Ensemble-Based Chemical Data Assimilation

A. Sandu, W. Liao, E. Constantinescu: Virginia Tech
G.R. Carmichael, T. Chai: University of Iowa
J.H. Seinfeld: Caltech
D. Daescu: Portland State University
NASA, NOAA, EPA for Trace-P and ICARTT data

NSF ITR AP&IM-0205198
Motto

“Prediction is very difficult - especially about the future.”

Niels Bohr
Dynamic data driven simulations of atmospheric chemistry, particles, and transport are challenging.
Main focus of our ITR work is on 4d-Var chemical data assimilation (tool development and applications)

ICARTT field campaign, NE US, July 2004
Airnow, P3, ozonesonde observations
Assimilation improves simulations
Assimilation of AIRNOW O₃ surface observations adjusts predictions considerably over NE US

July 20, 2004, 16:00 EDT
Observations: circles, color coded by O₃ mixing ratio

Ground O₃ (model only)

Ground O₃ (analysis)
Assimilation of elevated observations for July 20, 2004 leads to improved predictions

NOAA P3 flight observations

Ozonesonde observations (Brown, Rhode Island)
Ensemble-based chemical data assimilation techniques can complement the variational tools

**Motivation:**
- Ensemble-based d.a. generate a statistical sample of analyses
- Explicitly propagate (approximations of) the error statistics
- Can deal effectively with nonlinear dynamics
- Complement variational techniques

**Issues:**
- Initialization of the ensemble
- Rank-deficient covariance matrix

**Contributions:**
- Models of background error covariance
- Calculation of TESVs for reactive flows
- Targeted observations using TESVs
- Ensemble-based assimilation results
AR model of background errors accounts for flow-dependent correlations and is inexpensive

- Background error representation considerably impacts the assimilation results
- Typically estimated empirically from multiple model runs (NMC)
- “Correct” mathematical models of background errors are of great interest

\[
\delta y' = M' \delta y
\]
\[
(l - \Delta t \, M')^N \delta y = \xi
\]
\[
B^{-1} = \langle \delta y \, \delta y^T \rangle^{-1}
\]
\[
= (l - \Delta t \, M' \, * )^N (l - \Delta t \, M')^N
\]

- “Monotonic TLM discretization”
- AR model of background errors
- \( N\Delta t \approx \) lifetime of the species
- \( B \) is flow dep., cheap, full rank
4d-Var data assimilation results are visibly improved when using the new AR background covariance

Observation error 8%; I.C. error 10ppbv; Initial ozone is control

Tacoma, DC

McMillian Reservoir, DC
Singular vectors characterize the directions of maximum error growth (under different measures)

TESVs are obtained as solutions of generalized eigenvalue problem:

\[ \sigma^2 = \max \frac{\langle \delta y(t^f), E \delta y(t^f) \rangle}{\langle \delta y(t^0), C \delta y(t^0) \rangle} = \max \frac{\langle \delta y(t^0), M^* E M' \delta y(t^0) \rangle}{\langle \delta y(t^0), C \delta y(t^0) \rangle} \]

\[ \Rightarrow M^* E M' s_k = \sigma_k^2 C s_k \]

E=C=[I,0] \rightarrow Total energy singular vectors (TESV)
E=I, C=P_a^{-1}=0.5 \partial^2 J/\partial y^2 \rightarrow Hessian singular vectors (HSV)

**TESV/ HSV are useful to:**
- Understand areas of maximum sensitivity to uncertainty
- Construct the subspace of initial perturbations for EnKF
Computation of TESVs for stiff systems raises unexpected challenges

\[ M'^* \cdot E \cdot M' \text{ symmetric?} \]

Singular perturbation analysis

\[
y' = f(y, z), \quad \varepsilon z' = g(y, z), \quad t^0 \leq t \leq t^f
\]

\[
\delta y' = f_y \delta y + f_z \delta z, \quad \varepsilon \delta z' = g_y \delta y + g_z \delta z
\]

\[
\Rightarrow \begin{bmatrix} \delta y \\ \delta z \end{bmatrix} \approx \begin{bmatrix} \delta y \\ -g_z^{-1} g_y \delta y \end{bmatrix} \quad \text{and} \quad \lambda_z \approx 0
\]

Solution: TLM + ADJ chem. projections

\[
M' = \prod \left( T'_{h/2} C_h P'T'_{h/2} \right)
\]

\[
M'^* = \prod \left( T'^*_{h/2} P'^*C'^*_{h} T'^*_{h/2} \right)
\]
With target on $O_3$ and $NO_2$ concentrations, SVs decrease rapidly and few modes capture uncertainty.

Trace-P, 03/01/2001
Dominant TESVs (O₃ sections). The target is O₃ and NO₂ concentrations in three countries

- Neg. isosurf: −1.4e+01; −9.6e+00; −4.8e+00;
  - Pos. isosurf: 1.5e+01; 2.9e+01; 4.4e+01;

- Neg. isosurf: −5.7e+00; −3.8e+00; −1.9e+00;
  - Pos. isosurf: 1.9e+01; 3.8e+01; 5.6e+01;

- Neg. isosurf: −3.9e+00; −2.6e+00; −1.3e+00;
  - Pos. isosurf: 5.3e+00; 1.1e+01; 1.6e+01;

- Neg. isosurf: −5.2e+01; −3.5e+01; −1.7e+01;
  - Pos. isosurf: 7.0e+00; 1.4e+01; 2.1e+01;
TESVs are shaped by both meteorology and chemistry, as seen for different sections.

**NO₂ sections of first 3 dominant TESVs**

- **Neg. isosurf:** $-4.3 \times 10^{-2}; -2.8 \times 10^{-2}; -1.4 \times 10^{-2}$;  
  - **Pos. isosurf:** $2.6 \times 10^{-3}; 5.1 \times 10^{-3}; 7.7 \times 10^{-3}$;

- **Neg. isosurf:** $-7.9 \times 10^{-2}; -5.2 \times 10^{-2}; -2.6 \times 10^{-2}$;  
  - **Pos. isosurf:** $3.0 \times 10^{-3}; 5.9 \times 10^{-3}; 8.9 \times 10^{-3}$;

- **Neg. isosurf:** $-9.1 \times 10^{-3}; -6.1 \times 10^{-3}; -3.0 \times 10^{-3}$;  
  - **Pos. isosurf:** $3.9 \times 10^{-2}; 7.9 \times 10^{-2}; 1.2 \times 10^{-1}$;

**HCHO sections of first 3 dominant TESVs**

- **Neg. isosurf:** $-1.4 \times 10^{-3}; -9.3 \times 10^{-4}; -4.7 \times 10^{-4}$;  
  - **Pos. isosurf:** $2.2 \times 10^{-3}; 4.5 \times 10^{-3}; 6.7 \times 10^{-3}$;

- **Neg. isosurf:** $-1.1 \times 10^{-3}; -7.3 \times 10^{-4}; -3.6 \times 10^{-4}$;  
  - **Pos. isosurf:** $3.1 \times 10^{-3}; 6.1 \times 10^{-3}; 9.2 \times 10^{-3}$;

- **Neg. isosurf:** $-1.5 \times 10^{-2}; -1.0 \times 10^{-2}; -5.0 \times 10^{-3}$;  
  - **Pos. isosurf:** $2.4 \times 10^{-4}; 4.9 \times 10^{-4}; 7.3 \times 10^{-4}$;
TESVs are useful to target observations to areas that provide most information.

\[ T = \sum_{k \geq 1} \frac{\sigma_k}{\sigma_{\text{max}}} S_k \]  
(measure of influence)
Experimental setting of the ensemble-based data assimilation system

- 50 members, perturbed I.C., B.C., and emissions
- 30% initial std, AR correlations + TESV perturbations
- \( \text{O}_3 \) and \( \text{NO}_2 \) observations at 24 ground locations in 3 countries, and in one vertical column. Perturbation 0.1% std, uncorrelated
- Quality of analysis in a sub-domain including observation sites
Ensemble data assimilation is effective when the initial ensemble is based on TESV perturbations

50 members (TESV+bckg), 24 hrs., 30% initial std, 24 ground, 1 column O₃+NO₂ obs. sites, 0.1% obs. std.
Ensemble data assimilation improves the prediction of species which are not directly observed

50 members (TESV+bckg), 24 hrs., 30% initial std, 24 ground, 1 column $O_3+NO_2$ obs. sites, 0.1% obs. std.
Comparison of the solutions against the reference show marked improvements after assimilation.
Results for species not directly observed also show marked improvements after assimilation.
Dynamic integration of chemical data and atmospheric models is an important, growing field

- Assimilation of real-time chemical observations into CTMs:
  - improves reanalysis of fields and model forecast skills
  - provides top-down estimate of emission inventories
- During this ITR project we have:
  - developed the tools needed for 4d-Var chemical data assimilation
  - demonstrated them using real (field campaign) data
- This presentation focuses on new ensemble-based tools:
  - AR models for background errors
  - calculation of TESVs for stiff systems
  - demonstration of targeted obs. and ensemble data ass. on Trace-P
- Current and future work includes:
  1. hybrid methods (combining ensemble and variational approaches)
  2. second order adjoints and optimization and reduced order models
  3. close the feedback loop by targeted observations
  4. improved estimates of emission inventories
Ensemble bias and standard deviation

**O<sub>3</sub> bias**

**O<sub>3</sub> std**

**NO<sub>2</sub> bias**

**NO<sub>2</sub> std**
Ensemble bias and standard deviation

PAN bias (%)

HCHO bias (%)

PAN std (ppbv)

HCHO std (ppbv)
4D-Var applications during ICARTT

Domain: D02
(~300 stations)

Grid size: $25 \times 22 \times 20$
(60km $\times$ 60km)

DA window: 24 hrs
(0~23 EDT)

Forecasting: 72 hrs
(easy to change)
Atmospheric chemical data assimilation

Chemical model

#EQUATIONS { Small Stratospheric }
O2 + hv = 2O : 2.6E-10*S;
O + O2 = O3 : 8.0E-17;
O3 + hv = O + O2 : 6.1E-04*S;
O + O3 = 2O2 : 1.5E-15;
O3 + hv = O1D + O2 : 1.0E-03*S;
O1D + M = O + M : 7.1E-11;
O1D + O3 = 2O2 : 1.2E-10;
NO + O3 = NO2 + O2 : 6.0E-15;
NO2 + O = NO + O2 : 1.0E-11;
NO2 + hv = NO + O : 1.2E-02*S;

Aerosol model

Meteo model

CTM

Data Assimilation

Optimal analysis state

Observations

Forecast

Targeted observ.

Emissions

ICCS 2005 Dynamic Data Driven Application Systems
TESVs are shaped by both meteorology and chemistry

CO sections of first 3 dominant TESVs

- Neg. isosurf: $-2.7e+00; -1.8e+00; -9.0e-01$
- Pos. isosurf: $7.3e-01; 1.5e+00; 2.2e+00$

PAN sections of first 3 dominant TESVs

- Neg. isosurf: $-2.2e-03; -1.4e-03; -7.2e-04$
- Pos. isosurf: $1.7e-03; 3.3e-03; 5.0e-03$

- Neg. isosurf: $-3.2e+00; -2.1e+00; -1.1e+00$
- Pos. isosurf: $1.9e-01; 3.8e-01; 5.7e-01$

- Neg. isosurf: $-9.0e+00; -2.6e+00; -1.3e+00$
- Pos. isosurf: $1.5e+00; 3.1e+00; 4.6e+00$

- Neg. isosurf: $-3.9e+00; -2.6e+00; -1.3e+00$
- Pos. isosurf: $1.9e-01; 3.8e-01; 5.7e-01$

- Neg. isosurf: $-1.6e-02; -1.1e-02; -5.4e-03$
- Pos. isosurf: $1.9e-03; 3.8e-03; 5.7e-03$

- Neg. isosurf: $-8.3e-03; -5.5e-03; -2.8e-03$
- Pos. isosurf: $7.3e-04; 1.5e-03; 2.2e-03$
Our previous work focused on the variational approach for chemical data assimilation

- Discrete/continuous adjoint models, and analysis, for
  - stiff chemical systems
  - integral-partial-differential aerosol dynamic equations
  - upwind and slope/flux limited schemes for hyperbolic systems
  - second order adjoints and optimization algorithms
- Parallel adjoint STEM-III with 2 level, distributed checkpointing scheme
- Real applications with Trace-P, ICARTT data etc.