Uncertainty Reduction in Atmospheric Composition Models by Chemical Data Assimilation

Adrian Sandu

Computational Science Laboratory

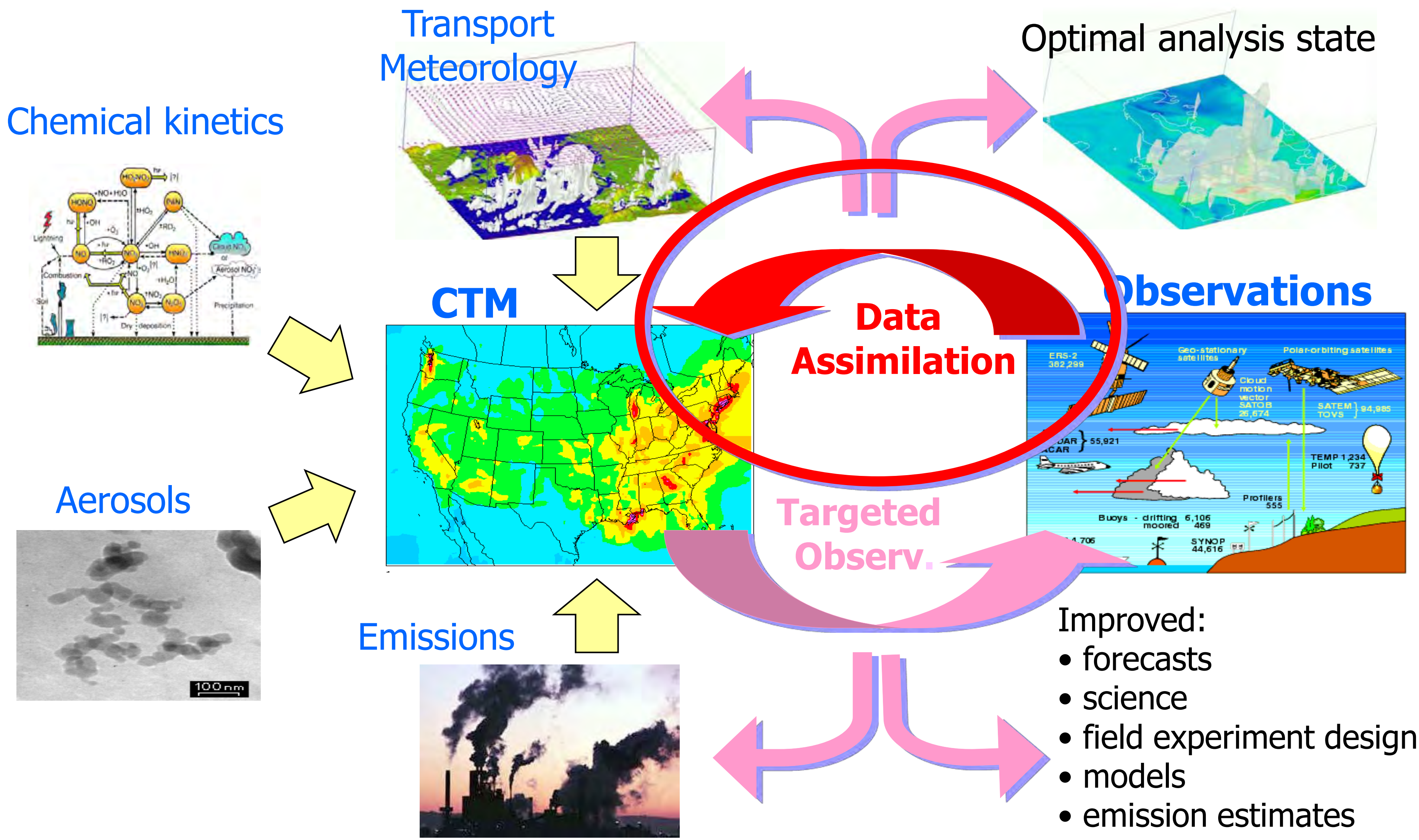
Virginia Polytechnic Institute

and State University

Aug. 4, 2011. IFIP UQ Workshop,

Boulder, CO

Information feedback loops between CTMs and observations: data assimilation and targeted meas.



CSL

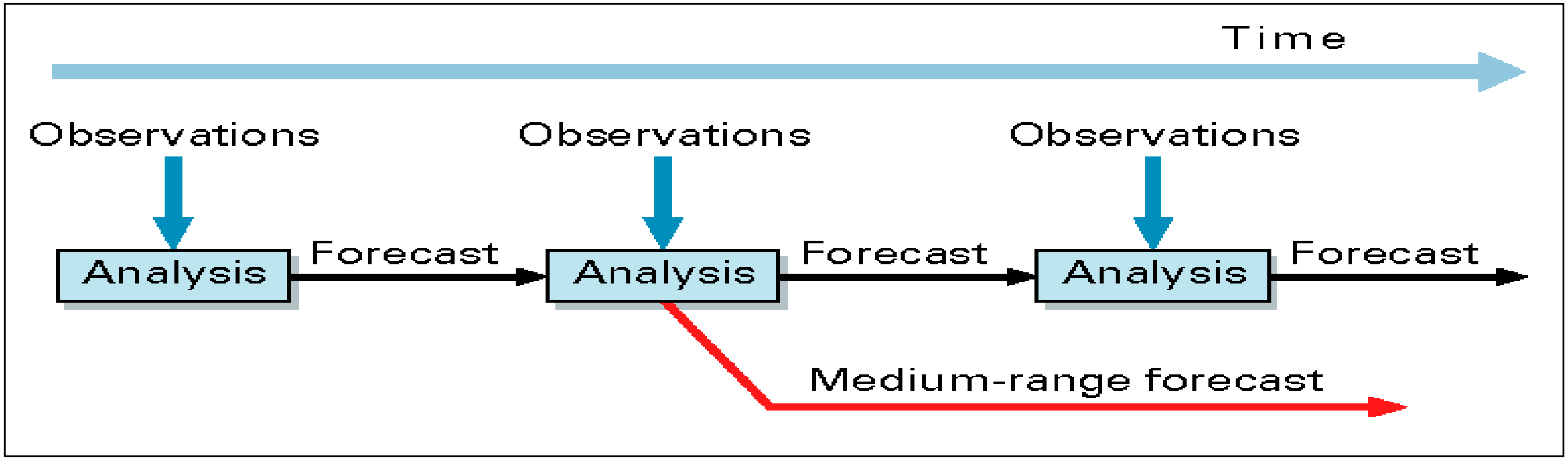
Virginia Tech

What is data assimilation?

The fusion of information from:

1. prior knowledge,
2. imperfect model predictions, and
3. sparse and noisy data,

to obtain a consistent description of the state of a physical system, such as the atmosphere.



Lars Isaksen (http://www.ecmwf.int)

Source of information #1: The prior encapsulates our *current knowledge* of the state

- The background (prior) probability density: $\mathcal{P}^{\rm b}(\mathbf{x})$
- The current best estimate: apriori (background) state $\mathbf{x}^{\rm b}$.
- Typical assumption on random background errors

$$\varepsilon^{\rm b} = \mathbf{x}^{\rm b} - \mathcal{S}(\mathbf{x}^{\rm true}) \in \mathcal{N}\left(\mathbf{0}, \mathbf{B}\right) .$$

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

Source of information #2: The model encapsulates our knowledge about physical and chemical laws that govern the evolution of the system

- The model evolves an initial state $\mathbf{x}_0 \in \mathbb{R}^n$ to future times

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

- Typical size of chemical transport models: $n \in \mathcal{O}\left(10^7\right)$ variables.
- The model is imperfect

$$\mathcal{S}\left(\mathbf{x}_i^{\text{true}}\right) = \mathcal{M}_{t_{i-1} \to t_i} \cdot \mathcal{S}\left(\mathbf{x}_{i-1}^{\text{true}}\right) - \eta_i ,$$

where η_i is the model error in step i.

Source of information #3: The observations are sparse and noisy snapshots of reality

- Measurements $\mathbf{y}_i \in \mathbb{R}^m$ ($m \ll n$) taken at times $t_1, \ldots, t_N$

$$\mathbf{y}_i = \mathcal{H}^{\mathrm{t}}\left(\mathbf{x}_i^{\mathrm{true}}\right) - \varepsilon_i^{\mathrm{instrument}} = \mathcal{H}\left(\mathcal{S}(\mathbf{x}_i^{\mathrm{true}})\right) - \varepsilon_i^{\mathrm{obs}}, \quad i = 1, \cdots, N\,.$$

- Observation operators
  - $\mathcal{H}^{\mathrm{t}}$: physical space $\rightarrow$ observation space, while
  - $\mathcal{H}$: the model space $\rightarrow$ observation space.
- The *observation error*

$$\varepsilon_i^{\mathrm{obs}} = \underbrace{\varepsilon_i^{\mathrm{instrument}}}_{\text{instrument error}} + \underbrace{\mathcal{H}\left(\mathcal{S}(\mathbf{x}_i^{\mathrm{true}})\right) - \mathcal{H}^{\mathrm{t}}\left(\mathbf{x}_i^{\mathrm{true}}\right)}_{\text{representativeness error}}$$

- Typical assumptions:

$$\varepsilon_i^{\mathrm{obs}} \in \mathcal{N}\left(\mathbf{0}, \mathbf{R}_i\right)\ ; \quad \varepsilon_i^{\mathrm{obs}}, \varepsilon_j^{\mathrm{obs}} \quad \text{independent for} \ \ t_i \neq t_j\,.$$

Result of data assimilation: The analysis encapsulates our *enhanced knowledge* of the state

- The analysis (posterior) probability density $\mathcal{P}^{\rm a}(\mathbf{x})$:

$$\text{Bayes:} \qquad \mathcal{P}^{\rm a}(\mathbf{x}) = \mathcal{P}(\mathbf{x}|\mathbf{y}) = \frac{\mathcal{P}(\mathbf{y}|\mathbf{x}) \cdot \mathcal{P}^{\rm b}(\mathbf{x})}{\mathcal{P}(\mathbf{y})} \,.$$

- The best state estimate $\mathbf{x}^{\rm a}$ is called the aposteriori, or the *analysis*.
- Analysis estimation errors $\varepsilon^{\rm a} = \mathbf{x}^{\rm a} - \mathcal{S}(\mathbf{x}^{\rm true})$ characterized by
  - *analysis mean error (bias)* $\beta^{\rm a} = \mathbb{E}^{\rm a}\left[\varepsilon^{\rm a}\right]$
  - *analysis error covariance matrix* $\mathbf{A} = \mathbb{E}^{\rm a}\left[\left(\varepsilon^{\rm a} - \beta^{\rm a}\right)\left(\varepsilon^{\rm a} - \beta^{\rm a}\right)^T\right] \in \mathbb{R}^{n \times n}$
- In the Gaussian, linear case, Bayes posterior admits an analytical solution by Kalman filter formulas

Extended Kalman filter

- The observations are considered successively at times $t_1, \cdots, t_N$.
- The background state at t_i given by the model forecast:

$$\mathbf{x}_i^{\mathrm{b}} \equiv \mathbf{x}_i^{\mathrm{f}} = \mathcal{M}_{t_{i-1}\to t_i} \cdot \mathbf{x}_{i-1}^{\mathrm{a}} \,.$$

- Model is imperfect, but is assumed unbiased

$$\eta_i \in \mathcal{N}(\mathbf{0}, \mathbf{Q}_i)$$

- Model error η_i and solution error $\varepsilon_{i-1}^{\mathrm{a}}$ are assumed independent; solution error small, propagated by linearized model $\mathbf{M} = \mathcal{M}'(\mathbf{x})$

$$\mathcal{O}(n^3): \quad \mathbf{B}_i \equiv \mathbf{P}_i^{\mathrm{f}} = \mathbf{M}_{t_{i-1}\to t_i}\, \mathbf{P}_{i-1}^{\mathrm{a}}\, \mathbf{M}_{t_i \to t_{i-1}}^{T} + \mathbf{Q}_i \,.$$

- EKF analysis uses $\mathbf{H}_i = \mathcal{H}'(\mathbf{x}_i^{\mathrm{f}})$:

$$\begin{aligned}
\mathcal{O}(nm): &\quad \mathbf{x}_i^{\mathrm{a}} = \mathbf{x}_i^{\mathrm{f}} + \mathbf{K}_i \left(\mathbf{y}_i - \mathcal{H}(\mathbf{x}_i^{\mathrm{f}})\right) \\
\mathcal{O}(nm^2 + n^2m + m^3): &\quad \mathbf{K}_i = \mathbf{P}_i^{\mathrm{f}}\, \mathbf{H}_i^{\mathrm{T}} \left(\mathbf{H}_i\, \mathbf{P}_i^{\mathrm{f}}\, \mathbf{H}_i^{\mathrm{T}} + \mathbf{R}_i\right)^{-1} \\
\mathcal{O}(n^2m + n^3): &\quad \mathbf{A}_i \equiv \mathbf{P}_i^{\mathrm{a}} = \left(\mathbf{I} - \mathbf{K}_i\, \mathbf{H}_i\right) \mathbf{P}_i^{\mathrm{f}} \,.
\end{aligned}$$

Practical Kalman filter methods

- EKF is not practical for very large systems
- Suboptimal KF approximate the covariance matrices e.g.,

$$\mathbf{B}_{(\ell),(k)} = \sigma_{(\ell)}\,\sigma_{(k)}\ \exp\left(\mathrm{distance}\{\mathrm{gridpoint}(\ell), \mathrm{gridpoint}(k)\}^2/L^2\right)$$

- Ensemble Kalman filters (EnKF) use a Monte-Carlo approach

$$\begin{aligned}
\mathbf{x}_i^{\mathrm{f}}[e] &= \mathcal{M}_{t_{i-1}\to t_i}\left(\mathbf{x}_{i-1}^{\mathrm{a}}[e]\right) + \underbrace{\eta_i[e]}_{\text{model error}}\ , \quad e = 1,\ldots,E \\
\mathbf{x}_i^{\mathrm{a}}[e] &= \mathbf{x}_i^{\mathrm{f}}[e] + \mathbf{K}_i\left(\mathbf{y}_i + \varepsilon_i^{\mathrm{obs}}[e] - \mathcal{H}_i(\mathbf{x}_i^{\mathrm{f}}[e])\right), \quad e = 1,\ldots,E\,.
\end{aligned}$$

- Error covariances $\mathbf{P}_i^{\mathrm{f}}$, $\mathbf{P}_i^{\mathrm{a}}$ estimated from statistical samples
- EnKF issues: rank-deficiency of the estimated $\mathbf{P}_i^{\mathrm{f}}$
- EnKF strengths: capture non-linear dynamics, doesn't need TLM, ADJ, accounts for model errors, almost ideally parallelizable

Maximum aposteriori estimator

Maximum aposteriori estimator (MAP) defined by

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

Using Bayes and assumptions for background, observation errors:

$$\begin{aligned} \mathcal{J}(\mathbf{x}) &= -\ln \mathcal{P}^{\mathrm{a}}(\mathbf{x}) = -\ln \mathcal{P}^{\mathrm{b}}(\mathbf{x}) - \ln \mathcal{P}(\mathbf{y}|\mathbf{x}) + \mathrm{const} \\ &\doteq \frac{1}{2} \left(\mathbf{x} - \mathbf{x}^{\mathrm{b}}\right)^T \mathbf{B}^{-1} \left(\mathbf{x} - \mathbf{x}^{\mathrm{b}}\right) + \frac{1}{2} \left(\mathcal{H}(\mathbf{x}) - \mathbf{y}\right)^T \mathbf{R}^{-1} \left(\mathcal{H}(\mathbf{x}) - \mathbf{y}\right) \end{aligned}$$

Optimization by gradient-based numerical procedure

$$\nabla_{\mathbf{x}} \mathcal{J}(\mathbf{x}^{\mathrm{a}}) = \mathbf{B}^{-1} \left(\mathbf{x}^{\mathrm{a}} - \mathbf{x}^{\mathrm{b}}\right) + \mathbf{H}^T \mathbf{R} \left(\mathcal{H}(\mathbf{x}^{\mathrm{a}}) - \mathbf{y}\right) \,; \quad \mathbf{H} = \mathcal{H}(\mathbf{x}^{\mathrm{b}}) \,.$$

Hessian of cost function approximates inverse analysis covariance

$$\nabla^2_{\mathbf{x},\mathbf{x}} \mathcal{J} = \mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \approx \mathbf{A}^{-1} \,.$$

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

- All observations at all times $t_1, \cdots, t_N$ are considered simultaneously
- The control variables (parameters $\mathbf{p}$, initial conditions $\mathbf{x}_0$, boundary conditions, etc) uniquely determine the state of the system at all future times
- 4D-Var MAP estimate via model-constrained optimization problem

$$
\begin{aligned}
\mathcal{J}(\mathbf{x}_0) &= \frac{1}{2}\left\|\mathbf{x}_0 - \mathbf{x}_0^{\mathrm{b}}\right\|^2_{\mathbf{B}_0^{-1}} + \frac{1}{2}\sum_{i=1}^{N}\|\mathcal{H}(\mathbf{x}_i) - \mathbf{y}_i\|^2_{\mathbf{R}_i^{-1}} \\
\mathbf{x}_0^{\mathrm{a}} &= \arg\min \mathcal{J}(\mathbf{x}_0) \\
& \quad \text{subject to: } \mathbf{x}_i = \mathcal{M}_{t_0 \to t_i}(\mathbf{x}_0)\,, \quad i = 1, \cdots, N
\end{aligned}
$$

- Formulation can be easily extended to other model parameters

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

- The large scale optimization problem is solved in a reduced space using a gradient-based technique.
- The 4D-Var gradient reads

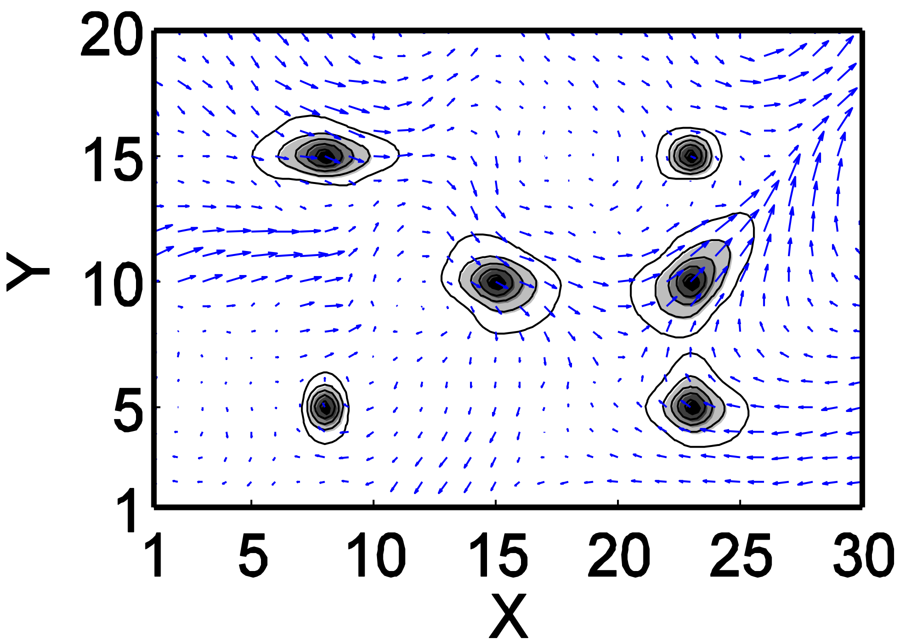
$$\nabla \mathcal{J}_{\mathbf{x}_0}(\mathbf{x}_0) = \mathbf{B}_0^{-1}\,\left(\mathbf{x}_0 - \mathbf{x}_0^{\mathrm{b}}\right) + \sum_{i=1}^{N} \left(\frac{\partial \mathbf{x}_i}{\partial \mathbf{x}_0}\right)^T \mathbf{H}_i^T\,\mathbf{R}_i^{-1}\,\left(\mathcal{H}(\mathbf{x}_i) - \mathbf{y}_i\right)$$

- Needs linearized observation operators $\mathbf{H}_i = \mathcal{H}'(\mathbf{x}_i)$
- Needs the transposed sensitivity matrix $(\partial \mathbf{x}_i / \partial \mathbf{x}_0)^T \in \mathbb{R}^{n \times n}$
- Adjoint models efficiently compute the transposed sensitivity matrix times vector products
- The construction of an adjoint model is a nontrivial task.

Correct models of background errors are of great importance for data assimilation

- Background error representation determines the spread of information, and impacts the assimilation results
- Needs: high rank, capture dynamic dependencies, efficient computations
- Traditionally estimated empirically (NMC, Hollingsworth-Lonnberg)

1. Tensor products of 1d correlations, decreasing with distance (Singh et al, 2010)
2. Multilateral AR model of background errors based on "monotonic TLM discretizations" (Constantinescu et al 2007)
3. Hybrid methods in the context of 4D-Var (Cheng et al, 2007)



[Constantinescu and Sandu, 2007]

What is the effect of mis-specification of inputs?

(Daescu, 2008) Consider a *verification functional* $\Psi(\mathbf{x}_v^{\mathrm{a}})$ defined on the optimal solution at a future time t_v. Ψ is a measure of the forecast error. What is the impact of small errors in the specification of covariances, background, and observation data?

$$
\begin{aligned}
\nabla_{\mathbf{y}_i}\Psi &= \mathbf{R}_i^{-1}\mathbf{H}_i\mathbf{M}_{t_0\to t_i}\left(\left(\nabla^2_{\mathbf{x}_0,\mathbf{x}_0}\mathcal{J}\right)^{-1}\nabla_{\mathbf{x}_0}\Psi\right)\\
\nabla_{\mathbf{R}_i(:)}\Psi &= \left(\mathbf{R}_i^{-1}(\mathcal{H}(\mathbf{x}_i^{\mathrm{a}})-\mathbf{y}_i)\right)\otimes\nabla_{\mathbf{y}_i}\Psi\\
\nabla_{\mathbf{x}^{\mathrm{b}}}\Psi &= \mathbf{B}_0^{-1}\left(\left(\nabla^2_{\mathbf{x}_0,\mathbf{x}_0}\mathcal{J}\right)^{-1}\nabla_{\mathbf{x}_0}\Psi\right)\\
\nabla_{\mathbf{B}_0(:)}\Psi &= \left(\mathbf{B}_0^{-1}(\mathbf{x}_0^{\mathrm{a}}-\mathbf{x}_0^{\mathrm{b}})\right)\otimes\nabla_{\mathbf{x}^{\mathrm{b}}}\Psi
\end{aligned}
$$

General framework for sensitivity analysis

Forward model equations link parameters and solutions:

$$\mathcal{F}(\mathbf{x},\theta) = \mathbf{0} \in \mathcal{H}_F \,. \quad \mathcal{H}_F\text{= model constraint space, Hilbert: } \langle\cdot,\cdot\rangle_{\mathcal{H}_F}$$

$$\mathbf{x} \in \mathcal{H}_{\mathbf{x}} \,. \quad \mathcal{H}_{\mathbf{x}}\text{= model state space, Hilbert: } \langle\cdot,\cdot\rangle_{\mathcal{H}_{\mathbf{x}}} \,,$$

$$\theta \in \mathcal{H}_\theta \,. \quad \mathcal{H}_\theta\text{= parameter space, Hilbert: } \langle\cdot,\cdot\rangle_{\mathcal{H}_\theta} \,.$$

The response functional (QoI) associates a real value to each state

$$\mathcal{J}(\mathbf{x}) : \mathcal{H}_{\mathbf{x}} \longrightarrow \mathbb{R} \quad \left(e.g.,\ \mathcal{J}(\mathbf{x}) = \frac{1}{2}\,\|\mathcal{H}(\mathbf{x}) - \mathbf{y}\|^2_{\mathbf{R}^{-1}} \right)$$

Assumptions:

1. $\mathcal{F}$, $\mathcal{J}$ are continuously Frèchet differentiable.
2. $\mathcal{F}_{\mathbf{x}}$ has a continuous linear inverse mapping. By IFT a Frèchet differentiable model solution operator $\mathbf{x} = \mathcal{M}(\theta)$ exists locally

$$\mathcal{M} : \mathcal{H}_\theta \to \mathcal{H}_{\mathbf{x}} \,; \quad \mathbf{x} = \mathcal{M}(\theta) \,; \quad \mathcal{M}'(\theta) = -\mathcal{F}_{\mathbf{x}}^{-1}(\mathbf{x},\theta)\cdot\mathcal{F}_\theta(\mathbf{x},\theta) \,.$$

Formulation of the inverse problem as a model-constrained optimization problem

Find the optimal vector of parameters θ_{opt} such that:

$$\theta^* = \arg\min_{\theta} \mathcal{J}(\mathbf{x})$$
$$\text{subject to } \mathcal{F}(\mathbf{x}, \theta) = 0 .$$

Comments.

1. The cost function depends implicitly on the parameters:

$$\mathcal{J}(\mathbf{x}) = \mathcal{J}\left(\mathcal{M}(\theta)\right) = \left(\mathcal{J} \circ \mathcal{M}\right)\ (\theta)\ .$$

2. Gradient-based optimization techniques require

$$\nabla_\theta \mathcal{J} = \mathcal{M}'^*(\theta) \cdot \nabla_{\mathbf{x}} \mathcal{J}$$

3. Difficulty: model solution operator is only defined implicitly.

Direct (forward) vs. adjoint sensitivity analysis

1. Tangent linear model is obtained by Frèchet differentiation

$$(TLM): \quad \mathcal{F}_\theta(\theta, \mathbf{x}) \cdot \delta\theta + \mathcal{F}_{\mathbf{x}}(\theta, \mathbf{x}) \cdot \delta\mathbf{x} = 0 \in \mathcal{H}_F \; .$$

$$\delta\mathcal{J} = \langle \nabla_{\mathbf{x}}\mathcal{J}, \delta\mathbf{x} \rangle_{\mathcal{H}_{\mathbf{x}}} = \langle \nabla_\theta\mathcal{J}, \delta\theta \rangle_{\mathcal{H}_\theta} \; .$$

Comment. $\nabla_{\mathbf{x}}\mathcal{J}$ by direct differentiation. One TLM solution provides one inner product . To find the entire gradient $\nabla_\theta\mathcal{J}$ ...

Direct (forward) vs. adjoint sensitivity analysis

1. Tangent linear model is obtained by Frèchet differentiation

$$(TLM): \quad \mathcal{F}_\theta(\theta, \mathbf{x}) \cdot \delta\theta + \mathcal{F}_{\mathbf{x}}(\theta, \mathbf{x}) \cdot \delta\mathbf{x} = 0 \in \mathcal{H}_F \,.$$

$$\delta\mathcal{J} = \langle \nabla_{\mathbf{x}}\mathcal{J}, \delta\mathbf{x} \rangle_{\mathcal{H}_{\mathbf{x}}} = \langle \nabla_\theta\mathcal{J}, \delta\theta \rangle_{\mathcal{H}_\theta} \,.$$

Comment. $\nabla_{\mathbf{x}}\mathcal{J}$ by direct differentiation. One TLM solution provides one inner product . To find the entire gradient $\nabla_\theta\mathcal{J}$ ...

2. *Adjoint model* obtained using duality:

$$\begin{aligned} (\lambda \in \mathcal{H}_F^* \equiv \mathcal{H}_F) \quad &\Leftrightarrow \quad \langle \lambda, \mathcal{F}_{\mathbf{x}} \cdot \delta\mathbf{x} \rangle_{\mathcal{H}_F} + \langle \lambda, \mathcal{F}_\theta \cdot \delta\theta \rangle_{\mathcal{H}_F} = 0 \in \mathbb{R} \\ (\textit{by adjoint}) \quad &\Leftrightarrow \quad \langle \mathcal{F}_{\mathbf{x}}^* \cdot \lambda, \delta\mathbf{x} \rangle_{\mathcal{H}_{\mathbf{x}}} + \langle \mathcal{F}_\theta^* \cdot \lambda, \delta\theta \rangle_{\mathcal{H}_\theta} = 0 \in \mathbb{R} \,. \end{aligned}$$

$$ADJ: \quad \mathcal{F}_{\mathbf{x}}^* \cdot \lambda = -\nabla_{\mathbf{x}}\mathcal{J}(\mathbf{x}) \,.$$

$$\langle \nabla_{\mathbf{x}}\mathcal{J}(\mathbf{x}), \delta\mathbf{x} \rangle_{\mathcal{H}_{\mathbf{x}}} = \langle (\mathcal{F}_\theta)^* \cdot \lambda, \delta\theta \rangle_{\mathcal{H}_\theta} = \langle \nabla_\theta\mathcal{J}, \delta\theta \rangle_{\mathcal{H}_\theta} = \delta\mathcal{J} \,.$$

Comment. Adjoint model does not depend on the particular perturbations $\delta\theta$, $\delta\mathbf{x}$, and needs to be solved *only once*.

Continuous and discrete adjoints of mass balance equations lead to different computational models

$$\nabla_{\mathbf{y}^0} \psi = \cdots + \sum_{k=1}^{N} \left(\partial \mathbf{y}^k / \partial \mathbf{y}^0\right)^{\mathbf{T}} \left(\mathbf{H}^k\right)^{\mathbf{T}} \mathbf{R}_k^{-1} \left(\mathbf{H}^k \mathbf{y}^k - \mathbf{z}_{obs}^k\right)$$

Continuous forward model

$$\frac{dC_i}{dt} = -\vec{u} \cdot \nabla C_i + \frac{1}{\rho} \nabla\left(\rho K \cdot \nabla C_i\right) + \frac{1}{\rho} f_i(\rho C) + E_i$$

$$C_i\left(t^0, x\right) = C_i^0(x), \qquad t^0 \le t \le t^F$$

$$C_i(t, x) = C_i^{IN}(t, x) \quad on \quad \Gamma^{IN}$$

$$K \frac{\partial C_i}{\partial n} = 0 \quad on \quad \Gamma^{OUT}$$

$$K \frac{\partial C_i}{\partial n} = V_i^{DEP} C_i - Q_i \quad on \quad \Gamma^{GROUND}$$

adj

Continuous adjoint model

$$\frac{d\lambda_i}{dt} = -\nabla \cdot \left(\vec{u} \lambda_i\right) - \nabla \cdot \left(\rho K \cdot \nabla \frac{\lambda_i}{\rho}\right) - \left(F^T(\rho C) \cdot \lambda\right)_i - \phi_i$$

$$\lambda_i\left(t^F, x\right) = \lambda_i^F(x), \qquad t^F \ge t \ge t^o$$

$$\lambda_i(t, x) = 0 \quad on \quad \Gamma^{IN}$$

$$\vec{u} \lambda_i + \rho K \frac{\partial\left(\lambda_i / \rho\right)}{\partial \vec{n}} = 0 \quad on \quad \Gamma^{OUT}$$

$$\rho K \frac{\partial\left(\lambda_i / \rho\right)}{\partial \vec{n}} = V_i^{DEP} \lambda_i \quad on \quad \Gamma^{GROUND}$$

discr

discr

Discrete forward model

$$C^{k+1} = N_{[t, t+\Delta t]} \circ C^k$$

$$N_{[t, t+\Delta t]} = T_{HOR}^{\Delta t} \circ T_{VERT}^{\Delta t} \circ R_{CHEM}^{\Delta t} \circ T_{VERT}^{\Delta t} \circ T_{HOR}^{\Delta t}$$

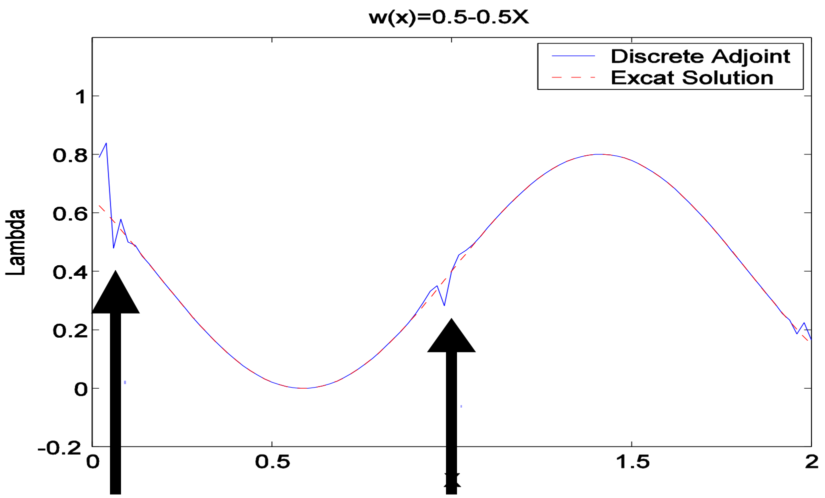
adj

Computational adjoint model

$$\lambda^k = N'^{*}_{[t, t+\Delta t]} \circ \lambda^{k+1} + \phi^{k+1}$$

$$N'^{*}_{[t, t+\Delta t]} = \left(T_{HOR}^{\Delta t}\right)'^{*} \circ \left(T_{VERT}^{\Delta t}\right)'^{*} \circ \left(R_{CHEM}^{\Delta t}\right)'^{*} \circ \left(T_{VERT}^{\Delta t}\right)'^{*} \circ \left(T_{HOR}^{\Delta t}\right)'^{*}$$

Discrete adjoints of advection numerical schemes can become inconsistent with the adjoint PDE



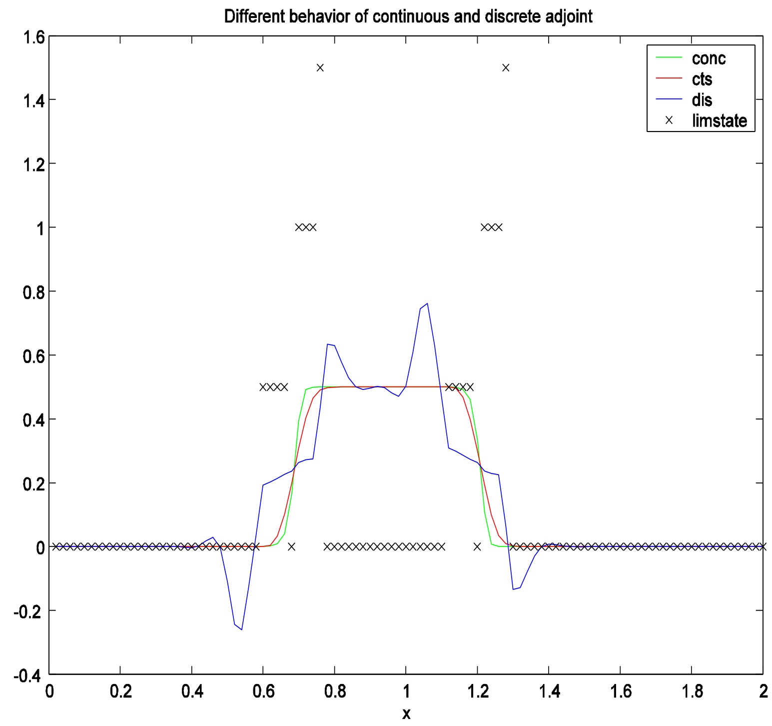
discretization change

upwind direction change

Change of forward scheme pattern:

- Change of upwinding
- Sources/sinks
- Inflow boundaries scheme

Example: 3rd order upwind FD



Active forward limiters act as pseudo-sources in adjoint

Example: vminmod

[Liu and Sandu, 2005]

Discrete Runge-Kutta adjoints can be regarded as "numerical methods" applied to the adjoint ODE

RK Method

$$\mathbf{y}^{\mathbf{n+1}} = \mathbf{y}^{\mathbf{n}} + h\sum_{i=1}^{s} b_i \mathbf{f}\left(\mathbf{Y}^{\mathbf{i}}\right),$$

$$\mathbf{Y}^{\mathbf{i}} = \mathbf{y}^{\mathbf{n}} + h\sum_{i=1}^{s} a_{i,j}\, \mathbf{f}\left(\mathbf{Y}^{\mathbf{j}}\right)$$

Discrete RK Adjoint

[Hager, 2000]

$$\lambda^{\mathbf{n}} = \lambda^{\mathbf{n+1}} + \sum_{\mathbf{i=1}}^{\mathbf{s}} \theta^{\mathbf{i}}$$

$$\theta^{\mathbf{i}} = h\,\mathbf{J}^{\mathrm{T}}\left(\mathbf{Y}^{\mathbf{i}}\right)\cdot\left[b_i\, \lambda^{\mathbf{n+1}} + \sum_{j=1}^{s} a_{j,i}\, \theta^{\mathbf{j}}\right]$$

Discrete Runge-Kutta adjoints: error analysis

Local error analysis: The discrete adjoint of RK method of order p ***is an order p*** *discretization of the adjoint equation. [Sandu, 2005]* . This:

- works for both explicit and implicit methods
- true for arbitrary orders p

Global error analysis: The discrete adjoint (of a RK method convergent with order p) ***converges with order p*** *to the solution of the adjoint ODE. [Sandu, 2005] The analysis accounts for:*

1. *the truncation error at each step, and*
2. *the different trajectories about which the continuous and the discrete adjoints are defined*

Stiff case: Consider a stiffly accurate Runge Kutta method **of order p** with invertible coefficient matrix A. The discrete adjoint provides:

1. an **order p** discretization of the adjoint of **nonstiff variable**
2. an **order min(p,q+1,r+1)** of the adjoint of **stiff variable**

[Sandu, 2005]

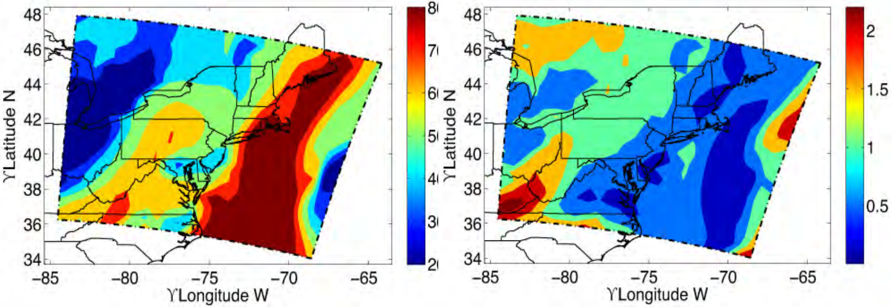
Properties of discrete adjoint LMM

1. For **fixed step sizes**
   - the discrete adjoint starting and ending steps, in general, are not consistent approximations of the adjoint ODE
   - the adjoint LMM is (at least) first order consistent with the adjoint ODE
2. For **variable step sizes** the adjoint LMM is not a consistent discretization of the adjoint ODE
3. The discrete **adjoint variable at the initial time** is an order p approximation of the continuous adjoint, where p is the order of the (forward) LMM method.

[Sandu, 2007]

Uncertainty quantification using polynomial chaos and the STEM model

Ground emissions	NO_x (NO, NO_2)	$\pm$20%
Ground emissions	$AVOC$ ($HCHO$, ALK, OLE, ARO)	$\pm$50%
Ground emissions	$BVOC$ ($ISOPRENE$, $TERPENE$, $ETHENE$)	$\pm$40%
Deposition velocity	O_3	$\pm$50%
Deposition velocity	NO_2	$\pm$50%
West Dirichlet B.C.	O_3	$\pm$5%
West Dirichlet B.C.	PAN	$\pm$5%



O_3 mean (48h)

O_3 standard dev. (48h)

Uncertainty apportionment with the STEM model

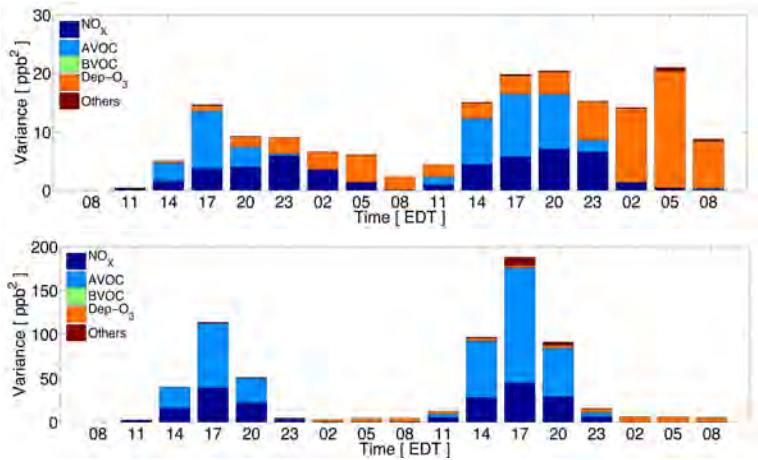


Figure: Top: New York. Bottom: Boston. 48 hrs ozone mean, standard deviation, and uncertainty (variance) apportionment.

Quantification of the probability of non-compliance with the NAAQS ozone maximum admissible levels

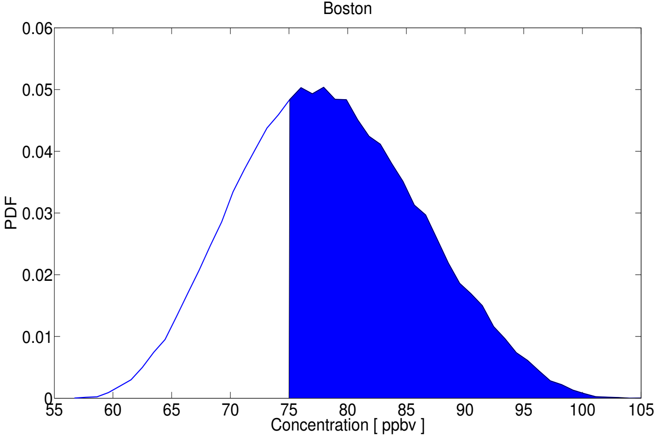
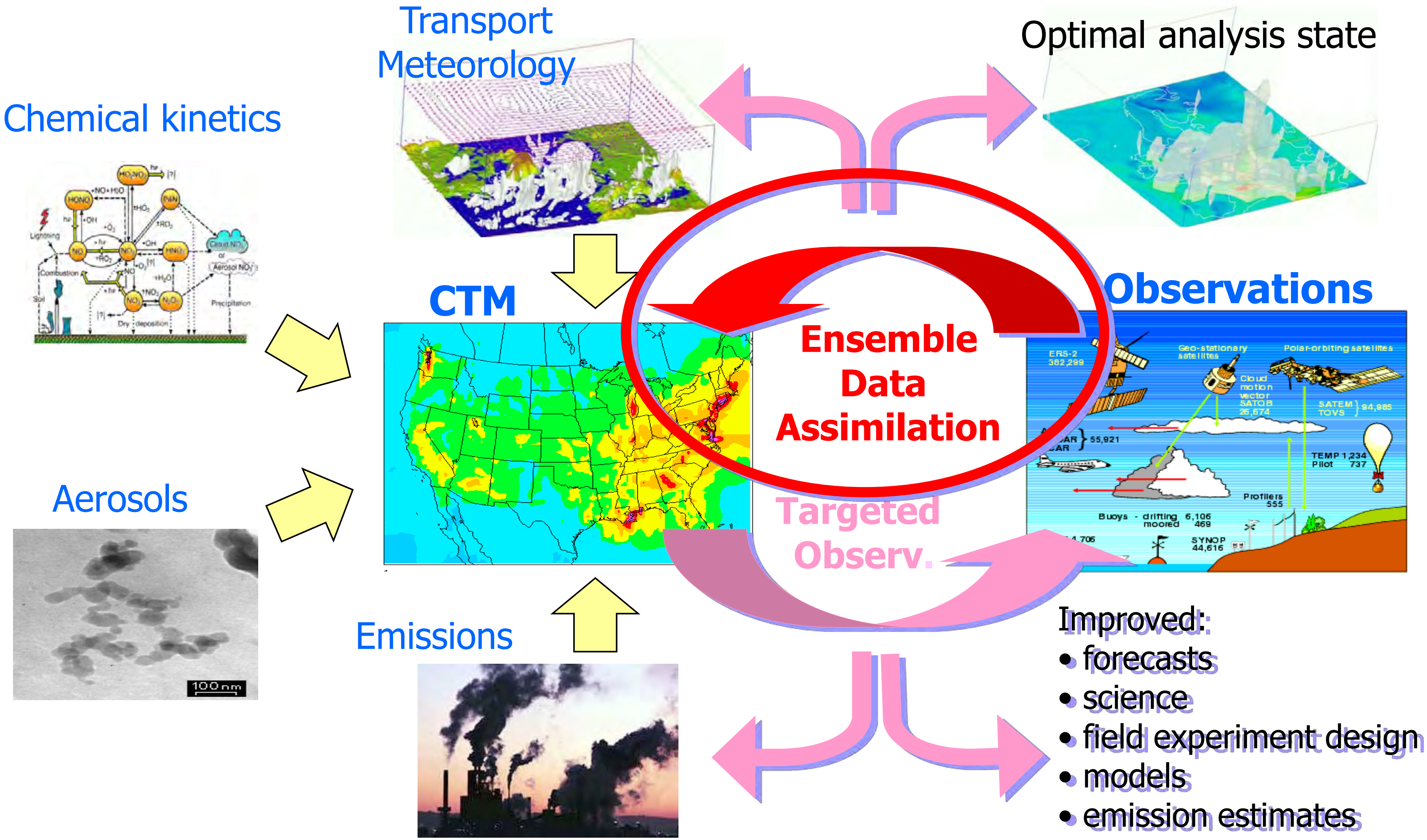


Figure: Boston 8hrs average ozone PDF shows a 68% probability of exceeding the maximum admissible level of 75 ppbv.

Ensemble-based chemical data assimilation is an alternative to variational techniques



The Ensemble Kalman Filter (EnKF) popular in NWP but not extensively used before with CTMs

$$\mathbf{y}_f^k = M\left(t^{k-1}, \mathbf{y}_a^{k-1}\right)$$

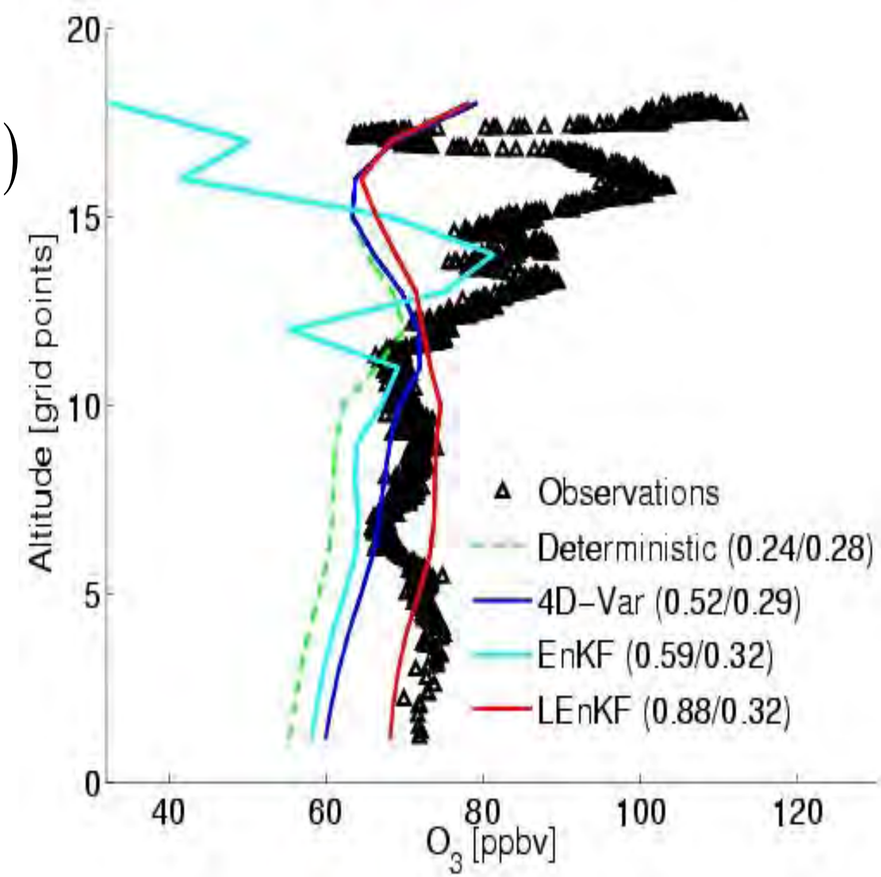
$$\mathbf{y}_a^k = \mathbf{y}_f^k + \mathbf{P}_f^k \mathbf{H}_k^T \left(\mathbf{R}_k + \mathbf{H}_k \mathbf{P}_f^k \mathbf{H}_k^T\right)^{-1} \left(\mathbf{z}_{obs}^k - \mathbf{H}_k \mathbf{y}_f^k\right)$$

Specify initial ensemble (sample B)

Covariance inflation: Prevents filter divergence (additive, multiplicative, model-specific)

Covariance localization (limit long-distance spurious correlations)

Correction localization (limit increments away from observations)



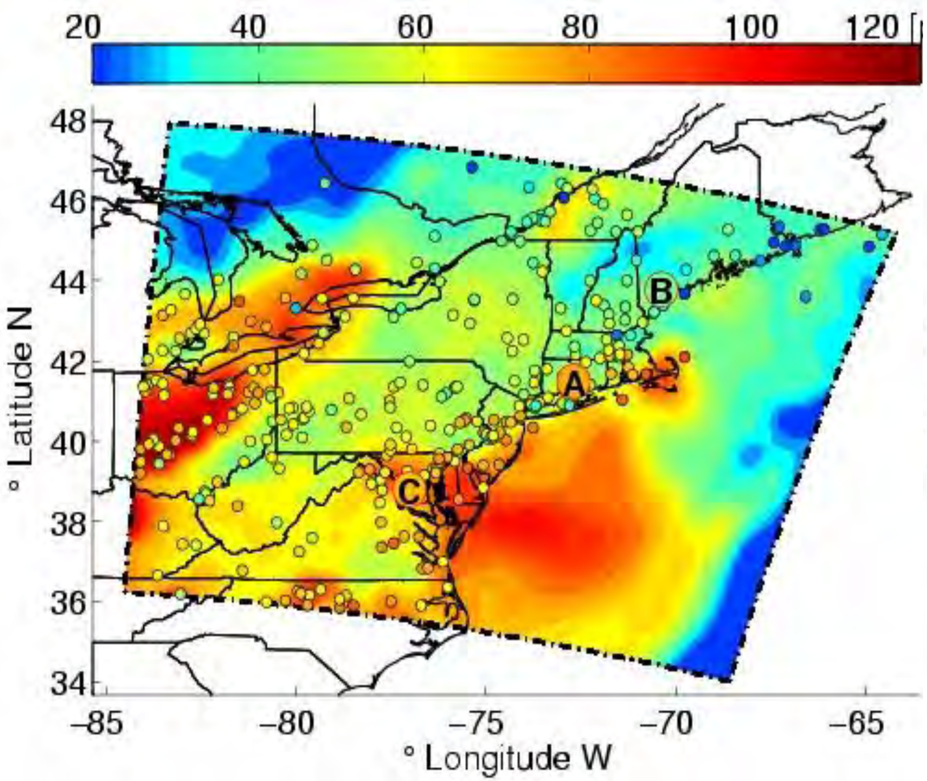
Ozonesonde S2 (18 EDT, July 20, 2004)

[Constantinescu et al., 2007]

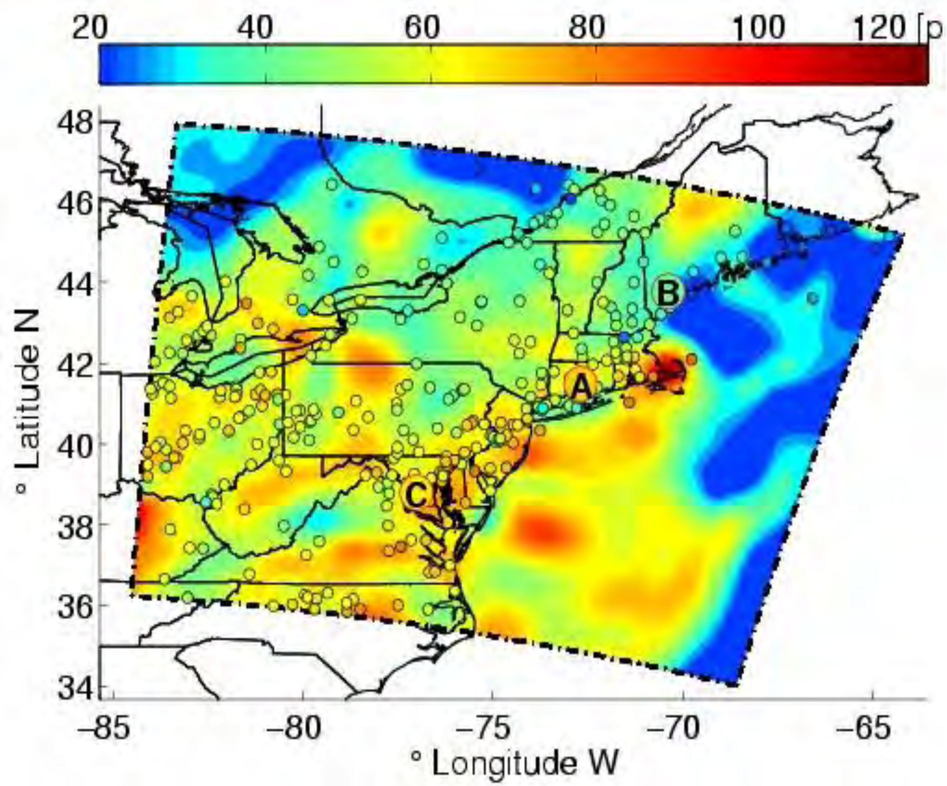
Ground level ozone at 2pm EDT, July 20, 2004, better matches observations after LEnKF data assimilation

Observations: circles, color coded by O_3 mixing ratio

Forecast (R^2=0.24/0.28)

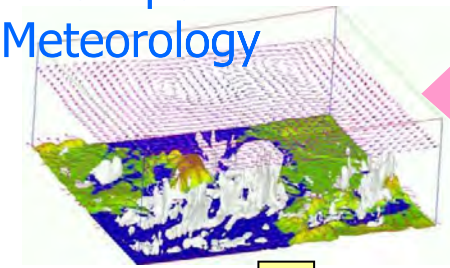


Analysis (R^2=0.88/0.32)

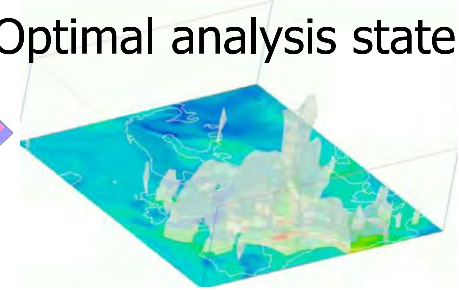


The use of adjoints in large scale simulations: atmospheric chemical transport models

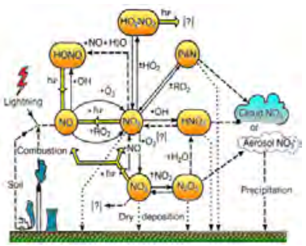
Transport
Meteorology



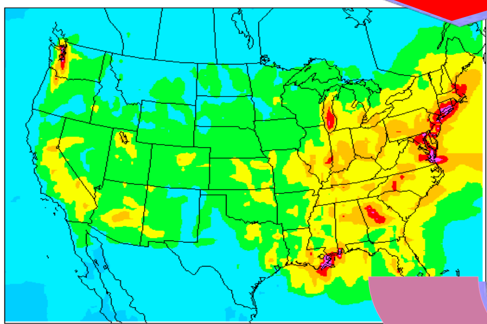
Optimal analysis state



Chemical kinetics



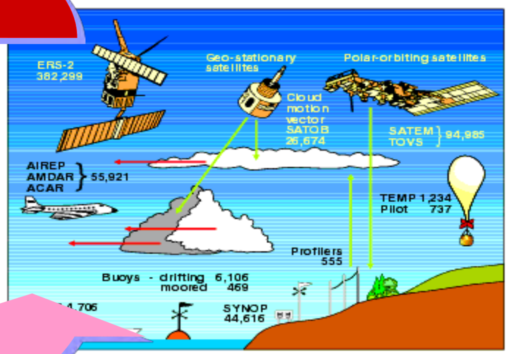
CTM



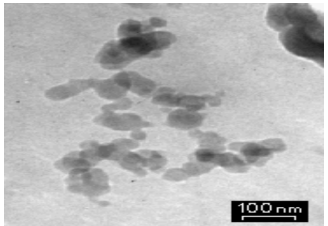
4D-Var Data Assimilation

Targeted Observ.

Observations



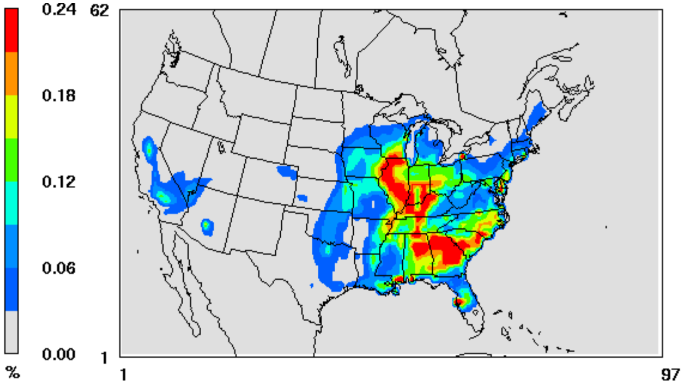
Aerosols



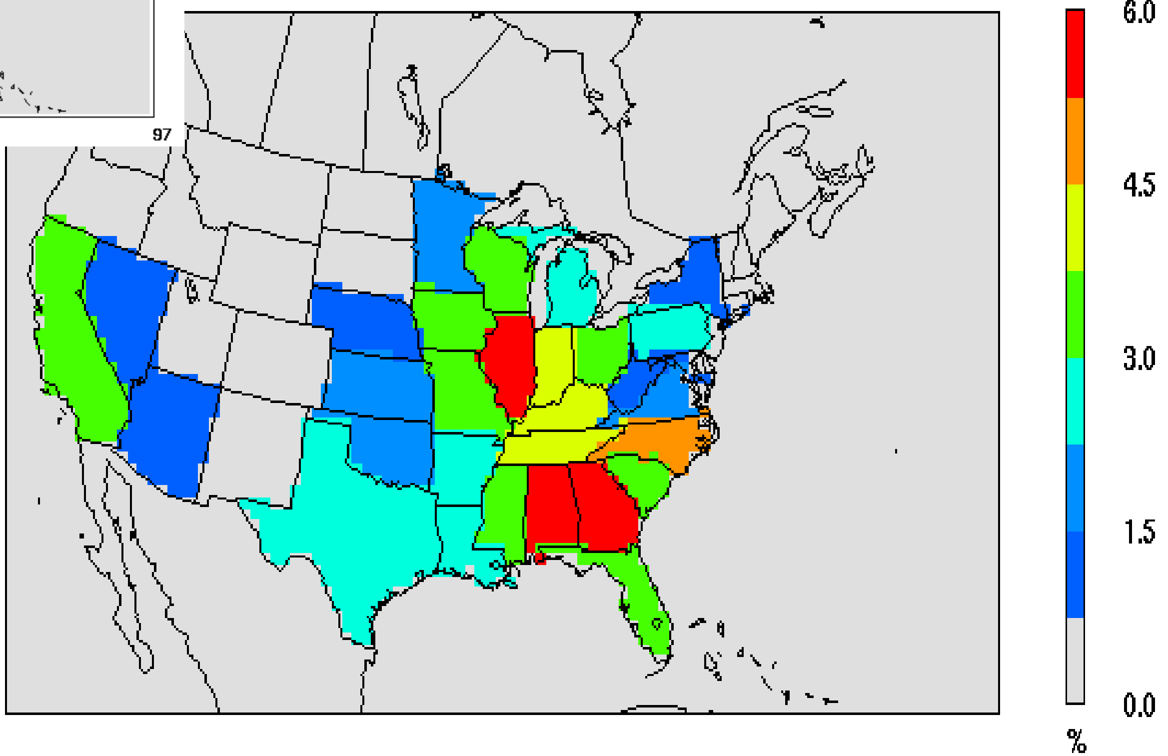
Emissions

Improved:
- forecasts
- science
- field experiment design
- models
- emission estimates

Adjoint sensitivity analysis of non-attainment metrics can help guide policy decisions



Estimated contributions by state to violating U.S. ozone NAAQS in July 2004

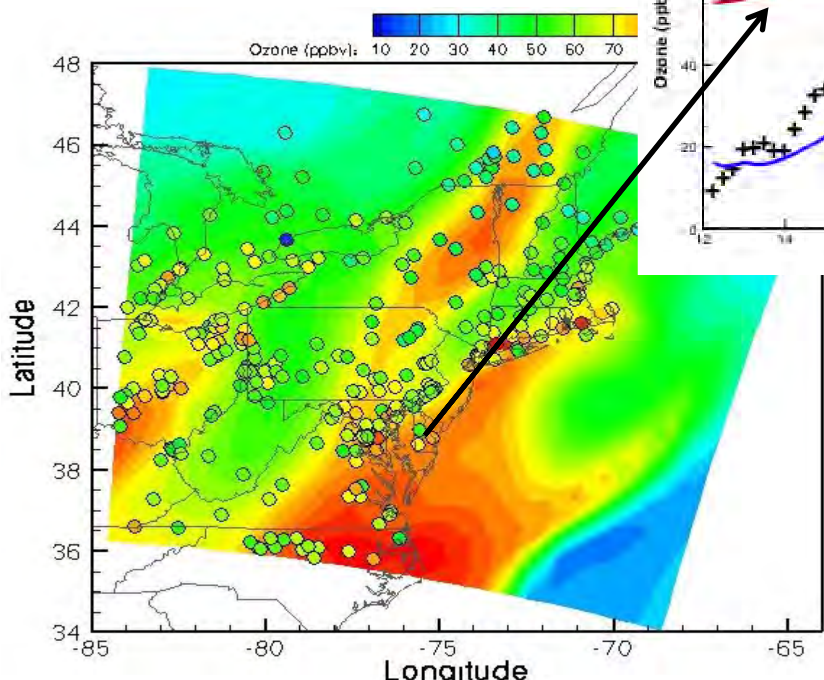


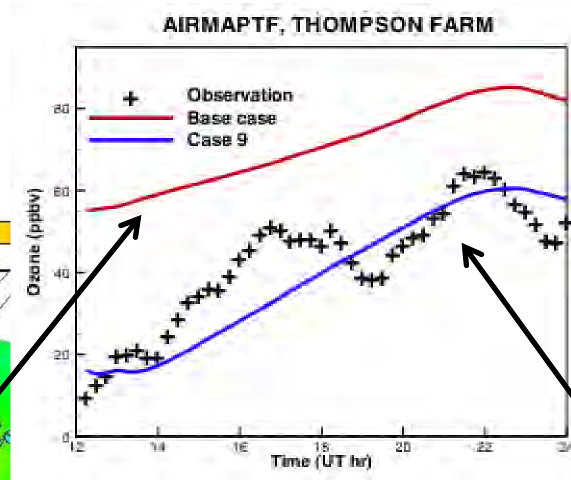
[Hakami et al., 2005]

STEM: Assimilation adjusts O_3 predictions considerably at 4pm EDT on July 20, 2004

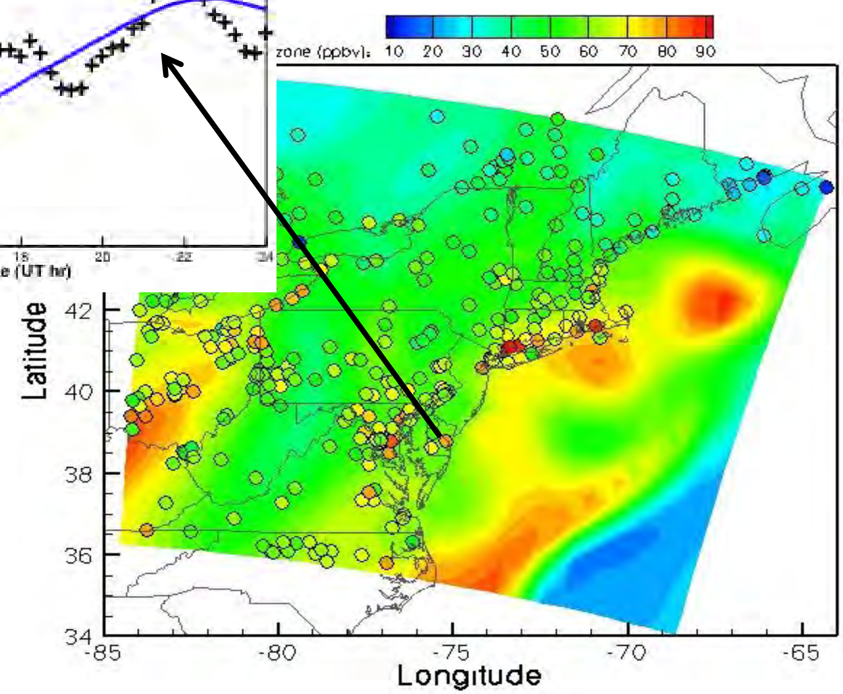
Observations: circles, color coded by O_3 mixing ratio

Ground O_3 (forecast)





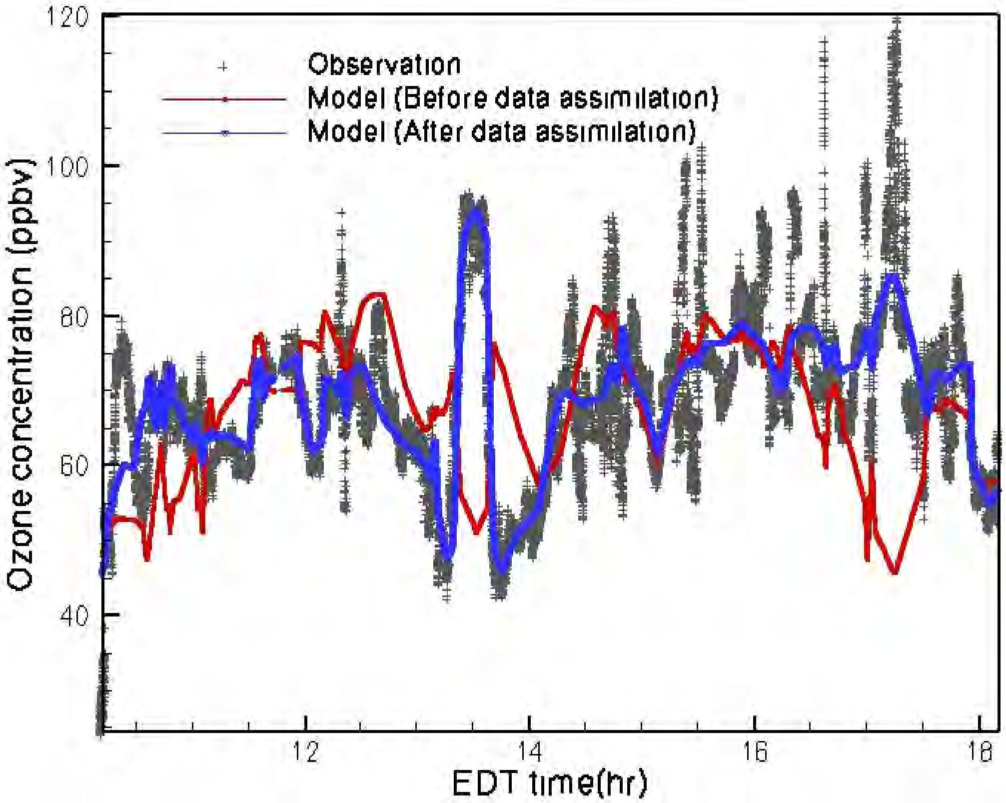
Ground O_3 (analysis)



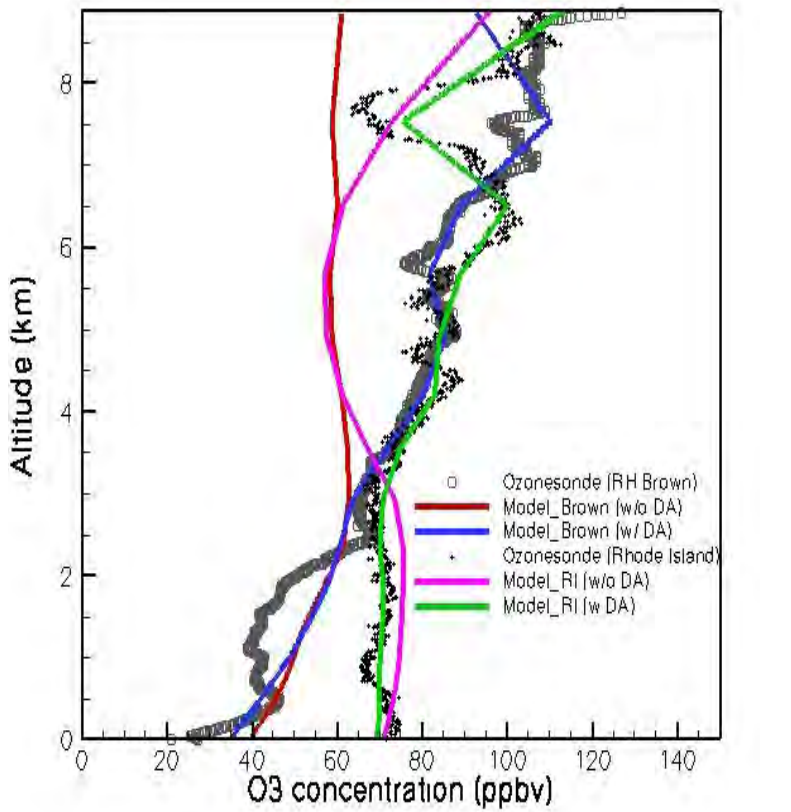
[Chai et al., 2006]

Assimilation of elevated observations for July 20, 2004

NOAA P3 flight observations



Ozonesonde observations (Rhode Island)

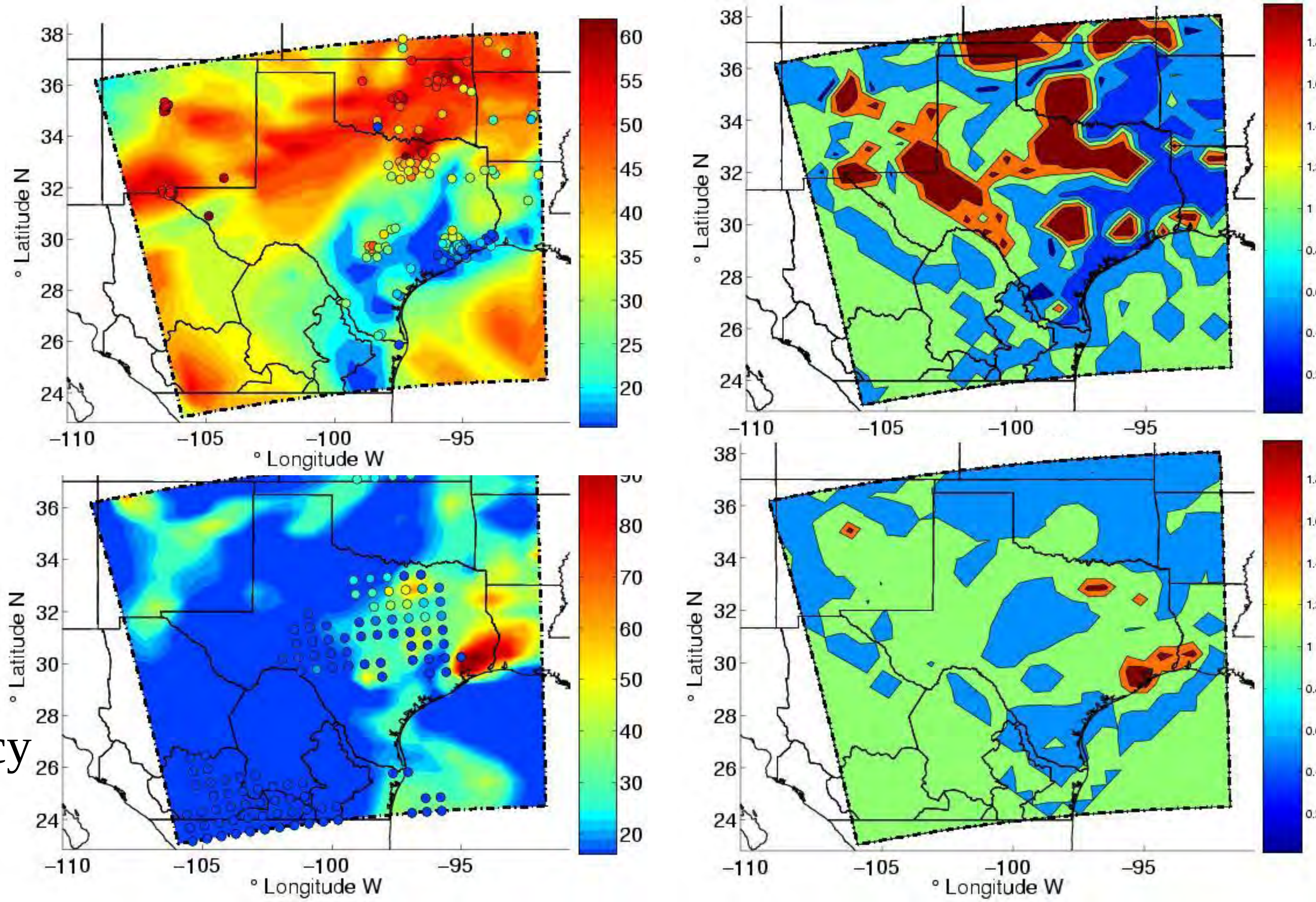


The inversion procedure can be extended to emissions, boundary conditions, etc.

Texas: 4am CST July 16 to 8pm CST on July 17, 2004.

O_3 AirNow

NO_2 emission corrections



NO_2 Schiamacy

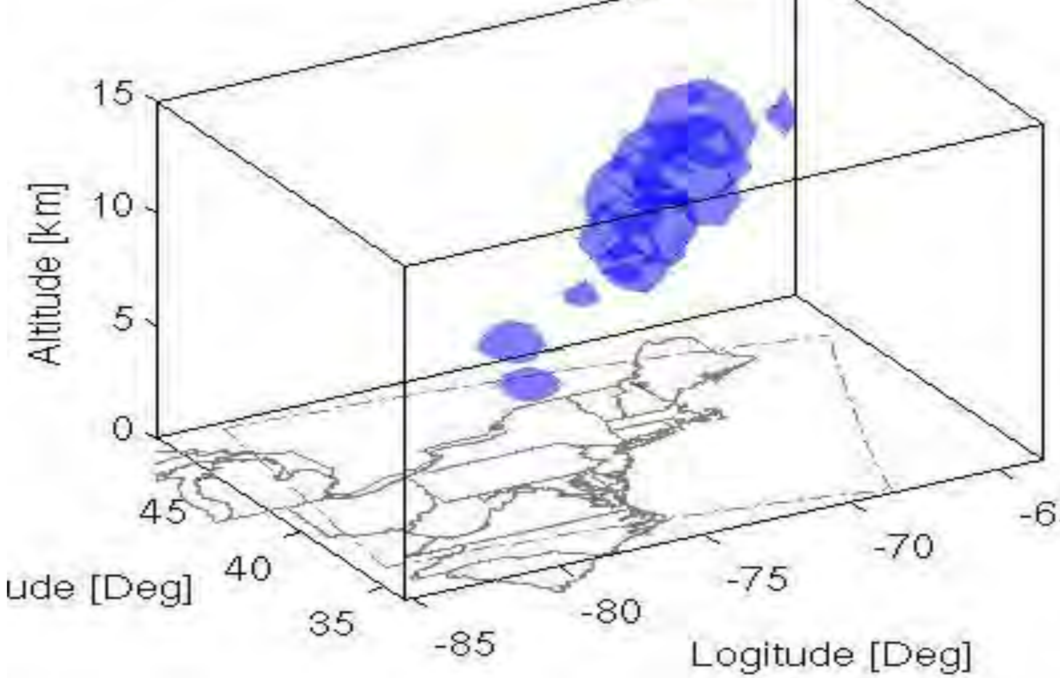
HCHO emission corrections

[Zhang, Sandu et al., 2006]

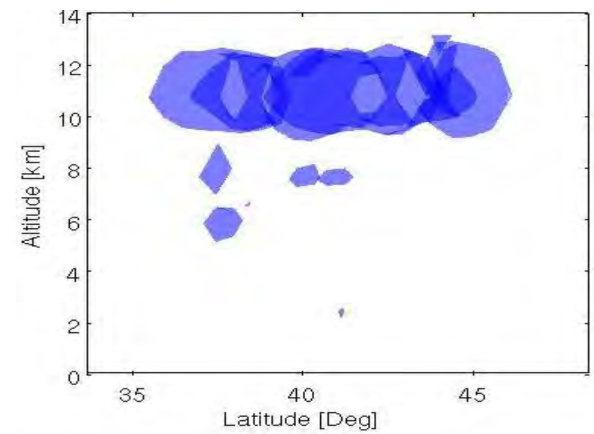
Smallest Hessian eigenvalues (vectors) approximate the principal aposteriori error components

$$\left(\nabla^2_{y^0,y^0}\Psi\right)^{-1} \approx \mathrm{cov}\left(y^{\mathrm{opt}}\right)$$

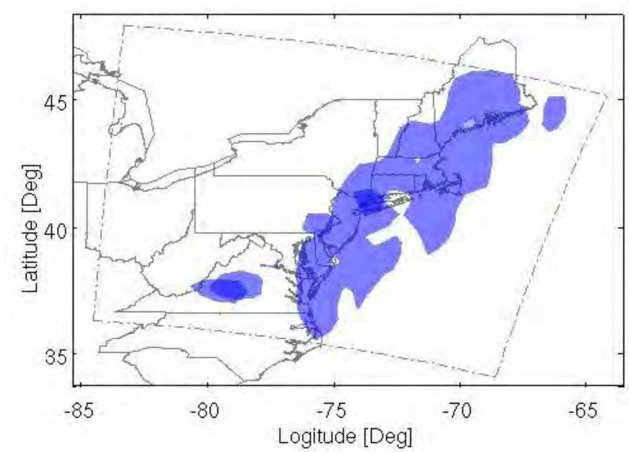
(a) 3D view (5ppb)



(b) East view



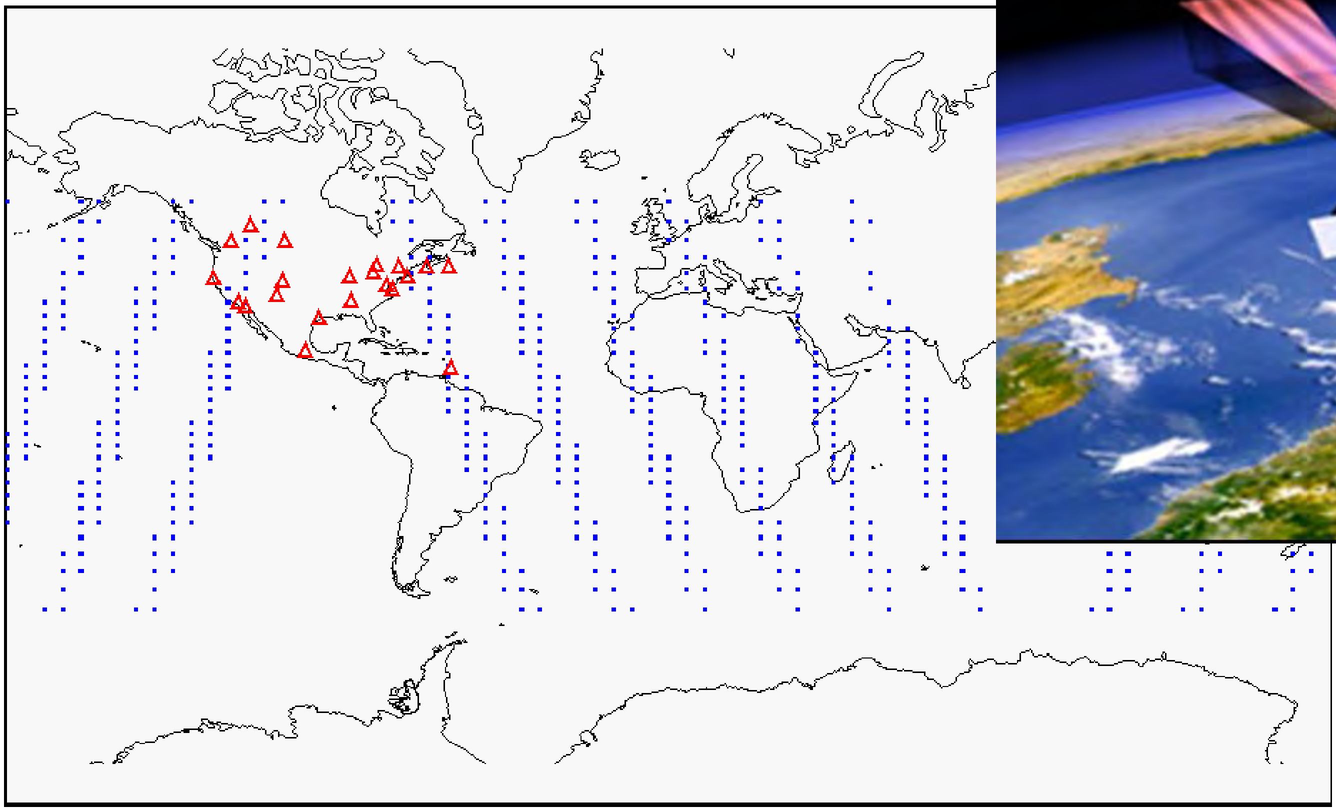
(c) Top view



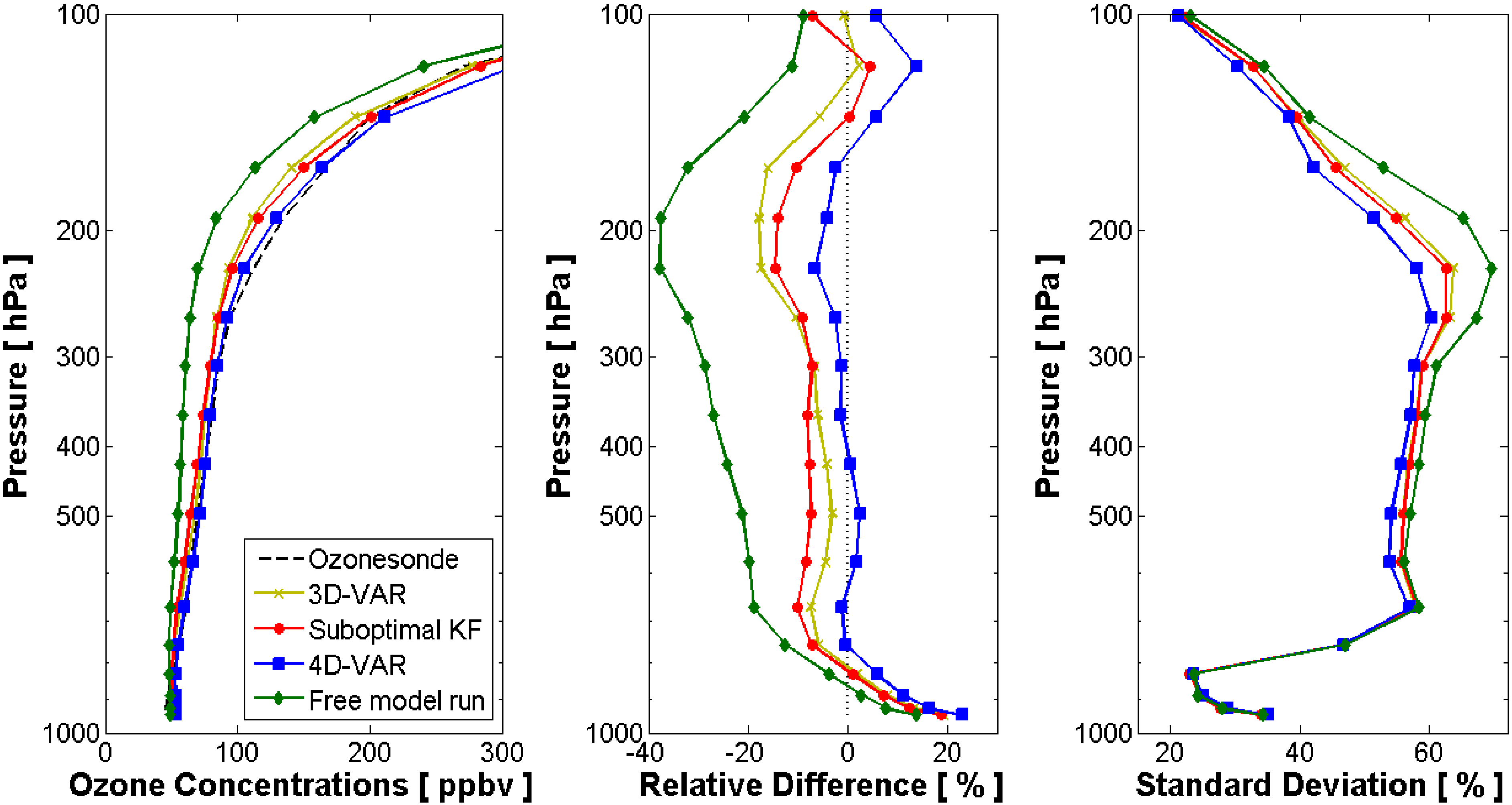
[Sandu et. al., 2007]

Assimilation of TES ozone column observations, August 2006. Lobatto-IIIC integrates stiff chemistry.

TES is one of four instruments on the NASA EOS Aura platform, launched July 14 2004



Quality of TES ozone column data assimilation results for several methods (August 1-15, 2006)



[Singh, Sandu et. al., 2010

Dynamic integration of chemical data and atmospheric models is an important, growing field

- the tools needed for 4D-Var chemical data assimilation are in place:
  - (adjoints for stiff systems, aerosols, transport; singular vectors, parallelization and multi-level checkpointing schemes, models of background errors)
- all algorithms are on a solid theoretical basis
- the ensemble filter methods show promise
- STEM, CMAQ, GEOS-CHEM have been endowed with data assimilation capabilities
- the tool strengths have been demonstrated using real (field campaign) data; ambitious science projects are ongoing