Total Energy 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
E-mail: {liao, 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
E-mail: {gcarmich,tchai}@cgrer.uiowa.edu

February 17, 2005

*Corresponding author.

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

Abstract

Total energy singular vectors (TESVs) 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 the TESV 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 target region and the selection of target species, the choice of energy norms, and the size of the target region. The calculation of singular vectors is as expensive as a full 4D-Var data assimilation cycle.

1 Introduction

Total energy singular vectors (TESVs) are the directions of fastest error growth over a finite time interval. Singular vector analysis was introduced in meteorology by Lorenz [23], who computed the largest error growth rates in an idealized model of the atmosphere. The adjoint technique was used by Molteni et al. [30] and Mureau et al. [31] to compute singular vectors for meteorological models. Singular vector analysis of large-scale general circulation models is now possible (see e.g., Navon et al. [22]).

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 [2, 13, 30, 32]; (2) initialization of ensemble forecasts [29, 31]; and (3) estimation of the optimal placement of adaptive observations [4, 15, 24, 33].

Numerous studies have shown that the structure of singular vectors in atmospheric general circulation models is determined by: (1) the atmospheric episode under consideration [1]; (2) model physics (e.g., the treatment of boundary layer processes) [1, 4, 5, 12, 25]; (3) model resolution [4, 5, 12, 25]; and (4) the particular choice of energy norms [20].

The objective of this work is to study the total energy 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.

The paper is organized as follows. In Section 2 we introduce the total energy singular vectors and introduce 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 energy 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 $x(t_{\mathrm{F}})$. With $\mathcal{M}$ denoting the model solution operator we have

$$x(t_{\mathrm{F}}) = \mathcal{M}_{t_0 \to t_{\mathrm{F}}} \left(x(t_0) \right) \ . \tag{1}$$

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

$$\delta x(t_{\rm F}) = M_{t_0 \to t_{\rm F}}\ \delta x(t_0)\ , \tag{2}$$

and adjoint variables according to the adjoint model

$$\lambda(t_0) = M^*_{t_{\rm F} \to t_0} \lambda(t_{\rm F})\ . \tag{3}$$

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 $P(t_{\rm F})$ according to

$$P(t_{\rm F}) = M_{t_0 \to t_{\rm F}}\ P(t_0)\ M^*_{t_{\rm F} \to t_0} + Q\ . \tag{4}$$

The additional term Q represents the covariance of the model errors.

2.1 Total Energy Singular Vectors

Total energy singular vectors (TESVs) 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 ("energy") norm defined by a symmetric positive definite matrix A

$$\| \delta x(t_0) \|^2_A = \langle\ \delta x(t_0)\ ,\ A\ \delta x(t_0)\ \rangle\ . \tag{5}$$

Similarly, the perturbation magnitude at the final time $t_{\rm F}$ is measured in a semi-norm defined by a semi-positive definite matrix B

$$\| \delta x(t_{\rm F}) \|^2_B = \langle\ \delta x(t_{\rm F})\ ,\ B\ \delta x(t_{\rm F})\ \rangle\ . \tag{6}$$

We call the norms (5) and (6) squared the "perturbation energies". The ratio between perturbation energies at $t_{\rm F}$ and t_0 offers a measure of error growth:

$$\begin{aligned} \sigma^2 &= \frac{\|\delta x(t_{\rm F})\|^2_B}{\|\delta x(t_0)\|^2_A} = \frac{\langle \delta x(t_{\rm F}), B\delta x(t_{\rm F})\rangle}{\langle \delta x(t_0), A\delta x(t_0)\rangle} \\ &= \frac{\langle M_{t_0 \to t_{\rm F}}\ \delta x(t_0)\ ,\ B\ M_{t_0 \to t_{\rm F}}\ \delta x(t_0)\rangle}{\langle \delta x(t_0)\ ,\ A\ \delta x(t_0)\rangle} \\ &= \frac{\langle \delta x(t_0), M^*_{t_{\rm F} \to t_0} B M_{t_0 \to t_{\rm F}} \delta x(t_0)\rangle}{\langle \delta x(t_0), A\delta x(t_0)\rangle} \end{aligned} \tag{7}$$

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

TESVs 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 (7). These directions are the solutions of the generalized eigenvalue problem

$$M^*_{t_{\rm F} \to t_0}\ B\ M_{t_0 \to t_{\rm F}}\ s_k(t_0) = \sigma_k^2\ A\ s_k(t_0)\ . \tag{8}$$

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 [21], or its parallel version PARPACK [27].

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

$$A^{-\frac{1}{2}} M^*_{t_{\rm F}\to t_0} B M_{t_0\to t_{\rm F}} A^{-\frac{1}{2}} v_k(t_0) = \sigma_k^2 v_k(t_0) \ , \quad v_k(t_0) = A^{\frac{1}{2}} s_k(t_0) \ . \tag{9}$$

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

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

The TESVs s_k are A-orthogonal at t_0 and B-orthogonal at $t_{\rm F}$

$$\big\langle s_k(t_0), A s_j(t_0) \big\rangle = 0 \qquad \text{and} \qquad \big\langle s_k(t_{\rm F}), B s_j(t_{\rm F}) \big\rangle = 0 \qquad \text{for} \quad j \neq k \ . \tag{11}$$

The equations (10) and (11) justify the name of "total energy singular vectors". The singular value decomposition of the linear operator $M_{t_0\to t_{\rm F}}$, with the A scalar product at t_0 and the B scalar product at $t_{\rm F}$, has the left singular vectors $s_k(t_0)$ and the right singular vectors $s_k(t_{\rm F})$. The singular values σ_k are the error amplification factors along each direction s_k.

A special set of energy 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 $s_k(t_{\rm F})$ of the forecast error covariance matrix $P(t_{\rm F})$,

$$P(t_{\rm F})\, s_k(t_{\rm F}) = \sigma_k^2 \, s_k(t_{\rm F}) \ . \tag{12}$$

The eigenvectors $s_k(t_{\rm F})$ are called "Hessian singular vectors" (HSVs). Since the leading eigenvectors of $P(t_{\rm F})$ are the directions of maximum variance of forecast error, HSVs define the directions along which we must do a good job of analysis in order to minimize the forecast error at $t_{\rm F}$. We assume that the model error in equation (14) is negligible over the period $[t_0, t_{\rm F}]$. From equation (12) it follows that Hessian singular vectors are the solutions of the following generalized eigenvalue problem

$$M^*_{t_{\rm F}\to t_0} M_{t_0\to t_{\rm F}} s_k(t_0) = \lambda_k P(t_0)^{-1} s_k(t_0) = \lambda_k \left(\nabla^2 J\right) s_k(t_0) \tag{13}$$

The second relation comes from the fact that the inverse of the analysis covariance matrix is equal to the Hessian matrix of the analysis cost function J in the variational analysis system. This motivates the name *Hessian singular vectors* for the solutions $s_k(t_0)$ of the eigenproblem (13).

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 [10, 19] provide a stochastic approach to the data assimilation problem. Consider a set of observations y at $t_{\rm F}$ (assumed, for simplicity, to be a linear function of model state, $y = Hx$). The extended Kalman filter uses the forecast state and its covariance $(x(t_{\rm F}), P(t_{\rm F}))$ and the observations and their covariance (y, R) to produce an optimal ("analyzed") estimation of the model state and its covariance $(x_{\rm A}(t_{\rm F}), P_{\rm A}(t_{\rm F}))$:

$$\begin{aligned} x_{\rm A}(t_{\rm F}) &= x(t_{\rm F}) + P(t_{\rm F})\, H^T \left(R + H\, P(t_{\rm F})\, H^T\right)^{-1} \left(y - Hx(t_{\rm F})\right) \\ P_{\rm A}(t_{\rm F}) &= P(t_{\rm F}) - P(t_{\rm F})\, H^T \left(R + H\, P(t_{\rm F})\, H^T\right)^{-1} H\, P(t_{\rm F}) \end{aligned} \tag{14}$$

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 [14]. The commonly used method to reduce the computational cost is to propagate (only) the projection of the covariance matrix onto a low-dimensional subspace ($\text{span}\{s_1, \cdots, s_k\}$). The ensemble Kalman filter [18] 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 $t_{\rm F}$) should contain the directions $s_k(t_{\rm F})$ along which the error has the maximal growth. Consequently the initial ensemble should be defined based on the total energy (or Hessian) 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 [3, 5, 16].

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.

Buizza and Montani [4] showed that TESVs can identify the most sensitive regions of the atmosphere for targeted observations. They are useful as long as the linearity assumption of error propagation holds [17]. Majudmar et al.[26] compare the TESV approach for observation targeting to the ensemble transform Kalman filter. Palmer et al. [33] argue that for

predictability studies an appropriate metric is the perturbation energy. Dăescu and Navon [9] discuss the adaptive observation problem in the context of 4D-Var data assimilation.

3 Computation of Chemical Singular Vectors

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

A numerical eigenvalue solver applied to equation (8) requires a symmetric matrix $M^* B M$ 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 [34]. 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,

$$u(\tau) = M_{t_0 \to \tau} u(t_0) \qquad \text{and} \qquad v(\tau) = M_{t_0 \to \tau} v(t_0) \ . \tag{15}$$

The symmetry residual is the difference

$$r(\tau) = \left\langle u(\tau), M^*_{t_{\rm F} \to \tau} M_{\tau \to t_{\rm F}} v(\tau) \right\rangle - \left\langle v(\tau), M^*_{t_{\rm F} \to \tau} M_{\tau \to t_{\rm F}} u(\tau) \right\rangle \tag{16}$$

If M^* is the exact discrete adjoint of M then $r(\tau) = 0$ for all τ. 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 [7] which has 93 species and 235 reactions. The forward, tangent linear, and adjoint models are implemented using the automatic code generator KPP [8, 11, 34]. 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 [34] 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 consider a singular perturbation model for the chemical system

$$y' = f(y,z) \ , \quad \epsilon \, z' = g(y,z) \ , \quad \epsilon \ll 1 \ . \tag{17}$$

Here y is the slow and z the fast component. Perturbations propagate through the tangent linear model

$$\delta y' = f_y(y,z)\,\delta y + f_z(y,z)\,\delta z\ , \quad \epsilon\,\delta z' = g_y(y,z)\,\delta y + g_z(y,z)\,\delta z\ . \tag{18}$$

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

$$\delta z = -g_z^{-1}(y,z)\,g_y(y,z)\,\delta y\ . \tag{19}$$

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 $M^*BM\,x$. 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

$$w = \Pi^* M^*_{t_{\rm F}\to t_0}\,B\,M_{t_0\to t_{\rm F}}\Pi\,x\ , \tag{20}$$

where Π and Π^* denote the projection operations performed with the tangent linear and the adjoint models respectively.

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.

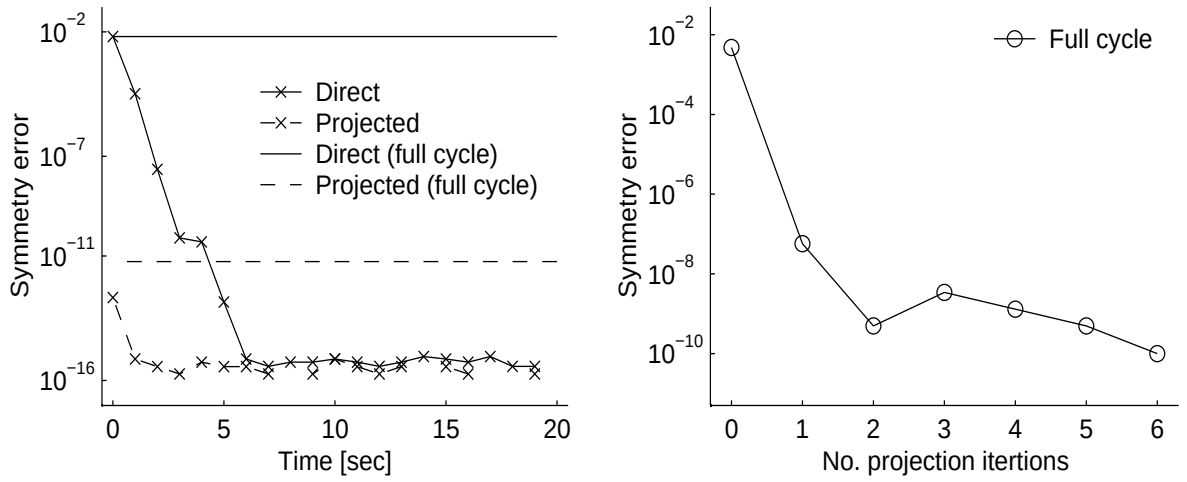
Fig.1 (a) (dashed) presents the evolution of the symmetry residual when 6 projection steps are performed with the very small stepsize of 10^{-9} seconds. The symmetry error during the stiff transient is reduced to 10^{-11}.

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 [35].

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.



(a) (b)

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.

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

$$\begin{aligned}
\frac{\partial c_i}{\partial t} &= -u\cdot\nabla c_i + \frac{1}{\rho}\nabla\cdot(\rho K\nabla c_i) + \frac{1}{\rho}f_i(\rho c) + E_i \ , \quad t_0 \le t \le t_{\rm F} \ , \\
c_i(t_0,x) &= c_i^0(x) \ , \\
c_i(t,x) &= c_i^{\rm in}(t,x) \quad \text{for} \quad x\in\Gamma^{\rm in} \ , \quad K\frac{\partial c_i}{\partial n} = 0 \quad \text{for} \quad x\in\Gamma^{\rm out} \ , \\
K\frac{\partial c_i}{\partial n} &= V_i^{\rm dep}c_i - Q_i \quad \text{for} \quad x\in\Gamma^{\rm ground} \ , \quad \text{for all} \ \ 1\le i\le N_{\rm spec} \ .
\end{aligned} \tag{21}$$

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

A perturbation δc^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:

$$\begin{aligned}
\frac{\partial \delta c_i}{\partial t} &= -u\cdot\nabla \delta c_i + \frac{1}{\rho}\nabla\cdot(\rho K\nabla \delta c_i) + F_{i,*}(\rho c)\delta c \ , \quad t_0 \le t \le t_{\rm F} \\
\delta c_i(t_0,x) &= \delta c_i^0(x) \ , \\
\delta c_i(t,x) &= 0 \quad \text{for} \quad x\in\Gamma^{\rm in} \ , \quad K\frac{\partial \delta c_i}{\partial n} = 0 \quad \text{for} \quad x\in\Gamma^{\rm out} \ , \\
K\frac{\partial \delta c_i}{\partial n} &= V_i^{\rm dep}\delta c_i \quad \text{for} \quad x\in\Gamma^{\rm ground} \ , \quad \text{for all} \ \ 1\le i\le N_{\rm spec} \ .
\end{aligned} \tag{22}$$

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 λ_i:

$$\begin{aligned}
\frac{\partial \lambda_i}{\partial t} &= -\nabla\cdot(u\lambda_i) - \nabla\cdot\left(\rho K \nabla \frac{\lambda_i}{\rho}\right) - \left(F^T(\rho c)\lambda\right)_i - \phi_i \ , \quad t_{\rm F} \geq t \geq t_0 \\
\lambda_i(t_{\rm F}, x) &= \lambda_i^{\rm F}(x) \ , \\
\lambda_i(t, x) &= 0 \quad \text{for} \quad x \in \Gamma^{\rm in} \ , \quad \lambda_i u + \rho K \frac{\partial(\lambda_i/\rho)}{\partial n} = 0 \quad \text{for} \quad x \in \Gamma^{\rm out} \ , \\
\rho K \frac{\partial(\lambda_i/\rho)}{\partial n} &= V_i^{\rm dep}\lambda_i \quad \text{for} \quad x \in \Gamma^{\rm ground} \ , \quad \text{for all} \ \ 1 \leq i \leq N_{\rm spec} \ .
\end{aligned} \tag{23}$$

The ground boundary condition is obtained from the fact that $\vec{u}\cdot\vec{n} = 0$ at ground level. The forcing function ϕ_i depends on the particular cost functional under consideration [35].

The numerical solution operator for (21) is based on an operator splitting approach, where the transport and the chemistry steps are taken successively. With $\mathcal{T}$ the numerical solution operator for transport, and $\mathcal{C}$ the solution operator for chemistry, one step of the solution operator reads

$$c^{m+1} = \mathcal{T}_{\Delta t/2}\cdot\mathcal{C}_{\Delta t}\cdot\mathcal{T}_{\Delta t/2}\cdot c^m \ . \tag{24}$$

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 (Π) is applied before each linearized chemistry,

$$\delta c^{m+1} = T_{\Delta t/2}\cdot C_{\Delta t}\cdot\Pi\cdot T_{\Delta t/2}\cdot\delta c^m \ . \tag{25}$$

The *discrete adjoint model* is based on the discrete adjoints of the transport (T^*) and chemistry (C^*) numerical schemes. A chemical adjoint projection (Π^*) is applied after each adjoint chemistry step,

$$\lambda^m = T^*_{\Delta t/2}\cdot\Pi^*\cdot C^*_{\Delta t}\cdot T^*_{\Delta t/2}\cdot\lambda^{m+1} + \phi^m \ . \tag{26}$$

5 Energy 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 will provide a reasonable measure of the "energy of the perturbation"

$$\| \delta c \|^2 = \sum_{i,j,k}\sum_{s=1}^{N_{\rm spec}} \left(\delta c^s_{i,j,k}\right)^2 = \left\langle \delta c \, , \ B\,\delta c \right\rangle \quad \text{with} \quad B = I \ .$$

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 target area, and a given set of chemical species. In this case the perturbation energy measure at the final time is defined in terms of a matrix with the diagonal entries equal to one for the selected target variables and equal to zero elsewhere,

$$\| \delta c(t_{\rm F}) \|^2 = \sum_{\substack{\text{target}\\ \text{area}}} \sum_{\substack{\text{target}\\ \text{spec.}}} \left(\delta c^s_{i,j,k}\right)^2 = \left\langle \delta c(t_{\rm F}) \,,\; B\, \delta c(t_{\rm F}) \right\rangle \quad \text{with} \quad B = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} . \tag{27}$$

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 energy 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

$$\sigma^2 = \frac{\left\langle \frac{\delta c(t_{\rm F})}{c(t_{\rm F})},\; B\,\frac{\delta c(t_{\rm F})}{c(t_{\rm F})} \right\rangle}{\left\langle \frac{\delta c(t_0)}{c(t_0)},\; A\frac{\delta c(t_0)}{c(t_0)} \right\rangle} . \tag{28}$$

This can be accomplished by using the logarithms of the concentrations $\log c^s_{i,j,k}$ as model variables. In practice it is advantageous to approximate the relative errors by the absolute errors δc^s_{ijk} scaled by "typical" concentration values w^s_{ijk}

$$\frac{\delta c^s_{i,j,k}(t_{\rm F})}{c^s_{i,j,k}(t_{\rm F})} \approx \frac{\delta c^s_{i,j,k}(t_{\rm F})}{w^s_{i,j,k}(t_{\rm F})} \quad \Rightarrow \quad \frac{\delta c(t_{\rm F})}{c(t_{\rm F})} \approx W^{-1}_{t_{\rm F}} \delta c(t_{\rm F}) \quad \text{where} \quad W_t = \mathrm{diag}_{i,j,k,s}\left\{ w^s_{i,j,k}(t) \right\} .$$

and maximize

$$\sigma^2 = \frac{\left\langle W^{-1}_{t_{\rm F}} \delta c(t_{\rm F}),\; B\, W^{-1}_{t_{\rm F}} \delta c(t_{\rm F}) \right\rangle}{\left\langle W^{-1}_{t_0} \delta c(t_0),\; A\, W^{-1}_{t_0} \delta c(t_0) \right\rangle} . \tag{29}$$

One reason for this approximation is that the "typical" concentrations $w^s_{i,j,k}$ 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

$$M^*_{t_{\rm F}\to t_0}\, W^{-1}_{t_{\rm F}}\, B\, W^{-1}_{t_{\rm F}}\, M_{t_0\to t_{\rm F}} s_k(t_0) = \sigma_k^2\, A\, W^{-2}_{t_0} s_k(t_0) \tag{30}$$

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

$$W_{t_0} M^*_{t_{\rm F}\to t_0}\, W^{-1}_{t_{\rm F}}\, B\, W^{-1}_{t_{\rm F}}\, M_{t_0\to t_{\rm F}} W_{t_0}\, v_k = \sigma_k^2\, v_k \;, \quad v_k = W^{-1}_{t_0} s_k(t_0) \;. \tag{31}$$

6 Numerical Results

The numerical tests use the state-of-the-art regional atmospheric chemical transport model STEM [6]. 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 [6].

The computational grid has $N_x \times N_y \times N_z$ nodes with N_x=30, N_y=20, N_x=18, and a horizontal resolution of 240 Km × 240 Km. The chemical mechanism is SAPRC-99 [7] 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 [35]. Both the forward and adjoint chemical models are implemented using KPP [8, 11, 34]. The forward and adjoint models are parallelized using PAQMSG [28]. PARPACK [21] was used to solve the symmetric generalized eigenvalue problems.

The singular vectors $s(N_x,N_y,N_z,N_{spec})$ in equation (8) are four-dimensional. 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 target regions, target species, simulation intervals, meteorological data, and energy norms are presented next.

6.1 TESVs for Different Simulation Intervals

We first consider the case where the target is the ground level ozone concentration in a 720 Km × 960 Km area covering Korea. The TESV 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

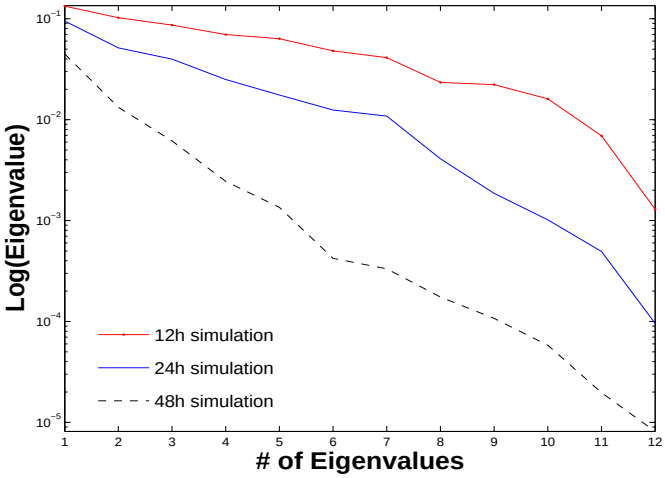


Figure 2: The dominant eigenvalues ($\lambda_k = \sigma_k^2$) for the 12h, 24h and 48h simulations. The rapid decrease in magnitude indicates that uncertainty in the target region can be captured by only a few TESVs.

that the uncertainty in the ground level ozone within the target region can be captured with only a few TESVs. The eigenvalues decrease faster for longer simulation intervals.

The top and South views for O_3 sections of the dominant TESVs for the 12h, 24h and 48h simulations are presented in Figure 3. TESVs are localized near the target area in both the horizontal and the vertical directions. As expected, for longer simulation intervals the singular vectors spread further away from the target region.

The first four dominant TESVs for the 12h, 24h, and 48h simulations starting at 0 GMT March 1, 2001 are shown in Figures 3, 4 and 5 (O_3 sections), Figures 6 and 7 (NO_2 sections), and Figures 8 and 9 ($HCHO$ sections). A close look at the structure of the dominant TESVs reveals that:

- For the same simulation the shapes of the four dominant TESVs are different. This illustrates the fact that different TESVs 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 target, illustrating that perturbations in a wider area at earlier times impact the target.

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

Initial Perturbation	Singular Values	Ratio of Perturbation Energies $(\|\Delta x(t_F)\|_B^2 / \|\Delta x(t_0)\|_A^2)$
along $s_1(t_0)$	$\sigma_1^2 = 9.494 \times 10^{-2}$	9.685×10^{-2}
along $s_2(t_0)$	$\sigma_2^2 = 5.146 \times 10^{-2}$	5.169×10^{-2}
along $s_3(t_0)$	$\sigma_3^2 = 3.982 \times 10^{-2}$	3.944×10^{-2}
along $s_4(t_0)$	$\sigma_4^2 = 2.499 \times 10^{-2}$	2.573×10^{-2}
along $s_5(t_0)$	$\sigma_5^2 = 1.756 \times 10^{-2}$	1.741×10^{-2}

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

In order to improve predictions of ground level Korean ozone 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 target area for longer intervals.

6.2 The Linearity Assumption

Inherent in the TESV 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 energy ratios approximate well the singular values. This confirms that the linearity assumption holds (at least) for the 24 hour simulation interval under consideration.

6.3 Evolved TESVs

The perturbations initialized along each dominant TESV develop in time becoming "evolved TESVs". We are interested in the shape of these perturbations at the end of the 24 hours simulation interval. The evolved TESVs (scaled to have the A-norm equal to one) are displayed in Figure 10. The largest values of TESVs are clustered above the target region. This is expected since the perturbations along dominant TESVs impact mostly the target

area at the end of the 24 hours interval.

6.4 TESVs versus Adjoints

To illustrate the difference between the information conveyed by the TESVs and by the adjoint variables we consider again the ground level O_3 in the Korea target 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 target 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 TESVs $s_k(t_0)$ we consider the correlation coefficients $\rho_k = \langle \lambda(t_0), A s_k(t_0) \rangle / (\|\lambda(t_0)\|_A \, \|s_k(t_0)\|_A)$. 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 TESV is the strongest. The O_3 section of the adjoint in particular is very weakly correlated with the remaining TESVs.

The comparison of adjoint variables with the TESVs (Figures 3–9) points to the following conclusions:

- The adjoints have a similar structure with the first TESVs. The next dominant TESVs(the second, the third, etc.) carry additional information about the areas where changes have a high impact on the target. 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.

6.5 Influence of the Meteorological Conditions

To assess the influence of different meteorological conditions on the TESVs we perform a 24h simulation starting at 0 GMT March 26, 2001, for the same target (ground level O_3 in the Korea area). The meteorological conditions on March 26 were considerably different than those on March 1, 2001.

The O_3, NO_2 and $HCHO$ sections of the four dominant TESVs are shown in Figure 13. The structure of the TESVs is different than for March 1 (Figures 3–9).

6.6 Influence of the Energy Norm

We expect that the total energy singular vectors (TESV), i.e. the directions of maximal error growth, will depend on the choice of error energy norms at the initial (A) and final

time (B). All previous numerical tests used an energy norm at the final time based on the ground level O_3 concentration in the Korea area. We now consider the energy norm defined over the same area by the ground level concentrations of 66 long-lived species. Since the concentrations of different species differ by orders of magnitude, we scale each concentration by a "typical" value for that species. This can be interpreted as a definition of the total energy of *relative errors* at the final time. The "typical" concentration values are the mean values of each species at ground level throughout the area of interest.

The first four dominant TESVs are shown in Figure 14. There are clear differences between these TESVs and the ones computed previously.

6.7 Influence of the Target Region

The O_3, NO_2 and $HCHO$ sections of the dominant eigenvector for another 24h, March 1 simulation are shown in Figure 15. The target 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, TESVs are localized over the target area.

Another numerical test is performed for a target 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 target 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 TESVs 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. TESVs are localized over China and Korea; there are no lobes localized over Japan. Due to the Eastward pattern of the flow, changes in initial concentration fields over Japan do not impact significantly ground O_3 concentration after 24h over the target region.

To further show the influence of the target 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 O_3, NO_2 and $HCHO$ sections of the dominant TESVs are shown in Figure 17. The dominant TESVs are also localized to Korea and South-East China. Note that the first few dominant TESVs are insufficient to accurately describe the uncertainty.

7 Conclusions

In this work we study the computational aspects of total energy singular vector analysis of chemical transport models. Singular vectors span the directions of maximal error growth in a finite time, as measured by specific energy 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 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 target region and the selection of target species, the choice of energy norms, and the size of the target region.

Most of the uncertainty in the target 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 target regions the singular values decrease rapidly, and a few dominant TESVs are sufficient to capture most of the uncertainty. For large target 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 target 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 target 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 target area for longer intervals.

The dominant TESV has a similar structure to the adjoint variable. The next TESVs 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 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 [35].

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 0093139 . We would like to thank Virginia Tech's laboratory for Advanced Scientific Computing (LASCA) for the use of the Anantham cluster.

References

[1] J. Barkmeijer, R. Buizza, T.N. Palmer, K. Puri, and J.F. Mahfouf. Tropical singular vectors computed with linearized diabatic physics. *Quarterly Journal of the Royal Meteorological Society*, 127:685–708, 2001.

[2] M.D. Borges and D.L. Hartmann. Barotropic instability and optimal perturbations of observed non-zonal flow. *Journal of the Atmospheric Sciences*, 49:335–354, 1992.

[3] R. Buizza, J. Barkmeijer, T.N. Palmer, and D.S. Richardson. Current status and future developments of the ECMWF ensemble prediction system. *Meteorological Applications*, 7:163–175, 2000.

[4] R. Buizza and A. Montani. Targeting observations using singular vectors. *Journal of the Atmospheric Sciences*, 56:2965–2985, 1999.

[5] R. Buizza and T.N. Palmer. The singular-vector structure of the atmospheric general circulation. *Journal of the Atmospheric Sciences*, 53:1434–1456, 1995.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] R. Daley. *Atmospheric Data Analysis.* Cambridge University Press, 1991.

[11] V. Damian, A. Sandu, M. Damian, F. Potra, and G.R. Carmichael. The kinetic preprocessor KPP - a software environment for solving chemical kinetics. *Computers and Chemical Engineering*, 26:1567–1579, 2002.

[12] M. Ehrendorfer, R. Errico, and K.D. Raeder. Singular vector perturbation growth in a primitive equation model with moist physics. *Journal of the Atmospheric Sciences*, 56:1627–1648, 1999.

[13] M. Ehrendorfer and J.J. Tribbia. Optimal prediction of forecast error covariance through singular vectors. *Journal of the Atmospheric Sciences*, 54:286–313, 1997.

[14] M. Fisher. Assimilation techniques(5): Approximate Kalman filters and singular vectors, 2001.

[15] 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.

[16] T. Hamill, C. Snyder, and J. Whitaker. Approximate analysis error covariance singular vectors in a simple GCM. *EGS - AGU - EUG Joint Assembly, Abstracts from the meeting held in Nice, France, 6 - 11 April 2003, abstract #13876*, pages 13876–+, 2003.

[17] J.A. Hansen and A. Smith. The role of operational constraints in selecting supplementary observations. *Journal of the Atmospheric Sciences*, 57:2859–2871, 2000.

[18] P.L. Houtekamer and H.L. Mitchell. A sequential ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*, 129(1):123–137, 2000.

[19] A.H. Jazwinski. *Stochastic Processes and Filtering Theory*. Academic Press, 1970.

[20] Z. Kuang. The norm dependence of singular vectors. *Journal of the Atmospheric Sciences*, 61:2943–2949, 2004.

[21] R. Lehoucq, K. Maschhoff, D. Sorensen, and C. Yang. ARPACK software home page. http:// www.caam.rice.edu/ software/ ARPACK.

[22] Z. Li, I.M. Navon, and M.Y. Hussaini. Analysis of the singular vectors of the full-physics FSU global spectral model. *Tellus*, accepted, 2005.

[23] E.N. Lorenz. A study of the predictability of a 28 variable atmospheric model. *Tellus*, 17:321–333, 1965.

[24] 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.

[25] J.F. Mahfouf. Influence of physical processes on the tangent-linear approximation. *Tellus*, 51(A):147–166, 1999.

[26] 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.

[27] K. Maschhoff and D. Sorensen. Parallel ARPACK home page. http:// www.caam.rice.edu/ ∼kristyn/ parpack_home.html.

[28] 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.

[29] F. Molteni, R. Buizza, T.N. Palmer, and T. Petroliagis. The new ECMWF ensemble prediction system: methodology and validation. *Quarterly Journal of the Royal Meteorological Society*, 122:73–119, 1996.

[30] 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.

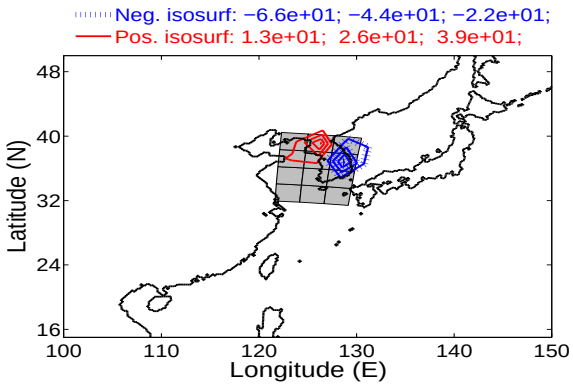
[31] R. Mureau, F. Molteni, and T.N. Palmer. Ensemble prediction using dynamically-conditioned perturbations. *Quarterly Journal of the Royal Meteorological Society*, 119:299–323, 1993.

[32] J. Oortwijn. Predictability of the onset of blocking and strong zonal flow regimes. *Journal of the Atmospheric Sciences*, 55:973–994, 1998.

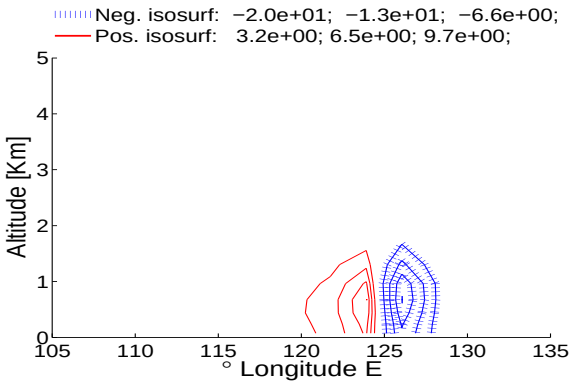
[33] T.N. Palmer, R. Gelaro, J. Barkmeijer, and R. Buizza. Singular vectors, metrics, and adaptive observations. *Journal of the Atmospheric Sciences*, 55(4):633–653, 1998.

[34] 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.

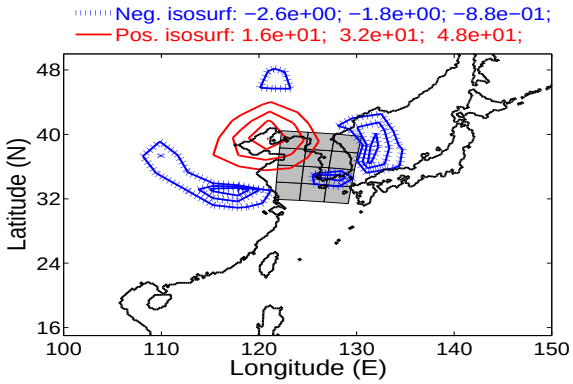
[35] A. Sandu, D. Daescu, G.R. Carmichael, and T. Chai. Adjoint sensitivity analysis of regional air quality models. *Journal of Computational Physics*, 2005. Accepted.



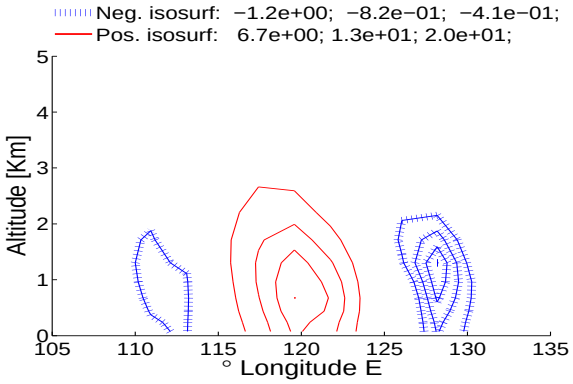
1st TESV (O_3), top view, 12h simulation



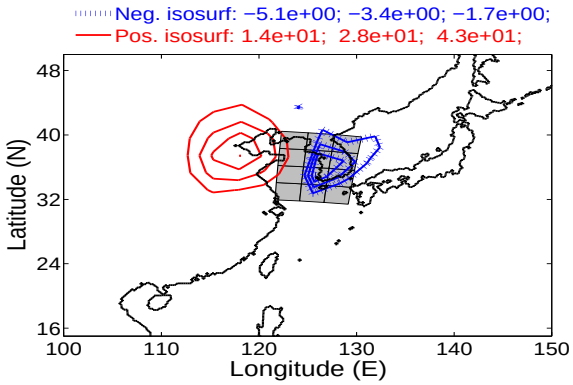
1st TESV (O_3), South view, 12h simulation



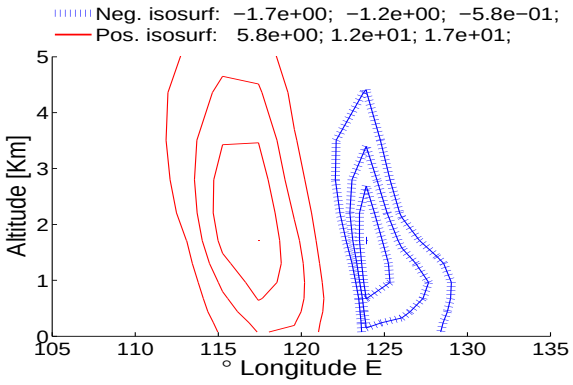
1st TESV (O_3), top view, 24h simulation



1st TESV (O_3), South view, 24h simulation

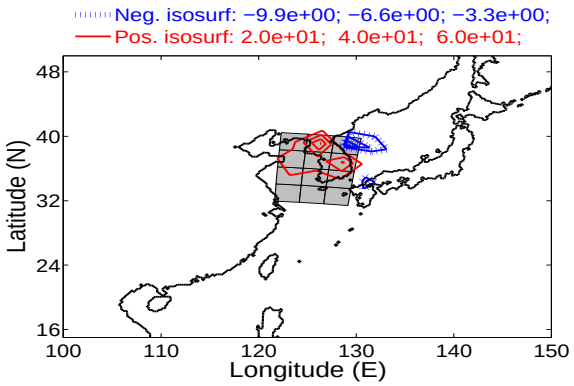


1st TESV (O_3), top view, 48h simulation

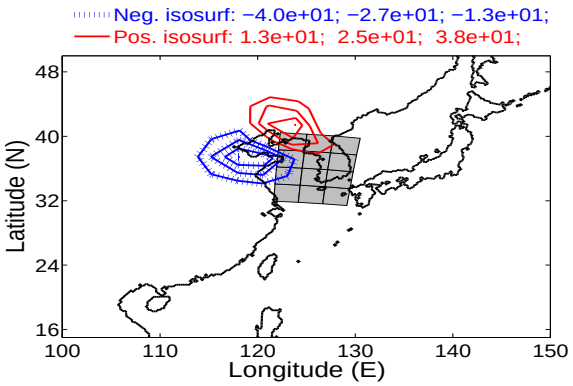


1st TESV (O_3), South view, 48h simulation

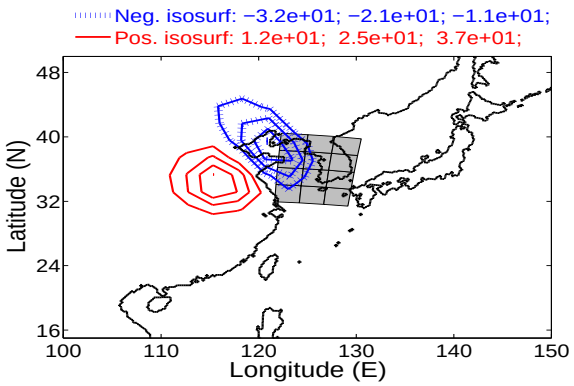
Figure 3: Dominant TESVs (O_3 sections) for the 12h, 24h and 48h simulations starting at 0 GMT March 1, 2001. The target is ground level O_3 value in the shaded area.



12h: 2nd TESV (O_3)

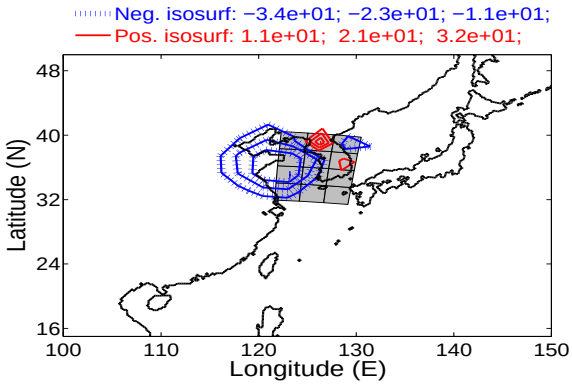


24h: 2nd TESV (O_3)

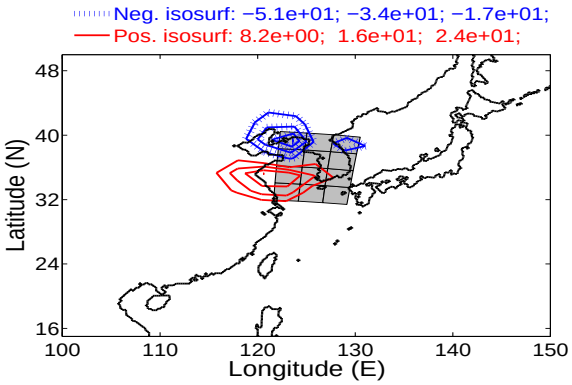


48h: 2nd TESV (O_3)

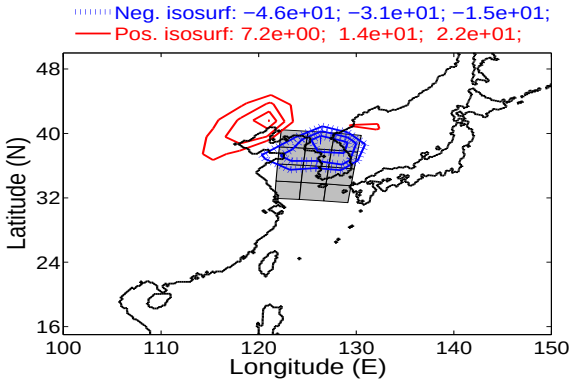
Figure 4: Second dominant TESVs (O_3 sections) for the 12h, 24h, and 48h simulations starting at 0 GMT March 1, 2001



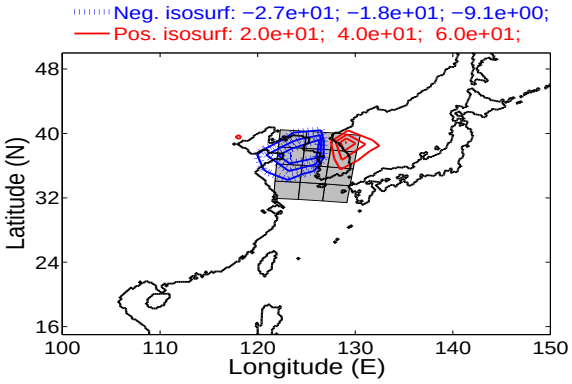
12h: 3rd TESV (O_3)



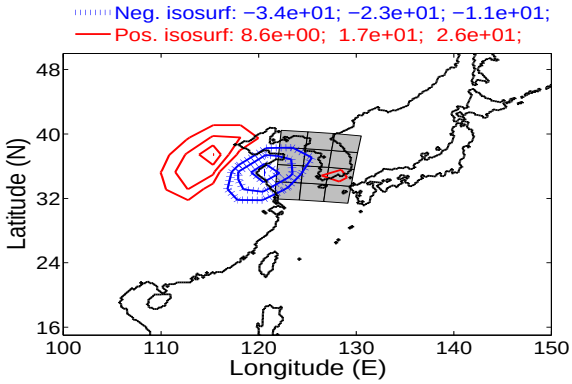
12h: 4th TESV (O_3)



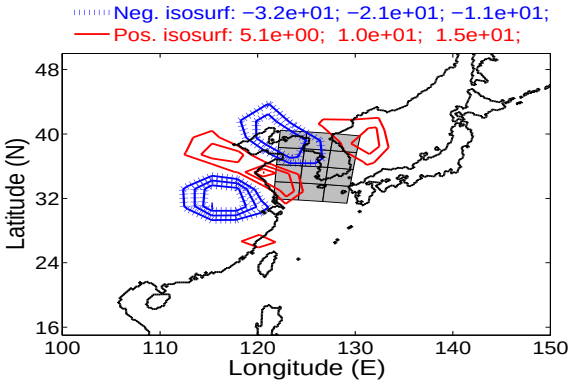
24h: 3rd TESV (O_3)



24h: 4th TESV (O_3)

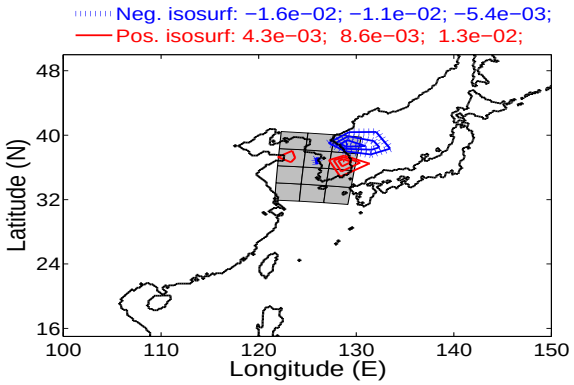


48h: 3rd TESV (O_3)

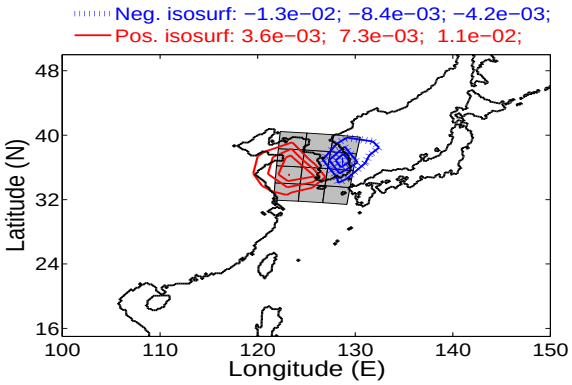


48h: 4th TESV (O_3)

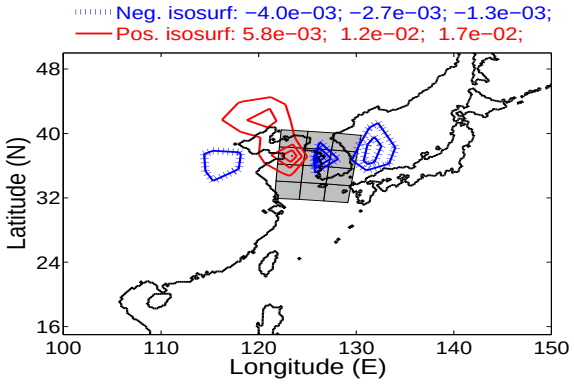
Figure 5: Third and fourth dominant TESVs (O_3 sections) for the 12h, 24h, and 48h simulations starting at 0 GMT March 1, 2001



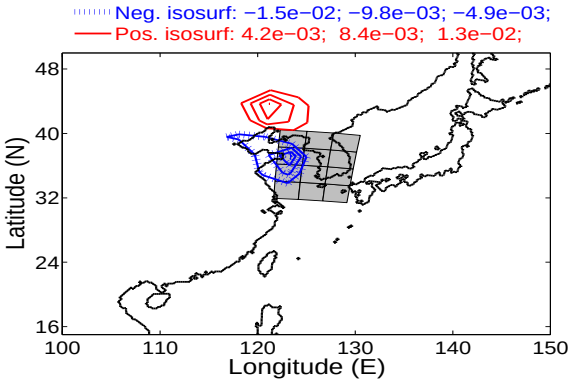
12h: 1st TESV (NO_2)



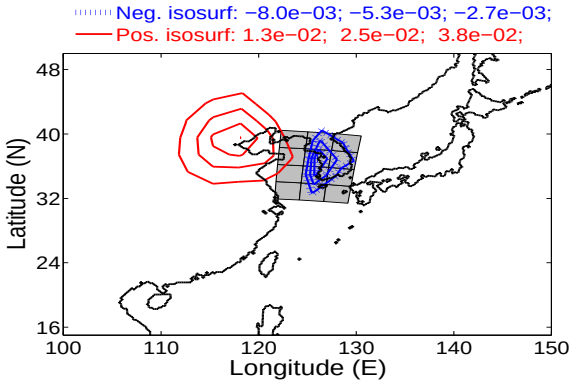
12h: 2nd TESV (NO_2)



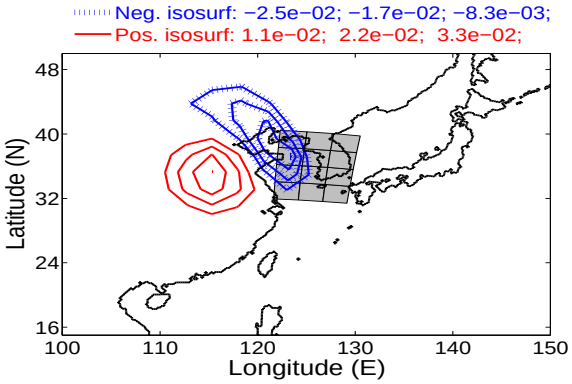
24h: 1st TESV (NO_2)



24h: 2nd TESV (NO_2)

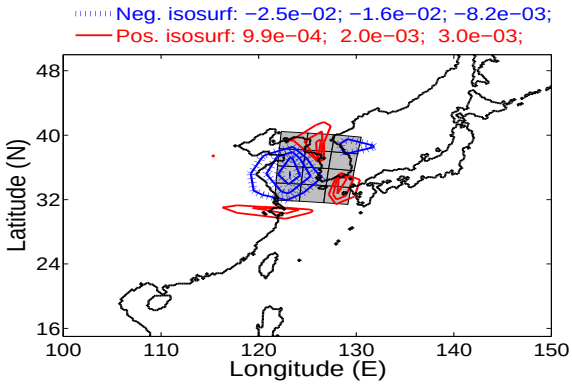


48h: 1st TESV (NO_2)

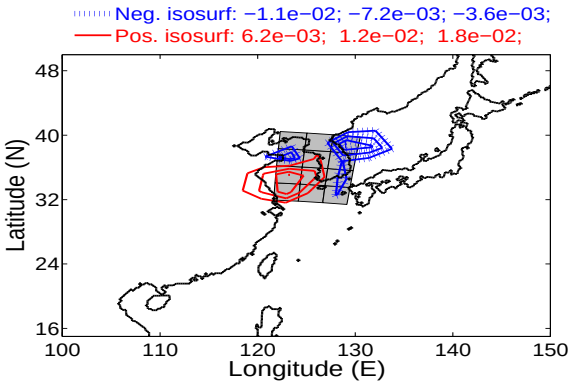


48h: 2nd TESV (NO_2)

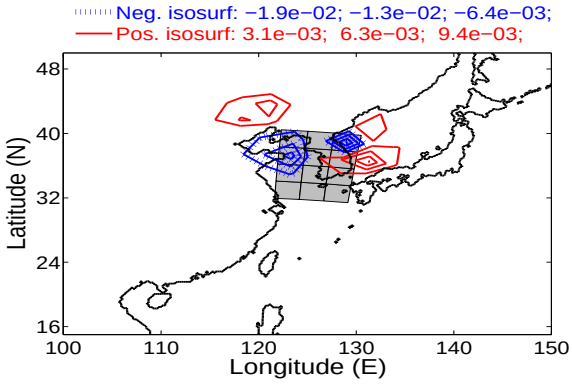
Figure 6: First and second dominant TESVs (NO_2 sections) for the 12h, 24h, and 48h simulations starting at 0 GMT March 1, 2001



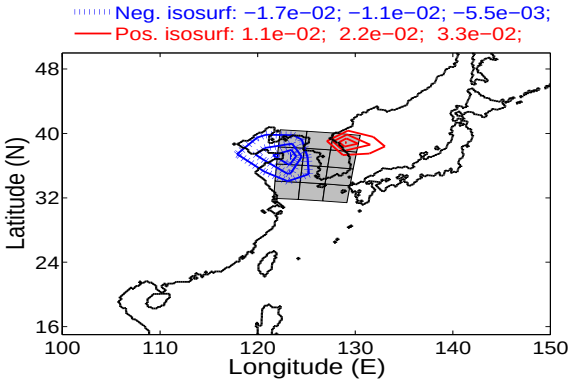
12h: 3rd TESV (NO_2)



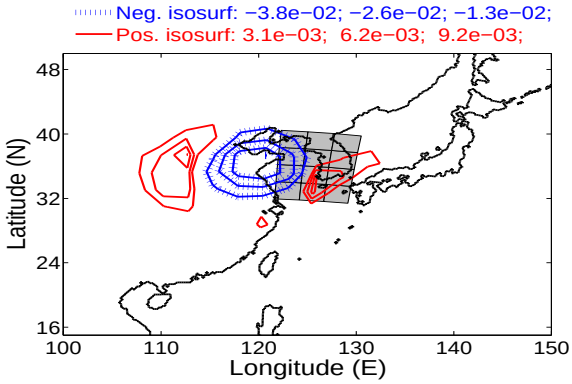
12h: 4th TESV (NO_2)



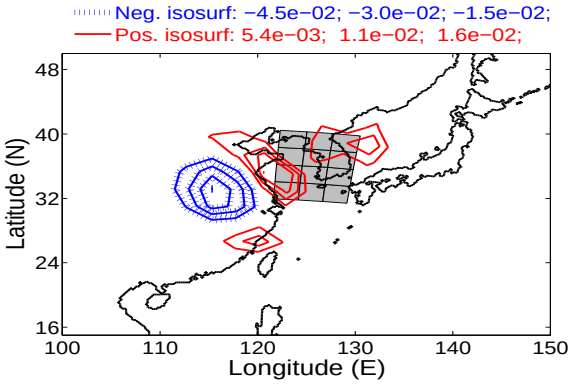
24h: 3rd TESV (NO_2)



24h: 4th TESV (NO_2)

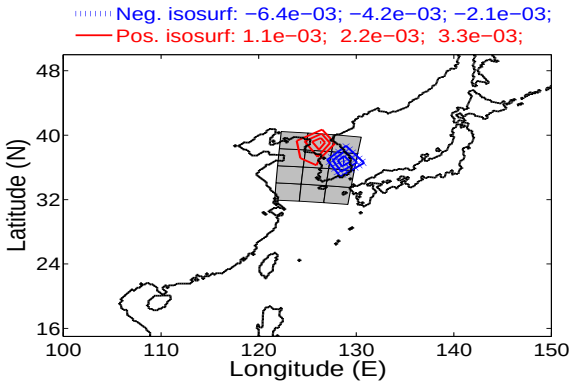


48h: 3rd TESV (NO_2)

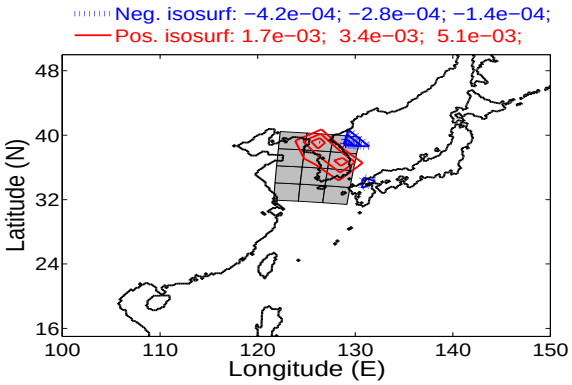


48h: 4th TESV (NO_2)

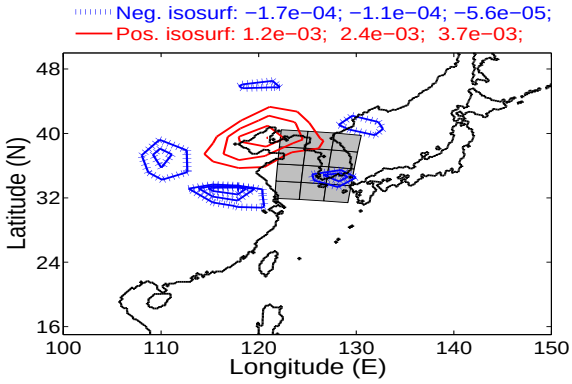
Figure 7: Third and fourth dominant TESVs (NO_2 sections) for the 12h, 24h, and 48h simulations starting at 0 GMT March 1, 2001



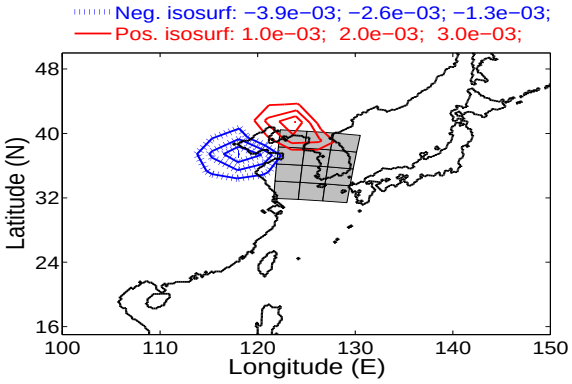
12h: 1st TESV ($HCHO$)



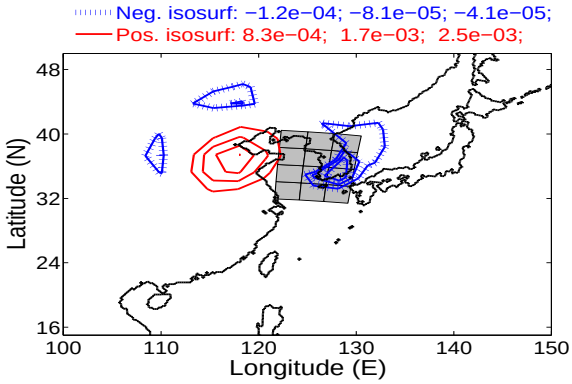
12h: 2nd TESV ($HCHO$)



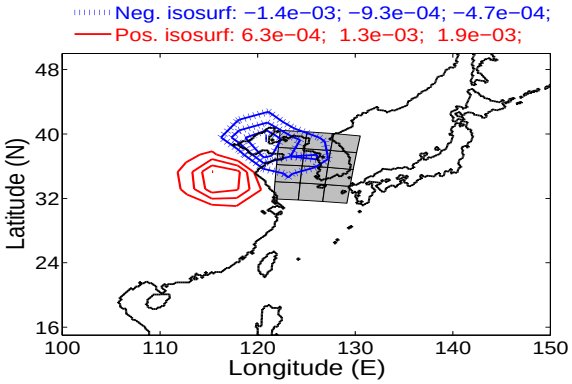
24h: 1st TESV ($HCHO$)



24h: 2nd TESV ($HCHO$)

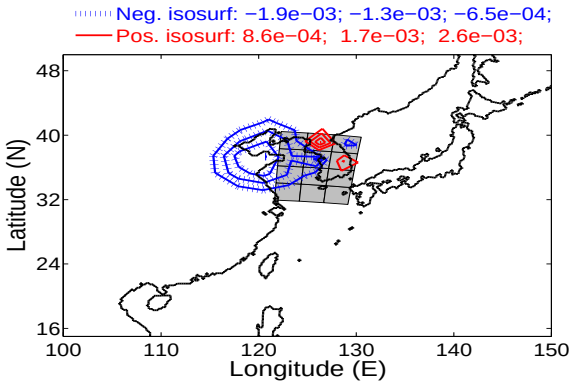


48h: 1st TESV ($HCHO$)

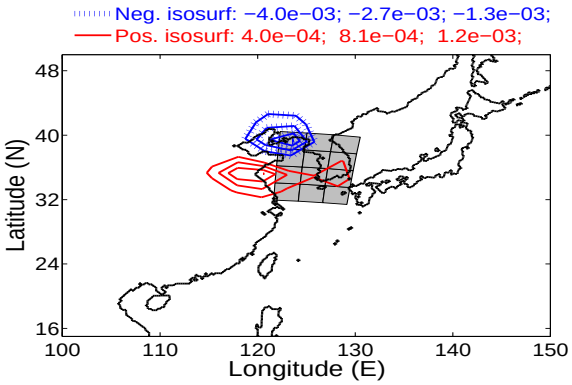


48h: 2nd TESV ($HCHO$)

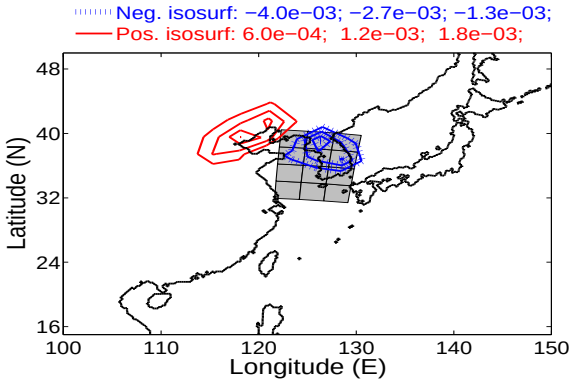
Figure 8: First and second dominant TESVs ($HCHO$ sections) for the 12h, 24h, and 48h simulations starting at 0 GMT March 1, 2001



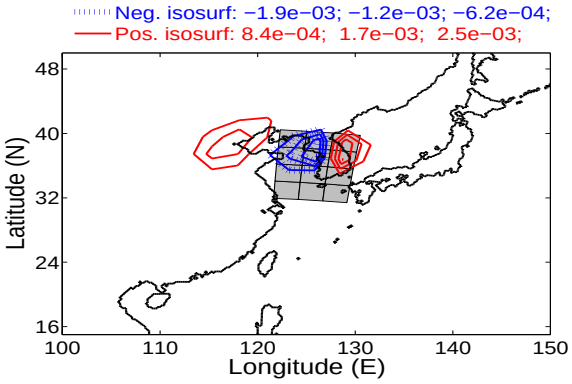
12h: 3rd TESV ($HCHO$)



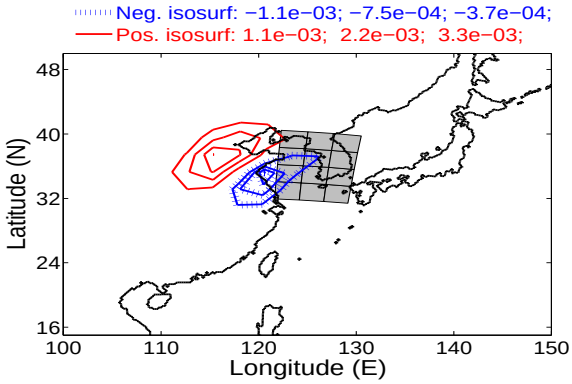
12h: 4th TESV ($HCHO$)



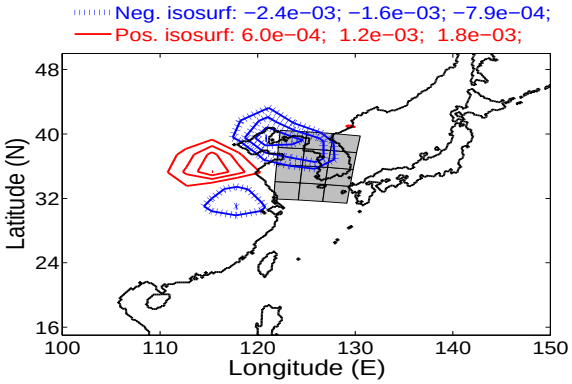
24h: 3rd TESV ($HCHO$)



24h: 4th TESV ($HCHO$)



48h: 3rd TESV ($HCHO$)



48h: 4th TESV ($HCHO$)

Figure 9: Third and fourth dominant TESVs ($HCHO$ sections) for the 12h, 24h, and 48h simulations starting at 0 GMT March 1, 2001

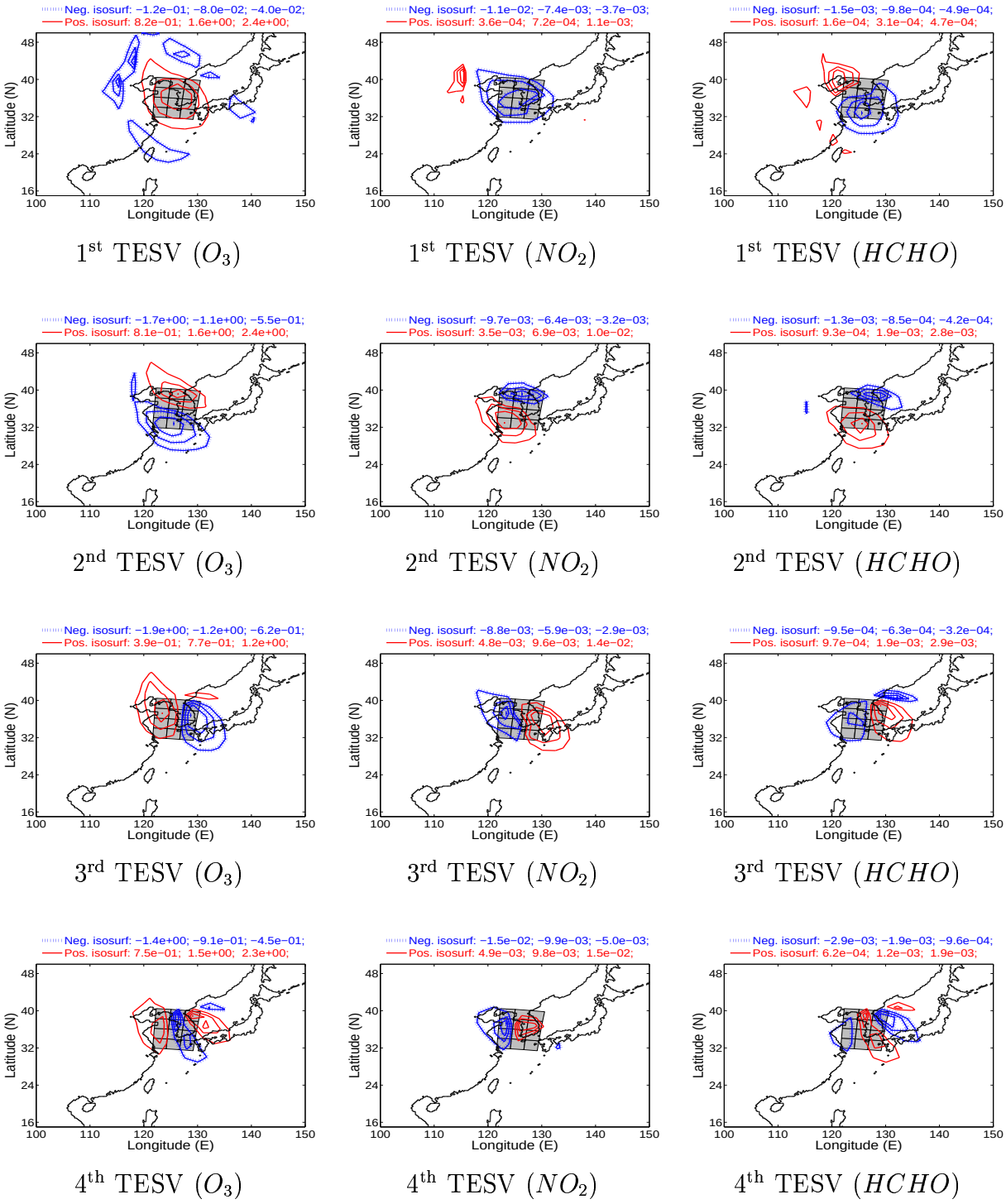


Figure 10: The evolved dominant TESVs after 24 hours of simulation (i.e., at 0 GMT, March 2, 2001).

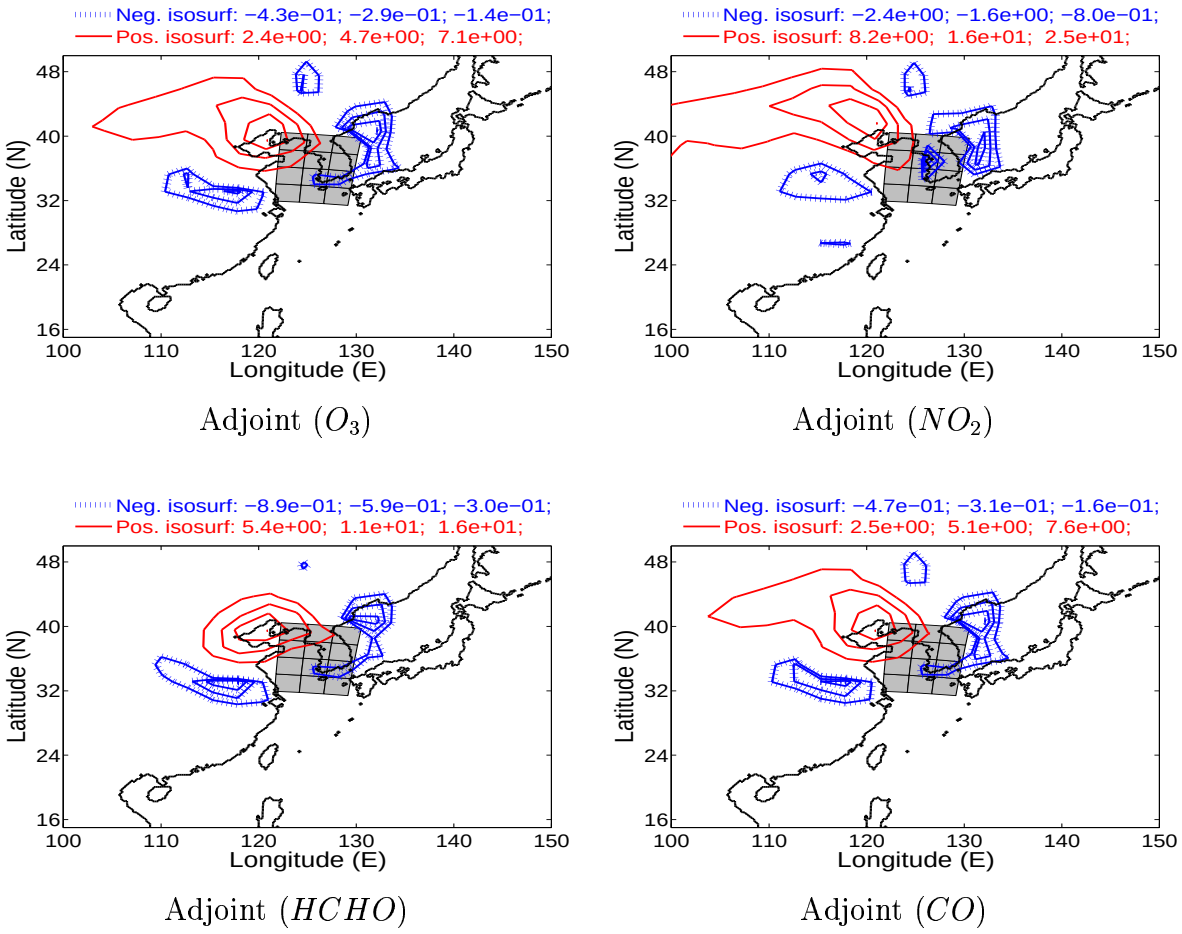


Figure 11: Adjoints for the 24h simulation starting at 0 GMT March 1, 2001

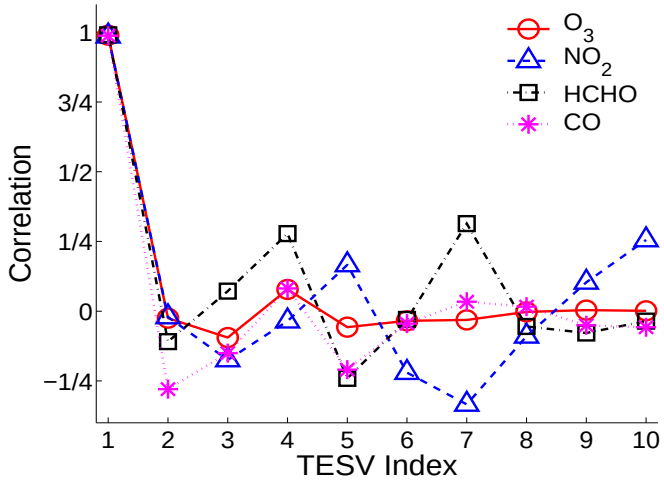
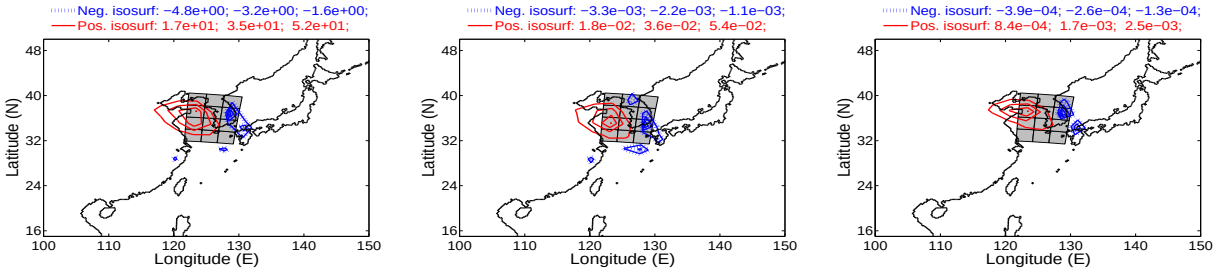
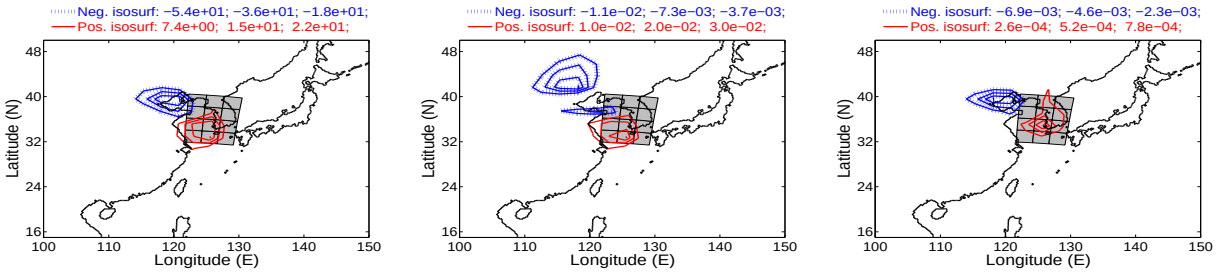


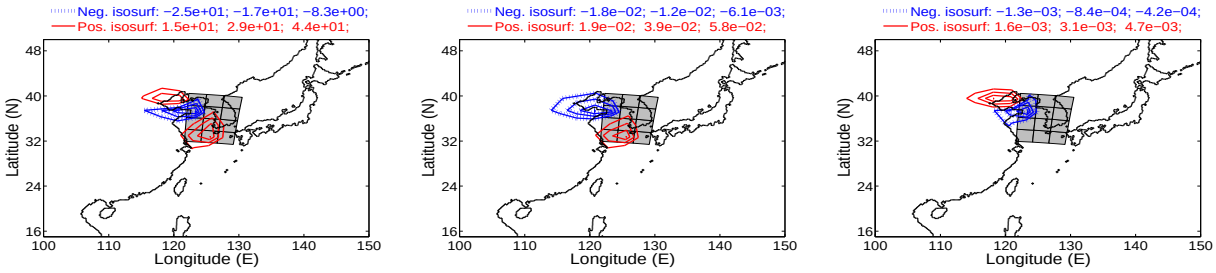
Figure 12: Correlations between homologous sections of the adjoint variable and of the dominant TESVs.



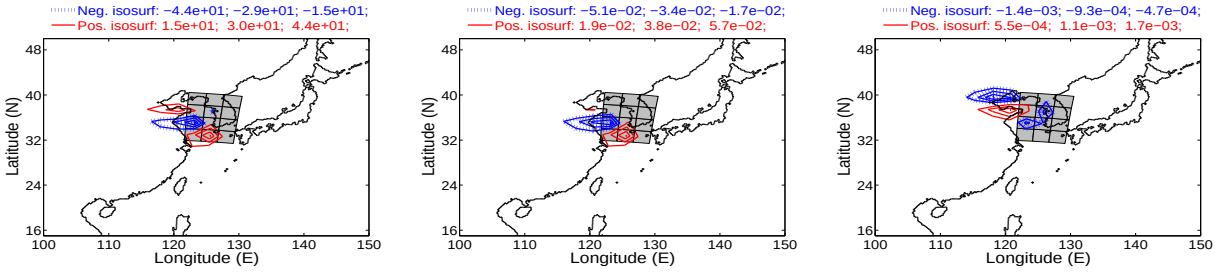
1st TESV (O_3) 1st TESV (NO_2) 1st TESV ($HCHO$)



2nd TESV (O_3) 2nd TESV (NO_2) 2nd TESV ($HCHO$)

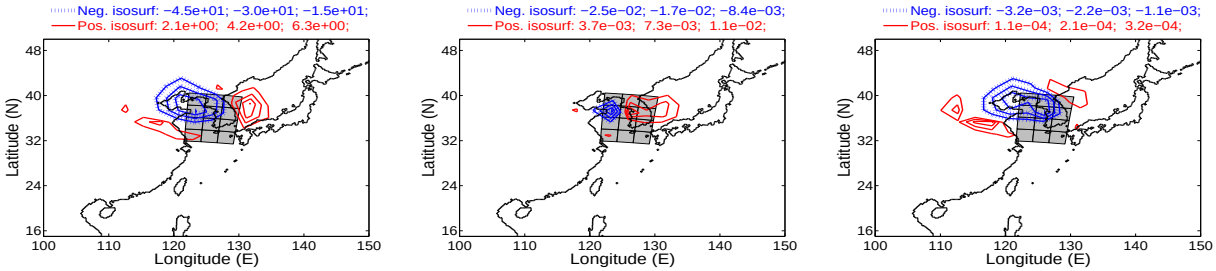


3rd TESV (O_3) 3rd TESV (NO_2) 3rd TESV ($HCHO$)

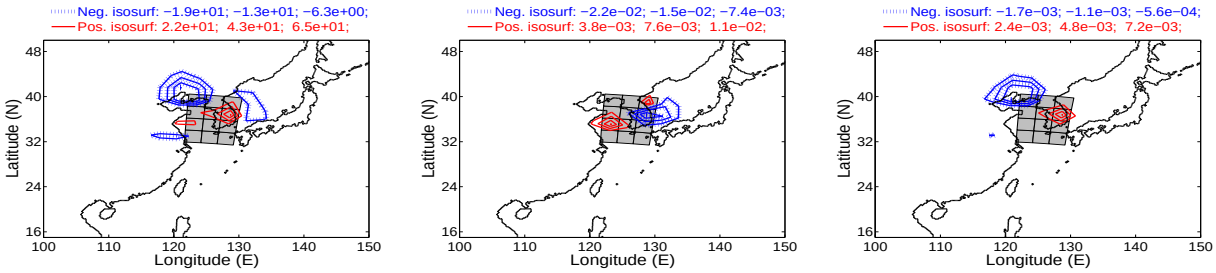


4th TESV (O_3) 4th TESV (NO_2) 4th TESV ($HCHO$)

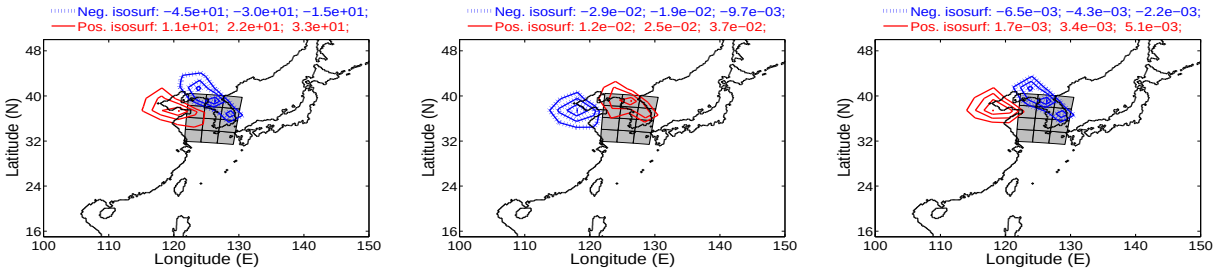
Figure 13: Dominant TESVs for the 24h simulation starting at 0 GMT, March 26, 2001



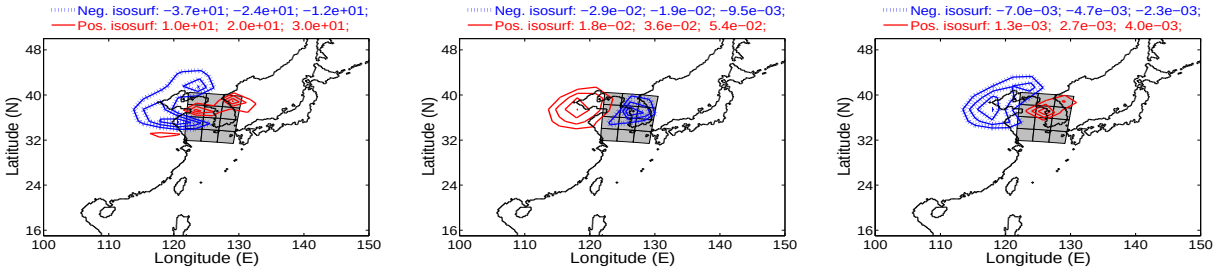
1st TESV (O_3) 1st TESV (NO_2) 1st TESV ($HCHO$)



2nd TESV (O_3) 2nd TESV (NO_2) 2nd TESV ($HCHO$)

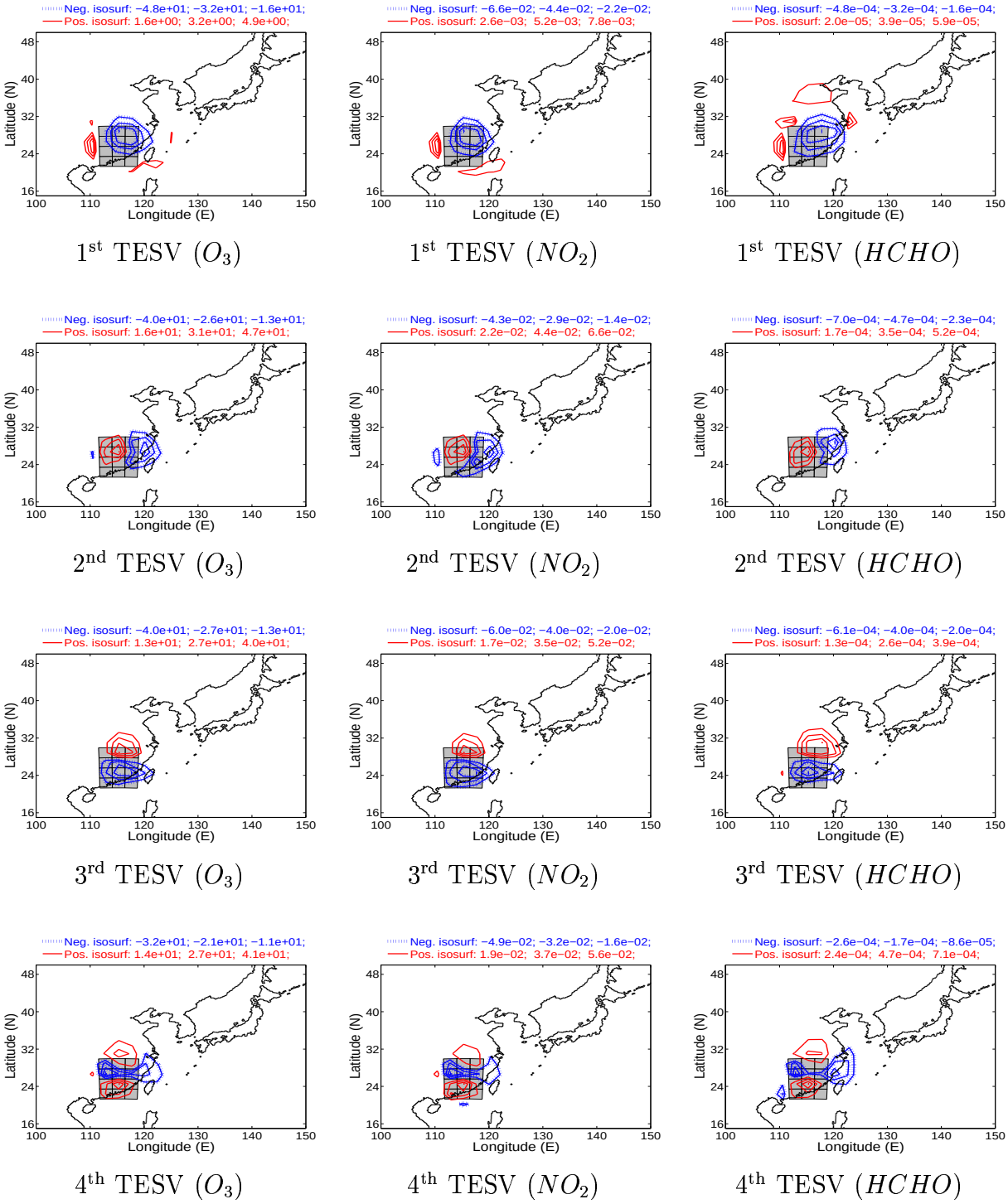


3rd TESV (O_3) 3rd TESV (NO_2) 3rd TESV ($HCHO$)



4th TESV (O_3) 4th TESV (NO_2) 4th TESV ($HCHO$)

Figure 14: Dominant TESVs for the 24h simulation for 66 species in the target



1st TESV (O_3) 1st TESV (NO_2) 1st TESV ($HCHO$)

2nd TESV (O_3) 2nd TESV (NO_2) 2nd TESV ($HCHO$)

3rd TESV (O_3) 3rd TESV (NO_2) 3rd TESV ($HCHO$)

4th TESV (O_3) 4th TESV (NO_2) 4th TESV ($HCHO$)

Figure 15: Dominant TESVs for the 24h simulation starting at 0 GMT March 1, 2001. The target is ground level O_3 over the gray region in South-East China

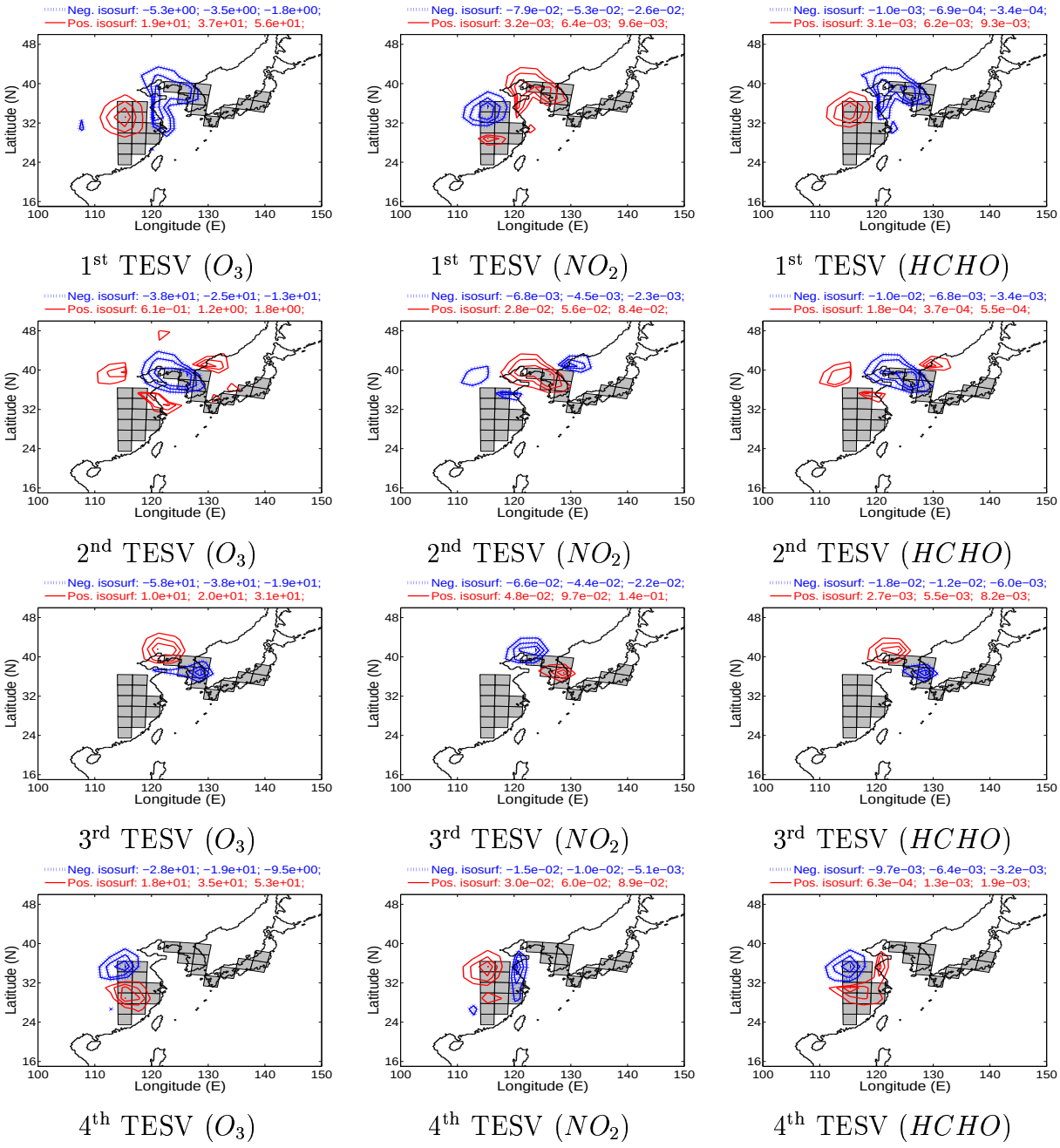


Figure 16: Dominant TESVs for the 24h simulation. The target is the 24-cell area covering Korea, Japan, and East China.

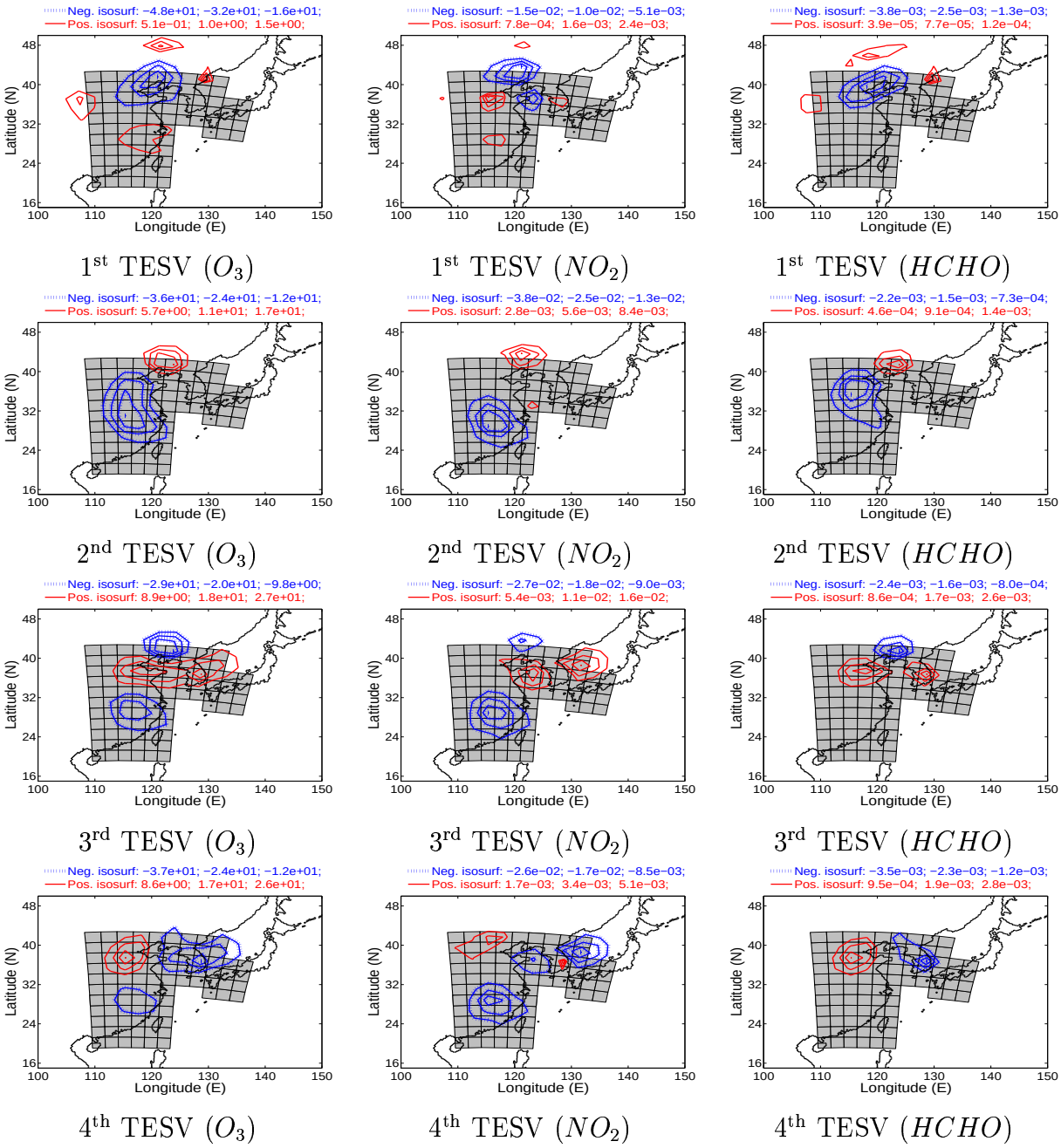


Figure 17: Dominant TESVs for the 24h simulation target at a Large area

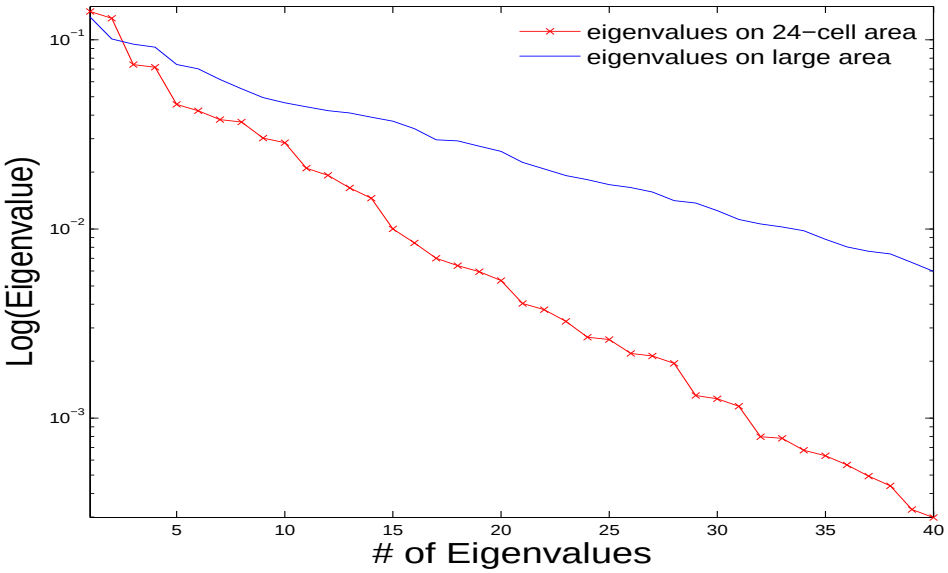


Figure 18: The dominant eigenvalues for simulations on 24-cell and large target areas.