left side of (8) involves one integration with the tangent linear model followed by one integration with the adjoint model. The eigenvalue problem (8) can be solved efficiently using the software package ARPACK (Lehoucq et al.), or its parallel version PARPACK (Maschhoff et al).

Using the square root of the the symmetric positive definite matrix $A$ the generalized eigenvalue problem (8) can be reduced to an ordinary eigenvalue problem

$$A^{-\frac{1}{2}} M_{t_F \rightarrow t_0}^* B M_{t_0 \rightarrow t_F} A^{-\frac{1}{2}} v_k(t_0) = \sigma_k^2 v_k(t_0) , \quad v_k(t_0) = A^{\frac{1}{2}} s_k(t_0) .$$

(9)
Furthermore, \( v_k(t_0) \) are the left singular vectors in the singular value decomposition

\[
 B^{\frac{1}{2}} M_{t_0 \to t_F} A^{-\frac{1}{2}} = U \cdot \text{diag}(\sigma_k) \cdot V^T \quad \text{where} \quad \sigma_k u_k(t_F) = B^{\frac{1}{2}} s_k(t_F). \tag{10}
\]

The singular vectors \( s_k \) are \( A \)-orthogonal at \( t_0 \) and \( B \)-orthogonal at \( t_F \)

\[
 \langle s_k(t_0), A s_j(t_0) \rangle = 0 \quad \text{and} \quad \langle s_k(t_F), B s_j(t_F) \rangle = 0 \quad \text{for} \quad j \neq k. \tag{11}
\]

The singular value decomposition of the linear operator \( M_{t_0 \to t_F} \), with the \( A \) scalar product at \( t_0 \) and the \( B \) scalar product at \( t_F \), has the left singular vectors \( s_k(t_0) \) and the right singular vectors \( s_k(t_F) \). If the same norms are used at the initial and at the final times the singular values \( \sigma_k \) can be interpreted as the error amplification factors along each direction \( s_k \).

A special set of perturbation norms is provided by the choice \( B = I \) and \( A = P(t_0)^{-1} \). In this case the resulting singular vectors \( s_k(t_0) \) evolve into the leading eigenvectors \( s_k(t_F) \) of the forecast error covariance matrix \( P(t_F) \),

\[
P(t_F) s_k(t_F) = \sigma_k^2 s_k(t_F). \tag{12}
\]

The eigenvectors \( s_k(t_F) \) are called analysis error covariance singular vectors (Ehrendorfer and Tribbia 1997; Barkmeijer et al. 1999). Since the leading eigenvectors of \( P(t_F) \) are the directions of maximum variance of forecast error, the singular vectors define the directions along which we must do a good job of analysis in order to minimize the forecast error at \( t_F \).

We assume that the model error in equation (4) is negligible over the period \([t_0, t_F]\). From equation (12) it follows that the singular vectors are the solutions of the following generalized eigenvalue problem

\[
 M^*_F M_{t_0 \to t_F} s_k(t_0) = \lambda_k P(t_0)^{-1} s_k(t_0) = \lambda_k \left( \nabla^2 J \right) s_k(t_0) \tag{13}
\]
When calculated as above the singular vectors are referred to as “Hessian singular vectors” (HSV) (Barkmeijer et al. 1998). The second relation comes from the fact that the Hessian matrix of the analysis cost function $J$ in the variational analysis system is an estimate of the inverse of the analysis covariance matrix. This motivates the name Hessian singular vectors for the solutions $s_k(t_0)$ of the eigenproblem (13).

In the study of atmospheric dynamics the dominant singular vectors are associated with unstable modes. In the study of chemical transport systems the dominant singular vectors are useful to describe the uncertainty in a limited subdomain (e.g., where the model prediction needs to be improved). Limited subdomain studies are also of interest in numerical weather prediction (Hersbach et al. 2003).

### 2.2 Initialization of Ensemble Forecasts

A critical element for accurate simulations is the use of observational data to constrain model predictions. Widely used data assimilation techniques include 3D-Var, 4D-Var, Kalman filter and ensemble nonlinear filters. Kalman filter techniques are discussed in (Daley, 1991; Jazwinski, 1970). Consider a set of observations $y$ at $t_F$ (assumed, for simplicity, to be a linear function of model state, $y = Hx$). The extended Kalman filter uses the forecast state and its covariance $(x(t_F), P(t_F))$ and the observations and their covariance $(y, R)$ to produce an optimal (“analyzed”) estimation of the model state and its covariance $(x_A(t_F), P_A(t_F))$:

$$
x_A(t_F) = x(t_F) + P(t_F) H^T \left( R + H P(t_F) H^T \right)^{-1} (y - H x(t_F))
$$

$$
P_A(t_F) = P(t_F) - P(t_F) H^T \left( R + H P(t_F) H^T \right)^{-1} H P(t_F)
$$

6
The computational expense of the Kalman filter (14) is extremely large because one needs to invert the matrix $R + HPH^T$ and apply the tangent linear model to each column and the adjoint model to each row of the covariance matrix (Fisher, 2001). The commonly used method to reduce the computational cost is to propagate (only) the projection of the covariance matrix onto a low-dimensional subspace ($\text{span}\{s_1, \ldots, s_k\}$). The ensemble Kalman filter (Houtekamer and Mitchell, 2000) uses a Monte-Carlo approach to define this subspace and to approximate the time-evolving covariance matrix.

The subspace (i.e., the ensemble of perturbations at the analysis time $t_F$) should contain the directions $s_k(t_F)$ along which the error has the maximal growth. Consequently the initial ensemble should be defined based on the singular vectors $s_k(t_0)$. This approach is used at the European Center for Medium-Range Weather Forecasts (ECMWF) to generate initial perturbations for ensemble forecasts (Buizza et al., 2000; Buizza and Palmer, 1995; Hamill et al., 2003).

### 2.3 Targeted Observations

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. Singular vectors identify sensitive regions of the atmospheric flow and can be used to optimally configure the observational network.

Singular vectors can identify the most sensitive regions of the atmosphere for targeted observations as long as the linearity assumption of error propagation holds (Hansen and

3 Computation of Singular Vectors for Chemical Models

In this section we discuss the computational challenges associated with the chemical singular vectors and propose an approach for calculating them accurately.

A numerical eigenvalue solver applied to equation (8) requires a symmetric matrix $M^*BM$ in order to successfully employ Lanczos iterations, and to guarantee that the numerical eigenvalues are real. There are two approaches to computing adjoints: continuous and discrete. In the continuous approach the adjoint of the continuous differential equations is derived, then solved numerically. In the discrete approach the numerical solution is (considered to be) the forward model and its adjoint is constructed. The symmetry requirement imposes to use the discrete adjoint $M^*$ of the tangent linear operator $M$ in equation (8). The computation of discrete adjoints for stiff systems is a nontrivial task (Sandu et al., 2003). In addition, computational errors (which can destroy symmetry) have to be small.

For a given model a symmetry indicator is constructed based on two random perturbation
vectors $u(t_0)$ and $v(t_0)$ which are propagated forward in time,

$$u(\tau) = M_{t_0\rightarrow \tau} u(t_0) \quad \text{and} \quad v(\tau) = M_{t_0\rightarrow \tau} v(t_0).$$

The symmetry residual is the difference

$$r(\tau) = \langle u(\tau), M^*_{t_\tau u} M_{t_\tau t_\tau} v(\tau) \rangle - \langle v(\tau), M^*_{t_\tau v} M_{t_\tau t_\tau} u(\tau) \rangle$$

If $M^*$ is the exact discrete adjoint of $M$ then $r(\tau) = 0$ for all $\tau$. However, both $M$ and $M^*$ are evaluated numerically and in practice we expect the symmetry residual $r(\tau)$ to have small (but nonzero) values.

To illustrate possible problems with losing the symmetry we consider the SAPRC-99 atmospheric gas-phase reaction mechanism (Carters, 2000) which has 93 species and 235 reactions. The forward, tangent linear, and adjoint models are implemented using the automatic code generator KPP (Damian et al., 2002; Daescu et al., 2003; Sandu et al., 2003). Several numerical experiments revealed that the magnitude of the symmetry residual depends on the choice of numerical integrator. Among the Rosenbrock integrators available in KPP we selected Rodas4 (Sandu et al., 2003) which performs best with respect to symmetry. The variation of $r(\tau)$ with time is shown in Figure 1(a) (solid line). Surprisingly, the symmetry is lost during the stiff transient at the beginning of the integration interval, where the symmetry residual jumps from $10^{-16}$ to $10^{-2}$.

The loss of symmetry is due to the stiffness of the chemical terms. To understand this behavior we consider a singular perturbation model for the chemical system of the form

$$\frac{dy}{dt} = f(y, z), \quad \epsilon \frac{dz}{dt} = g(y, z), \quad \epsilon \ll 1.$$

A singular perturbation problem (17) is a model ordinary differential equation that explicitly
separates the dynamics of the slow component $y$ and of the fast component $z$. This model problem is widely used in the theoretical study of the behavior of stiff systems and of the stiff numerical methods (Hairer et al., 2004).

Perturbations propagate through the tangent linear model of (17)

$$
\frac{d\delta y}{dt} = f_y(y,z)\delta y + f_z(y,z)\delta z, \quad \epsilon \frac{d\delta z}{dt} = g_y(y,z)\delta y + g_z(y,z)\delta z.
$$

(18)

For $\epsilon \to 0$, the perturbation vectors in (18) are of the form

$$
\delta z = -g_z^{-1}(y,z)g_y(y,z)\delta y.
$$

(19)

During the numerical computation of the eigenvectors ARPACK (or any other solver) generates vectors $x = [\delta y, \delta z]^T$ which do not satisfy equation (19). These vectors are the initial conditions for the tangent linear model and are propagated forward, then backward through the adjoint model, in order to evaluate the matrix-vector products $M^*B M x$. Strong, artificial transients appear in the tangent linear model due to the fact that the initial perturbations are away from the slow manifold described by equation (19).

To correct this we apply the tangent linear model on the initial perturbation for a short time, which is equivalent to “projecting” the initial perturbation onto the slow evolution manifold (19). The result is then used to initialize the subsequent tangent linear model run. In order to preserve operator symmetry, another projection using the adjoint model needs to be performed at the end of the adjoint integration. Consequently the matrix-vector products are computed as

$$
w = \Pi^* M_{t_f\to t_0}^* B M_{t_0\to t_f} \Pi x,
$$

(20)

where $\Pi$ and $\Pi^*$ denote the projection operations performed with the tangent linear and the
adjoint models respectively. Note that in the 3D model the projections will be performed at the beginning and at the end of each timesplit interval.

Numerical tests revealed that a small number of projection steps is sufficient in practice to substantially enhance symmetry. Fig.1(b) presents the evolution of the symmetry residual with the number of projection steps. The symmetry is markedly improved after only 2 projection steps.

Fig.1 (a) (dashed) presents the evolution of the symmetry residual when 6 projection steps are performed with the very small stepsize of $10^{-9}$ seconds. The symmetry error during the stiff transient is reduced to $10^{-11}$. Note that the projection time step is of the order of the fastest scales in the system, and is much smaller than the numerical integration time step.

We next extend these results to 3D chemical transport models.

4 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. In this section we briefly describe the governing mass balance equations and the tangent linear and adjoint models; a detailed discussion is given in (Sandu et al., 2005). Tangent-linear and adjoints of transport models are discussed in (Daley, 1995; Vukicevic and Hess, 2000).

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, $V_{i}^{\text{dep}}$ the deposition velocity, and $f_i$ be the rate of chemical transformations. Further, the inflow/outflow/ground boundaries of the
computational domain are denoted by \( \Gamma^\text{in}/\Gamma^\text{out}/\Gamma^\text{ground} \) respectively, \( u \) is the wind field vector, \( K \) the turbulent diffusivity tensor, and \( \rho \) the air density. The evolution of \( c_i \) is described by the following equations

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

\[
c_i(t_0, x) = c^0_i(x),
\]

\[
c_i(t, x) = c^\text{in}_i(t, x) \quad \text{for} \quad x \in \Gamma^\text{in}, \quad K \frac{\partial c_i}{\partial n} = 0 \quad \text{for} \quad x \in \Gamma^\text{out},
\]

\[
K \frac{\partial c_i}{\partial n} = V_i^\text{dep} c_i - Q_i \quad \text{for} \quad x \in \Gamma^\text{ground}, \quad \text{for all} \quad 1 \leq i \leq N^\text{spec}.
\]

We refer to the equations (21) as the **forward model**.

A perturbation \( \delta c^0 \) of the initial conditions will result in perturbations \( \delta c(t) \) of the concentration field at later times. The evolution of these perturbations is governed by the equations:

\[
\frac{\partial \delta c_i}{\partial t} = -u \cdot \nabla \delta c_i + \frac{1}{\rho} \nabla \cdot (\rho K \nabla \delta c_i) + F_i(\rho c) \delta c, \quad t_0 \leq t \leq t_F
\]

\[
\delta c_i(t_0, x) = \delta c^0_i(x),
\]

\[
\delta c_i(t, x) = 0 \quad \text{for} \quad x \in \Gamma^\text{in}, \quad K \frac{\partial \delta c_i}{\partial n} = 0 \quad \text{for} \quad x \in \Gamma^\text{out},
\]

\[
K \frac{\partial \delta c_i}{\partial n} = V_i^\text{dep} \delta c_i \quad \text{for} \quad x \in \Gamma^\text{ground}, \quad \text{for all} \quad 1 \leq i \leq N^\text{spec}.
\]

Equations (22) are referred to as the **tangent linear model** associated with the forward model (21). Here \( F = \partial f / \partial c \) denotes the Jacobian of the chemical rate function \( f \), and \( F_i,^* \) is its \( i \)-th row.

The **continuous adjoint model** associated with the forward model (21) (or, more exactly, the adjoint of the tangent linear model (22)) describes the evolution of the adjoint variables
\[ \lambda_i: \]
\[ \frac{\partial \lambda_i}{\partial t} = -\nabla \cdot (u \lambda_i) - \nabla \cdot \left( \frac{\rho K \nabla \lambda_i}{\rho} \right) - (F^T(\rho c) \lambda_i) - \phi_i, \quad t_F \geq t \geq t_0 \]
\[ \lambda_i(t_F, x) = \lambda_i^F(x), \]
\[ \lambda_i(t, x) = 0 \quad \text{for} \quad x \in \Gamma^{\text{in}}, \quad \lambda_i u + \rho K \frac{\partial (\lambda_i/\rho)}{\partial n} = 0 \quad \text{for} \quad x \in \Gamma^{\text{out}}, \quad (23) \]
\[ \rho K \frac{\partial (\lambda_i/\rho)}{\partial n} = V_i^{\text{dep}} \lambda_i \quad \text{for} \quad x \in \Gamma^{\text{ground}}, \quad \text{for all} \quad 1 \leq i \leq N_{\text{spec}}. \]

The ground boundary condition is obtained from the fact that \( \vec{u} \cdot \vec{n} = 0 \) at ground level. The forcing function \( \phi_i \) depends on the particular cost functional under consideration (Sandu et al., 2005).

The numerical solution operator for (21) is based on an operator splitting approach, where the transport and the chemistry steps are taken successively. With \( T \) the numerical solution operator for transport, and \( C \) the solution operator for chemistry, one step of the solution operator reads
\[ c^{m+1} = T_{\Delta t/2} \cdot C_{\Delta t} \cdot T_{\Delta t/2} \cdot c^m. \quad (24) \]

The tangent linear model of (24) is constructed from the tangent linear transport \( (T) \) and chemistry \( (C) \) operators. As explained in Section 3 a projection onto the chemical slow manifold \( (\Pi) \) is applied before each linearized chemistry,
\[ \delta c^{m+1} = T_{\Delta t/2} \cdot C_{\Delta t} \cdot \Pi \cdot T_{\Delta t/2} \cdot \delta c^m. \quad (25) \]

The discrete adjoint model is based on the discrete adjoints of the transport \( (T^*) \) and chemistry \( (C^*) \) numerical schemes. A chemical adjoint projection \( (\Pi^*) \) is applied after each adjoint chemistry step,
\[ \lambda^m = T^*_{\Delta t/2} \cdot \Pi^* \cdot C^*_{\Delta t} \cdot T^*_{\Delta t/2} \cdot \lambda^{m+1} + \phi^m. \quad (26) \]
The projection $\Pi (\Pi^*)$ is applied at every operator split step, since the transport step perturbs the chemical quasi-equilibria.

5 Error Norms

In numerical weather prediction models variables have different physical units (wind velocity, temperature, air density, etc). The energy norms provide a unified measure for the magnitude of perturbations in variables of different dimensions.

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 will provide a reasonable measure of the magnitude of the perturbation

$$\| \delta c \|^2_A = \sum_{i,j,k} \sum_{s=1}^{N_{\text{spec}}} (\delta c^s_{i,j,k})^2 = \langle \delta c, A \delta c \rangle$$

Here $c^s_{i,j,k}$ denotes the concentration of chemical species $s$ at the grid point $(i, j, k)$ in the discrete model, and $\delta c^s_{i,j,k}$ is its perturbation.

In practice we are interested to assess the influence of perturbations onto a well-defined optimization area, and a given set of chemical species. In this case the perturbation magnitude at the final time is defined in terms of a local projection operator (Buizza 1994), i.e. a matrix with the diagonal entries equal to one for the selected optimization variables and equal to zero elsewhere,

$$\| \delta c(t_F) \|^2_B = \sum_{\text{optimization area}} \sum_{\text{optimization spec.}} (\delta c^s_{i,j,k})^2 = \langle \delta c(t_F), B \delta c(t_F) \rangle$$

with $B = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}$. (27)

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 error.
norm (27). To have a balanced account for the influence of all species it is of interest to consider the directions of maximal relative error growth, i.e. the directions which maximize

$$\sigma^2 = \frac{\langle \delta c(t_F) / c(t_F), B \delta c(t_F) / c(t_F) \rangle}{\langle \delta c(t_0) / c(t_0), A \delta c(t_0) / c(t_0) \rangle} .$$

(28)

This can be accomplished by using the logarithms of the concentrations $\log c_{i,j,k}$ as model variables. In practice it is advantageous to approximate the relative errors by the absolute errors $\delta c_{i,j,k}$ scaled by “typical” concentration values $w_{i,j,k}$

$$\frac{\delta c_{i,j,k}(t_F)}{c_{i,j,k}(t_F)} \approx \frac{\delta c_{i,j,k}(t_F)}{w_{i,j,k}(t_F)} \Rightarrow \frac{\delta c(t_F)}{c(t_F)} \approx W_{t_F}^{-1} \delta c(t_F) \text{ where } W_t = \text{diag}_{i,j,k,s} \{ w_{i,j,k}(t) \} .$$

and maximize

$$\sigma^2 = \frac{\langle W_{t_F}^{-1} \delta c(t_F), B W_{t_F}^{-1} \delta c(t_F) \rangle}{\langle W_{t_0}^{-1} \delta c(t_0), A W_{t_0}^{-1} \delta c(t_0) \rangle} .$$

(29)

One reason for this approximation is that the “typical” concentrations $w_{i,j,k}$ can be chosen to be bounded away from zero. More importantly, having the weights independent of the system state $c$ keeps the maximization problem (29) equivalent to a generalized eigenvalue problem

$$M_{t_F}^{*} \rightarrow t_0 W_{t_F}^{-1} B W_{t_F}^{-1} M_{t_0} \rightarrow t_F s_k(t_0) = \sigma_k^2 A W_{t_0}^{-2} s_k(t_0) .$$

(30)

For the important case where $A = I$ the generalized eigenvalue problem (30) is equivalent to the following eigenvalue problem:

$$W_{t_F}^{-1} M_{t_F}^{*} \rightarrow t_0 W_{t_F}^{-1} B W_{t_F}^{-1} M_{t_0} \rightarrow t_F W_{t_0} v_k = \sigma_k^2 v_k , \quad W_{t_0} v_k = W_{t_0}^{-1} s_k(t_0) .$$

(31)

The weight matrices $W$ can be different at the initial and at the final time in order to account for different “typical concentration” levels at night time vs. day time.
6 Numerical Results

The numerical tests use the state-of-the-art regional atmospheric chemical transport model STEM (Carmichael et al., 2003). The simulation covers a region of 7200 Km × 4800 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 (Carmichael et al., 2003).

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 Km × 240 Km. The chemical mechanism is SAPRC-99 (Carters, 2000) which considers the gas-phase atmospheric reactions of volatile organic and nitrogen oxides in urban and regional settings. The meteorological fields have been computed using the Regional Atmospheric Modeling System (RAMS), and analyzed off-line (data assimilation of the meteorological observations has been performed before the chemical transport simulations). The initial and boundary conditions have been obtained from a long run of the model before the start time of the current computations; details can be found in (Carmichael et al., 2003). While the simulations described in (Carmichael et al., 2003) use a grid resolution of 80 Km × 80 Km, in the current paper we use a coarser grid in order to reduce the cpu time needed by the singular vector calculations.

The adjoint of the comprehensive model STEM is discussed in detailed in (Sandu et al., 2005). Both the forward and adjoint chemical models are implemented using KPP (Damian et al., 2002; Daescu et al., 2003; Sandu et al., 2003). The forward and adjoint models are parallelized using PAQMSG (Miehe et al., 2002). PARPACK (Lehoucq et al.) was used to solve the symmetric generalized eigenvalue problems.
The singular vectors $s(N, N_Y, N_Z, N_{\text{spec}})$ in equation (8) are represented by four-dimensional arrays. To visualize them we consider separately the vector sections corresponding to different chemical species. Further, each three-dimensional section is reduced to a two-dimensional “top” view by adding the values in each vertical column, or to a two-dimensional “South” view by adding the values in each North-South column.

Numerical results for different optimization regions, optimization species, simulation intervals, meteorological data, and error norms are presented next.

6.1 Singular Vectors for Different Simulation Intervals

We first consider the case where the optimization criterion is the ground level ozone concentration in a $720 \text{Km} \times 960 \text{Km}$ area covering Korea. The singular vector analysis presented next will help answer the following questions:

- In which areas will small changes in the initial conditions grow fastest to impact the ozone levels over Korea after 12, 24, and 48 hours? What is their rate of growth?

- How should the initial perturbations be constructed for ensemble simulations in order to properly describe the uncertainty in ground ozone predictions over Korea after 12, 24, and 48 hours?

- Where are additional observations needed the most in order to improve 12, 24, and 48 hours predictions of ground level Korean ozone?

The largest 12 singular values for 12h, 24h and 48h simulations started at 0 GMT March 1st, 2001 are shown in Figure 2. The rapid decrease of the eigenvalue magnitude indicates that
the uncertainty in the ground level ozone within the optimization region can be captured with only a few singular vectors. The eigenvalues decrease faster for longer simulation intervals.

The meteorological conditions are an important factor in determining the singular vectors. The 2 Km level wind fields for the simulation interval under consideration are shown in Figure 3.

The top and South views for $O_3$ sections of the dominant et als for the 12h, 24h and 48h simulations are presented in Figure 4. Singular vectors are localized near the optimization area in both the horizontal and the vertical directions. As expected, for longer simulation intervals the singular vectors spread further away from the optimization region. The singular vectors are not confined to the lowest layers, but also show important regions located between 1 – 3 km. This is due to the transport processes, which exchange material from the surface into the free troposphere, and which bring free tropospheric air back to the surface. This has important implications for the design of measurement systems, since it shows that surface measurements alone are not sufficient for a correct representation of ground level concentrations.

Several dominant singular vectors for the 12h, 24h, and 48h simulations starting at 0 GMT March 1, 2001 are shown in Figures 4, 5 and 6 ($O_3$ sections), Figure 7 ($NO_2$ sections), and Figure 8 ($HCHO$ sections). $NO_2$ and $HCHO$ are important species involved in the photochemical production of ozone. They also are species that are directly emitted into the atmosphere as a result of combustion processes.

A close look at the structure of the dominant singular vectors reveals that:

- Since different singular vectors are orthogonal they contain different information about
the areas of maximal error growth.

- The eigenvectors evolve in time as the length of the simulation interval increases. They tend to expand further away from the optimization area, illustrating that perturbations in a wider area at earlier times impact the optimization area.

- The shapes and the magnitudes of the $O_3$, $NO_2$ and $HCHO$ sections show subtle differences, illustrating the different influences that these species have on ground level $O_3$ after 12h, 24h and 48h.

Singular vectors contain valuable information for the placement of additional observations in order to improve predictions of ground level Korean ozone. According to this analysis, additional $O_3$ measurements may have to be placed in a different location than additional $NO_2$ or $HCHO$ observations. Moreover, the optimal location of observations changes in time and drifts away from the optimization area for longer intervals. In conclusion, what is needed is a well thought out distribution of sites measuring many parameters simultaneously.

### 6.2 Evolved Singular Vectors

The perturbations initialized along each dominant singular vectors develop in time becoming “evolved singular vectors”. We are interested in the shape of these perturbations at the end of the 24 hours simulation interval. The evolved singular vectors (scaled to have the A-norm equal to one) are displayed in Figure 9. The largest values of singular vectors are clustered above the optimization region. This is expected since the SVs are constructed to optimize final-time perturbation norm in the optimization region.
6.3 The Linearity Assumption

Inherent in the singular vector calculation is the assumption that small perturbations propagate according to the tangent linear model dynamics. To assess the validity of this linearity assumption we perturb the initial state with scaled versions each of the first five singular vectors. The scaling is chosen such that the ground level ozone perturbations are $\sim 5 - 10\%$ of the reference ozone values. The perturbed initial state is propagated forward for 24 hours using the full, nonlinear model. The perturbation at the final time is the difference between the perturbed and the reference final states. The B-norms of the evolved perturbations are divided by the A-norms of the initial perturbations. The results shown in Table 1 reveal that the perturbation magnitude ratios approximate well the singular values.

The structure of the nonlinearly evolved perturbations at final time are shown in Figure 10. The evolved perturbation structure is similar to that of the evolved singular vectors shown in Figure 9. Since both the magnitude and the structure of the linearly evolved perturbations match those of the nonlinearly evolved perturbations we conclude that the linearity assumption holds (at least) for the 24 hour simulation interval under consideration.

We next consider a random perturbation vector $r$ with components drawn from a uniform distribution with amplitude $\pm 10\%$ of the initial concentrations $(r_{i,j,k}^s(t_0) \in \mathcal{U} [0.9 r_{i,j,k}^s(t_0), 1.1 r_{i,j,k}^s(t_0)])$. The perturbation evolved for 24 hours and its B-norm was taken. The perturbation components along each of the singular vectors, $r_i(t_0) = \langle r(t_0), A s_i(t_0) \rangle s_i(t_0)$, evolve into $r_i(t_F) = \sigma_i \langle r(t_0), A s_i(t_0) \rangle s_i(t_F)$. The total B-norm of the first $n$ components of the perturbation is $\sum_{i=1}^{n} \sigma_i^2 \langle r(t_0), A s_i(t_0) \rangle^2$. The results in Table 1 (last row) show that the evolved perturbation components along the first 12 singular vectors account for virtually
all the B-norm of the random perturbation at final time. This result confirms the fact that perturbation effects can be captured using only a small subspace of dominant singular vectors.

6.4 Singular Vectors versus Adjoint

To illustrate the difference between the information conveyed by the singular vectors and by the adjoint variables we consider again the ground level $O_3$ in the Korea optimization area and focus on the 24h simulation starting at 0 GMT March 1, 2001. The cost function in the adjoint calculation is the sum of squared ground level $O_3$ concentrations in the optimization area. The adjoint variables are computed through a 24h backward integration and are shown in Figure 11.

To assess the relationship between the adjoint variable $\lambda(t_0)$ and different singular vectors $s_k(t_0)$ we consider the correlation coefficients $p_k = (\lambda(t_0), As_k(t_0))/\|\lambda(t_0)\|_A \|s_k(t_0)\|_A)$. Specifically, we compute the correlations between individual (and homologous) sections of $\lambda(t_0)$ and $s_k(t_0)$. The results are shown in Figure 12. For all sections the correlation of the adjoint and the first singular vector is the strongest. The $O_3$ section of the adjoint in particular is very weakly correlated with the remaining singular vectors.

The comparison of adjoint variables with the singular vectors (Figures 4–8) points to the following conclusions:

- The adjoints have a similar structure with the first singular vectors. The next dominant singular vectors (the second, the third, etc.) carry additional information about the areas where changes have a high impact on the optimization area. This additional
information is not captured by the adjoint.

- The adjoint covers a wider area following the flow pattern, while the singular vectors remain localized, even for larger simulation interval.

These findings are in good agreement with the results reported for meteorological models by Gelaro et al. (1998).

6.5 Influence of the Meteorological Conditions

To assess the influence of different meteorological conditions on the singular vectors we perform a 24 hours simulation starting at 0 GMT March 26, 2001, for the same optimization criterion (ground level $O_3$ in the Korea area). The 2 Km level wind fields on March 26 are shown in Figure 13. A comparison with the conditions present on March 1st (shown in Figure 3) reveals that the meteorological conditions were considerably different during these two days.

The $O_3$, $NO_2$ and $HCHO$ sections of the four dominant singular vectors are shown in Figure 14. There are clear differences between the structure of the singular vectors at March 26 and at March 1st (Figures 4–8). This shows the important role that meteorology plays in the distribution of pollutant concentrations.

6.6 Influence of the Optimization Region

In the study of chemical transport systems the dominant singular vectors are useful to describe the uncertainty in a limited subdomain. We now analyze how the choice of the optimization region impacts the singular vectors.
The $O_3$, $NO_2$ and $HCHO$ sections of the dominant eigenvector for another 24h, March 1 simulation are shown in Figure 15. The optimization criterion is ground level $O_3$ over a region of the same area, but located in South-East China (the gray area on the map). As expected, singular vectors are localized over the optimization area.

Another numerical test is performed for a optimization area that covers 24 grid cells over Japan, Korea and South-East China. The magnitude of the largest eigenvalues decreases at a slower rate, as shown in Figure 18. This is due to the optimization area being larger than in the previous numerical experiments. About 30 eigenvalues are needed for a 2 orders of magnitude decrease in the magnitude of the eigenvalues; therefore about 30 singular vectors are needed to accurately capture the uncertainty. The $O_3$, $NO_2$ and $HCHO$ sections of the dominant singular vectors are shown in Figure 16. Singular vectors are localized over China and Korea; there are no lobes localized over Japan. Due to the Westerly flows changes in initial concentration fields over Japan do not impact significantly ground $O_3$ concentration after 24h over the optimization region.

To further show the influence of the optimization region we consider a very large area with over 100 cells covering parts of China, Korea, and Japan (the gray area on the map in Figure 17). The magnitude of the largest 40 eigenvalues is shown in Figure 18. The decrease of eigenvalue magnitude is slower for the larger regions, and therefore, more eigenvectors are needed to capture the uncertainty. The ratio of the smallest to largest computed eigenvalues is $\lambda_1/\lambda_{40} = 22$ for the 100 cells region, compared to $\lambda_1/\lambda_{40} = 472$ for the 24 cells region. The $O_3$, $NO_2$ and $HCHO$ sections of the dominant singular vectors are shown in Figure 17. The dominant singular vectors are also localized to Korea and South-East China. Note that the first few dominant singular vectors are insufficient to accurately describe the uncertainty.
The singular vectors also depend on the choice of error norms at the initial \((A)\) and
final time \((B)\). Additional numerical tests (not shown here) have been performed using the
ground level concentrations of 66 long-lived species over Korea. The new singular vectors
displayed clear differences from the ones computed based on ozone ground level over Korea.

7 Conclusions

In this work we study the computational aspects of singular vector analysis of chemical
transport models. Singular vectors span the directions of maximal error growth in a finite
time, as measured by specific error norms.

To maintain the symmetry of the tangent linear-adjoint operator \((M^*M)\) it is necessary
to employ discrete adjoints. A projection method is proposed to preserve the symmetry of
\(M^*M\) operators for stiff chemical systems. The application of this technique is extended
to 3D chemical transport models. Different definitions of the perturbation error norms are
discussed.

Numerical results are presented for a 3D chemical transport simulation of atmospheric
currents in East Asia in March 2001. The assumption of linear propagation of perturbations,
intrinsic in the singular vector calculation, was checked numerically for a 24 hours simulation
interval. The singular values and the structure of the singular vectors depend on the length
of the simulation interval, the meteorological data, the location of optimization region and
the selection of optimized species, the choice of error norms, and the size of the optimization
region.

While in the study of atmospheric dynamics the dominant singular vectors are associated
with unstable modes, in the study of chemical transport systems the dominant singular vectors are useful to describe the uncertainty in a limited subdomain (e.g., where the model prediction needs to be improved). In this paper we illustrate the use of singular vectors to describe uncertainties in the initial conditions. Other very important sources of uncertainty in air quality models that need to be quantified in real applications are the emissions, the meteorological fields, the deposition velocities, and the top and lateral boundary conditions for regional models.

The predictions of air quality models are corrupted by uncertainties coming from the initial distribution of the chemical fields, the boundary conditions, the rates of emission of pollutants, and the meteorological fields. In this paper we illustrate the use of singular vectors to quantify the propagation of uncertainties from the initial conditions; but the other sources need to be accounted for in a real data assimilation setting. The decrease of the singular values for longer simulation intervals is due to the fact that, as time progresses, the final solution is driven more by emissions and less by the initial conditions. Consequently, the effect of uncertainties in emission sources on the final state becomes more important.

Most of the uncertainty in the optimization 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. For limited optimization regions the singular values decrease rapidly, and a few dominant singular vectors are sufficient to capture most of the uncertainty. For large optimization regions the singular values decrease slowly. The areas of influence are no longer localized, and uncertainty from all over the computational domain contribute to the uncertainty in the optimization area at the final time. As a consequence for data assimilation, small ensembles
are sufficient if the observations are localized, or if one seeks improved predictions over a relatively small, well defined region.

In order to improve predictions within the optimization region additional observations are needed in the areas described by the dominant singular vectors. Additional $O_3$ measurements have to be placed in a different location than additional $NO_2$ or $HCHO$ observations. The optimal location of observations changes in time and drifts away from the optimization area for longer intervals.

The dominant singular vector has a similar structure to the adjoint variable. The next singular vectors carry additional information about the high sensitivity areas, which is not captured by a simple adjoint analysis.

The computation of singular vectors is computationally intensive. In our experiments 40 to 100 iterations were necessary for PARPACK to converge (taking between 8–16 hours cpu time for a parallel run on 30 opteron processors). Each iteration includes one forward and tangent linear model run and one adjoint run. The cost of one forward and tangent linear model (using a direct-decoupled approach and reusing the matrix factorizations) followed by one backward adjoint integration is less than three times the cost of the forward trajectory calculation (Sandu et al., 2005). The calculation of singular vectors is at least as expensive as a full 4D-Var data assimilation cycle, where 20–30 iterations are typically sufficient for a substantial decrease in the cost function (Sandu et al., 2005). Note that the cost of performing the chemical projections accounts for only a small percent of the total computational time.

As the field of chemical weather forecasting grows, it can be anticipated that singular vectors will find many applications. The results presented in this paper are a first step in
this direction.

Acknowledgements

This work was supported by the National Science Foundation through the award NSF ITR AP&IM 0205198. A. Sandu’s work was also partially supported by the award NSF CAREER ACI-0413872. We would like to thank Virginia Tech’s laboratory for Advanced Scientific Computing (LASCA) for the use of the Anantham cluster. The authors would also like to thank the anonymous reviewers for their constructive comments which helped improve this work.
References


Lehoucq, R., Maschhoff, K., Sorensen, D., and Yang, C. Arpack software.


Leutbecher, M. 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.


Maschhoff, K. and Sorensen, D. Parallel arpack home page.


List of Figure Captions

Figure 1: (a) The variation of the symmetry residual with time shows a loss of symmetry during the initial transient of the stiff ODE. (b) The symmetry residual decreases rapidly with the number of projection iterations.

Figure 2: The dominant eigenvalues ($\lambda_k = \sigma_k^2$) for the 12h, 24h and 48h simulations. The rapid decrease in magnitude indicates that uncertainty in the optimization region can be captured by only a few singular vectors.

Figure 3: The 2 Km level wind fields during the 12h, 24h and 48h simulations starting at 0 GMT March 1, 2001. The optimization area is shaded.

Figure 4: Dominant singular vectors (SV) ($O_3$ sections) for the 12h, 24h and 48h simulations starting at 0 GMT March 1, 2001. The optimized criterion is ground level $O_3$ value in the shaded area.

Figure 5: Second dominant singular vectors (SV) ($O_3$ sections) for the 12h, 24h, and 48h simulations starting at 0 GMT March 1, 2001.

Figure 6: Third and fourth dominant singular vectors (SV) ($O_3$ sections) for the 12h, 24h, and 48h simulations starting at 0 GMT March 1, 2001.

Figure 7: First and second dominant singular vectors (SV) ($NO_2$ sections) for the 12h, 24h, and 48h simulations starting at 0 GMT March 1, 2001.

Figure 8: First and second dominant singular vectors (SV) ($HCHO$ sections) for the 12h, 24h, and 48h simulations starting at 0 GMT March 1, 2001.
Figure 9: The evolved dominant singular vectors (SV) after 24 hours of simulation (i.e., at 0 GMT, March 2, 2001).

Figure 11: Adjoint for the 24h simulation starting at 0 GMT March 1, 2001.

Figure 12: Correlations between homologous sections of the adjoint variable and of the dominant singular vectors.

Figure 13: The 2 Km level wind fields during the simulation starting at 0 GMT March 26, 2001. The optimization area is shaded.

Figure 14: Dominant singular vectors (SV) for the 24h simulation starting at 0 GMT, March 26, 2001.

Figure 15: Dominant singular vectors (SV) for the 24h simulation starting at 0 GMT March 1, 2001. The optimized criterion is ground level $O_3$ over the gray region in South-East China.

Figure 16: Dominant singular vectors (SV) for the 24h simulation. The optimization area is the 24-cell area covering Korea, Japan, and East China.

Figure 17: Dominant singular vectors (SV) for the 24h simulation over a large optimization area.

Figure 18: The dominant eigenvalues for simulations on 24-cell and large optimization areas.
List of Table Captions

Table 1. Singular values are well approximated by the ratios of perturbation energies for initial perturbations along the dominant singular vectors and also for random perturbations.
Figure 1: (a) The variation of the symmetry residual with time shows a loss of symmetry during the initial transient of the stiff ODE. (b) The symmetry residual decreases rapidly with the number of projection iterations.
Figure 2: The dominant eigenvalues ($\lambda_k = \sigma_k^2$) for the 12h, 24h and 48h simulations. The rapid decrease in magnitude indicates that uncertainty in the optimization region can be captured by only a few singular vectors.
Figure 3: The 2 Km level wind fields during the 12h, 24h and 48h simulations starting at 0 GMT March 1, 2001. The optimization area is shaded.
Figure 4: Dominant singular vectors (SV) ($O_3$ sections) for the 12h, 24h and 48h simulations starting at 0 GMT March 1, 2001. The optimized criterion is ground level $O_3$ value in the shaded area.
Figure 5: Second dominant singular vectors (SV) ($O_3$ sections) for the 12h, 24h, and 48h simulations starting at 0 GMT March 1, 2001.
Figure 6: Third and fourth dominant singular vectors (SV) ($O_3$ sections) for the 12h, 24h, and 48h simulations starting at 0 GMT March 1, 2001.
Figure 7: First and second dominant singular vectors (SV) ($NO_2$ sections) for the 12h, 24h, and 48h simulations starting at 0 GMT March 1, 2001.
Figure 8: First and second dominant singular vectors (SV) (HCHO sections) for the 12h, 24h, and 48h simulations starting at 0 GMT March 1, 2001.
Figure 9: The (linearly) evolved dominant singular vectors (SV) after 24 hours of simulation (i.e., at 0 GMT, March 2, 2001).
Figure 10: The nonlinearly evolved perturbations after 24 hours of simulation (i.e., at 0 gmt, March 2, 2001).
Figure 11: Adjoints for the 24h simulation starting at 0 GMT March 1, 2001.
Figure 12: Correlations between homologous sections of the adjoint variable and of the dominant singular vectors.
Figure 13: The 2 Km level wind fields during the simulation starting at 0 GMT March 26, 2001. The optimization area is shaded.
Figure 14: Dominant singular vectors (SV) for the 24h simulation starting at 0 GMT, March 26, 2001.
Figure 15: Dominant singular vectors (SV) for the 24h simulation starting at 0 GMT March 1, 2001. The optimized criterion is ground level $O_3$ over the gray region in South-East China.
Figure 16: Dominant singular vectors (SV) for the 24h simulation. The optimization area is the 24-cell area covering Korea, Japan, and East China.
Figure 17: Dominant singular vectors (SV) for the 24h simulation over a large optimization area.
Figure 18: The dominant eigenvalues for simulations on 24-cell and large optimization areas.
<table>
<thead>
<tr>
<th>Initial Perturbation</th>
<th>Singular Values</th>
<th>Ratio of Perturbation Energies</th>
</tr>
</thead>
<tbody>
<tr>
<td>along $s_1(t_0)$</td>
<td>$\sigma_1^2 = 9.494 \times 10^{-2}$</td>
<td>$9.685 \times 10^{-2}$</td>
</tr>
<tr>
<td>along $s_2(t_0)$</td>
<td>$\sigma_2^2 = 5.146 \times 10^{-2}$</td>
<td>$5.169 \times 10^{-2}$</td>
</tr>
<tr>
<td>along $s_3(t_0)$</td>
<td>$\sigma_3^2 = 3.982 \times 10^{-2}$</td>
<td>$3.944 \times 10^{-2}$</td>
</tr>
<tr>
<td>along $s_4(t_0)$</td>
<td>$\sigma_4^2 = 2.499 \times 10^{-2}$</td>
<td>$2.573 \times 10^{-2}$</td>
</tr>
<tr>
<td>along $s_5(t_0)$</td>
<td>$\sigma_5^2 = 1.756 \times 10^{-2}$</td>
<td>$1.741 \times 10^{-2}$</td>
</tr>
<tr>
<td>random $r$</td>
<td>$\sum_{i=1}^{12} \sigma_i^2 \langle r(t_0), A s_i(t_0) \rangle^2 = 3.174 \times 10^{-22}$</td>
<td>$3.210 \times 10^{-22}$</td>
</tr>
</tbody>
</table>

Table 1: Singular values are well approximated by the ratios of perturbation energies for initial perturbations along the dominant singular vectors and also for random perturbations.