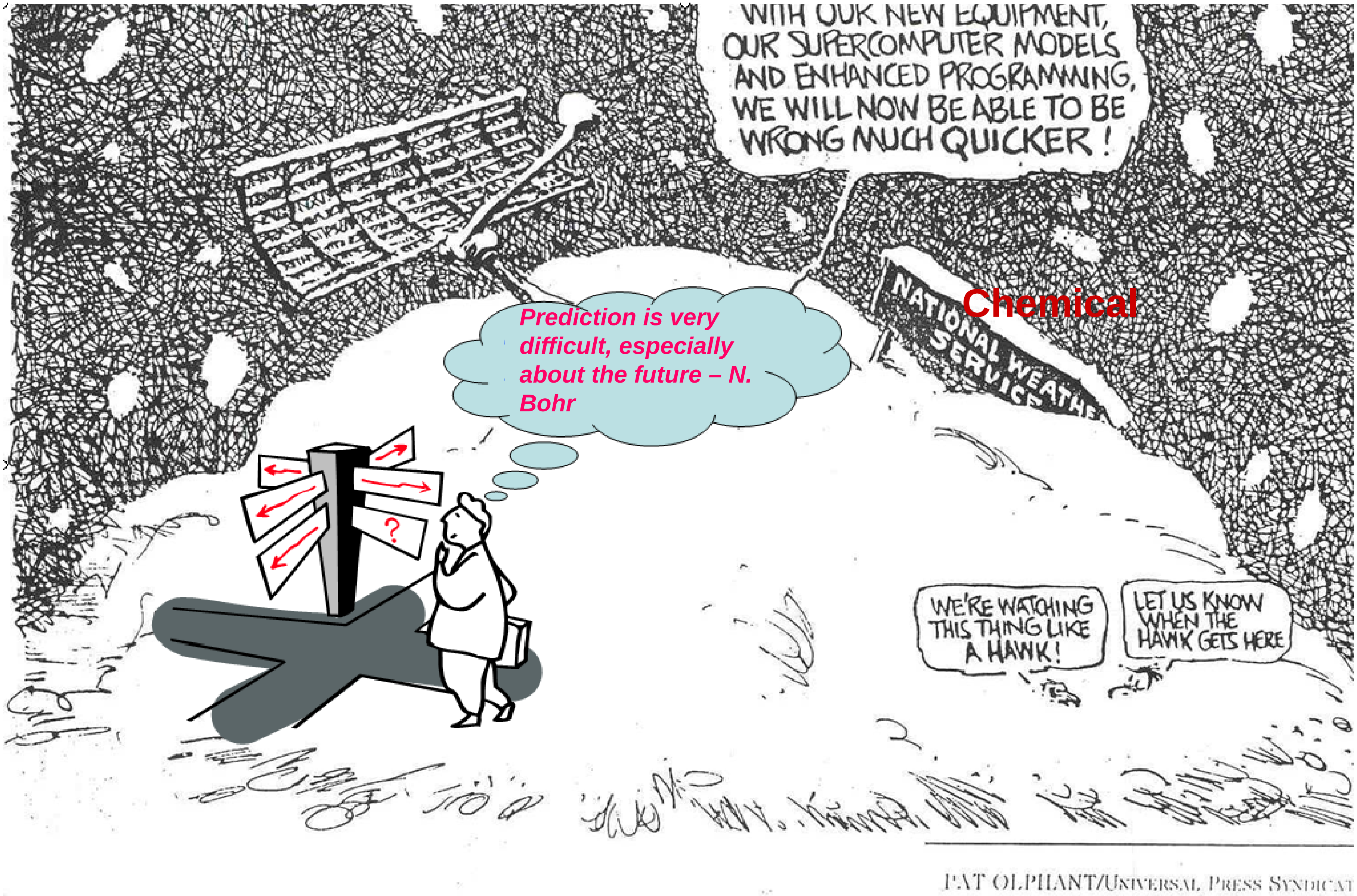
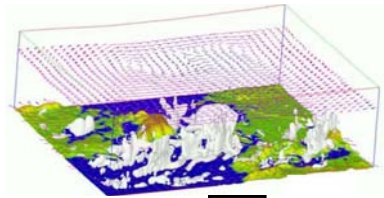
Chemical Weather Prediction

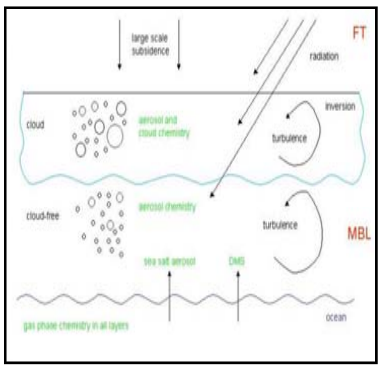


Air Quality Modeling: Improving Predictions of Air Quality (analysis and forecasting perspectives)

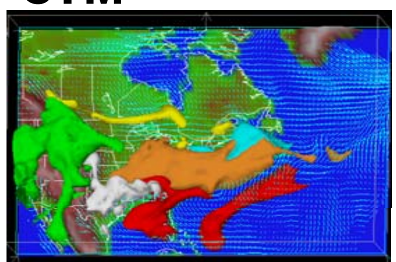
Met model



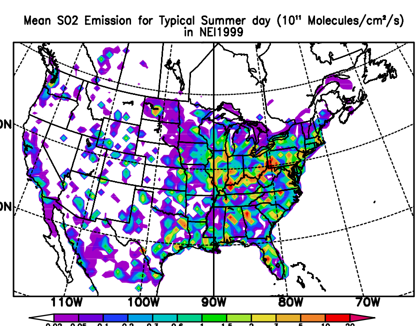
Chemical, Aerosol, Removal modules



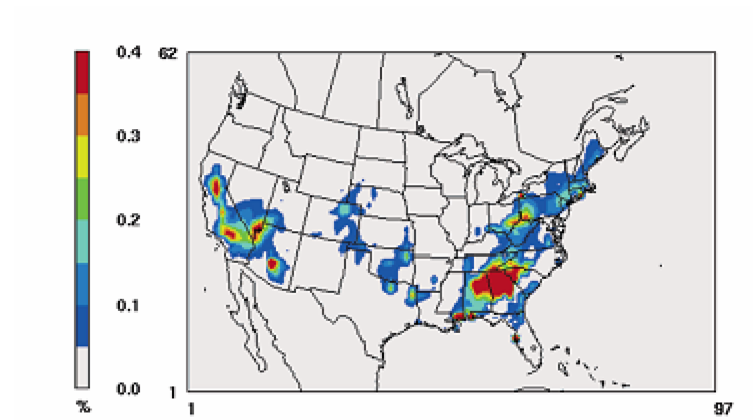
CTM



Emissions

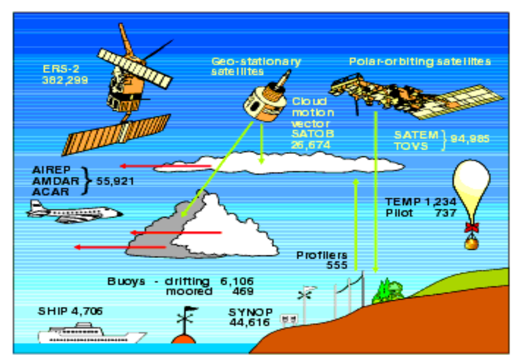


Predicted Quantity: e.g., *ozone AQ violation*



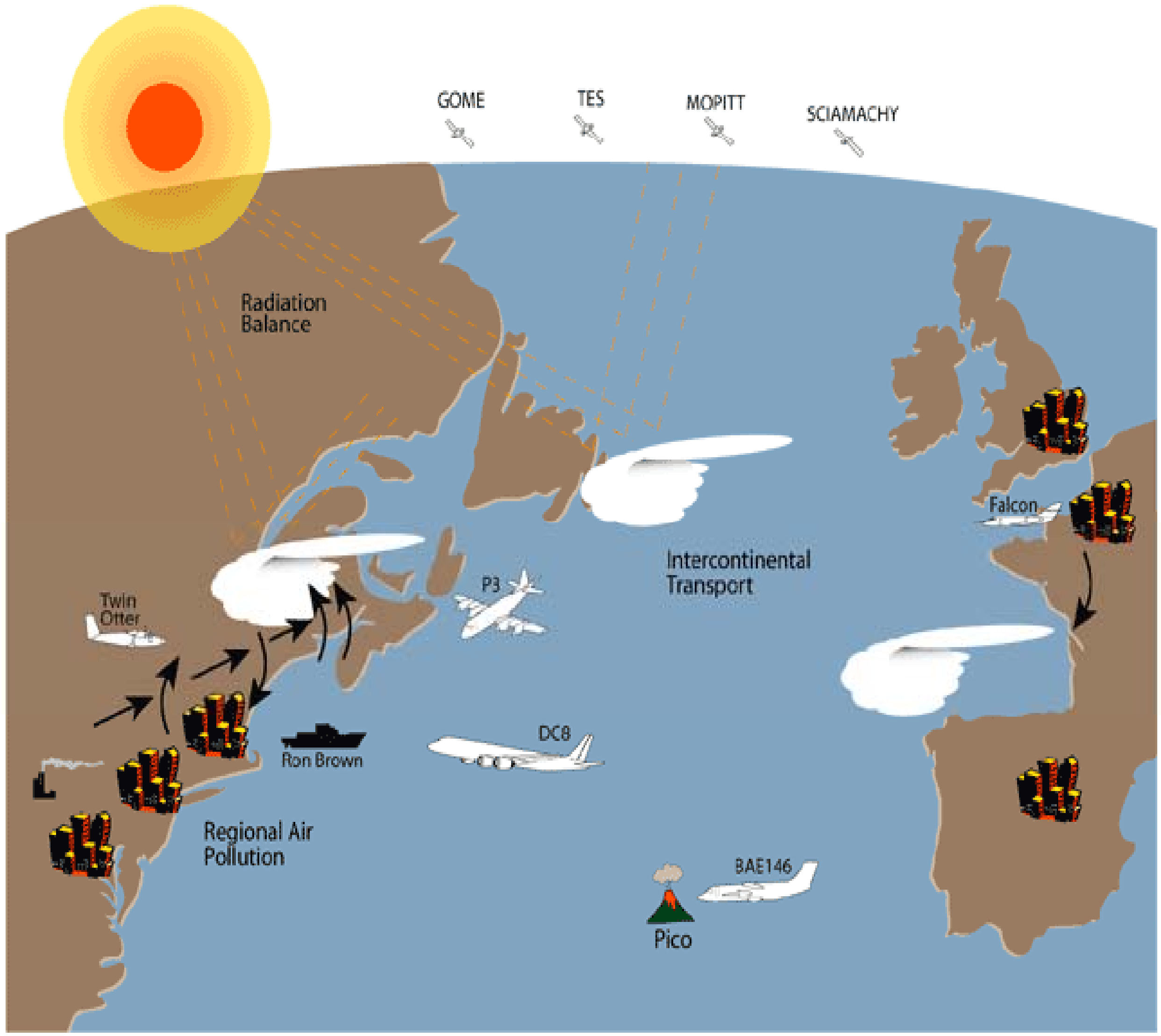
How confident are we in the models & predictions?

Observations

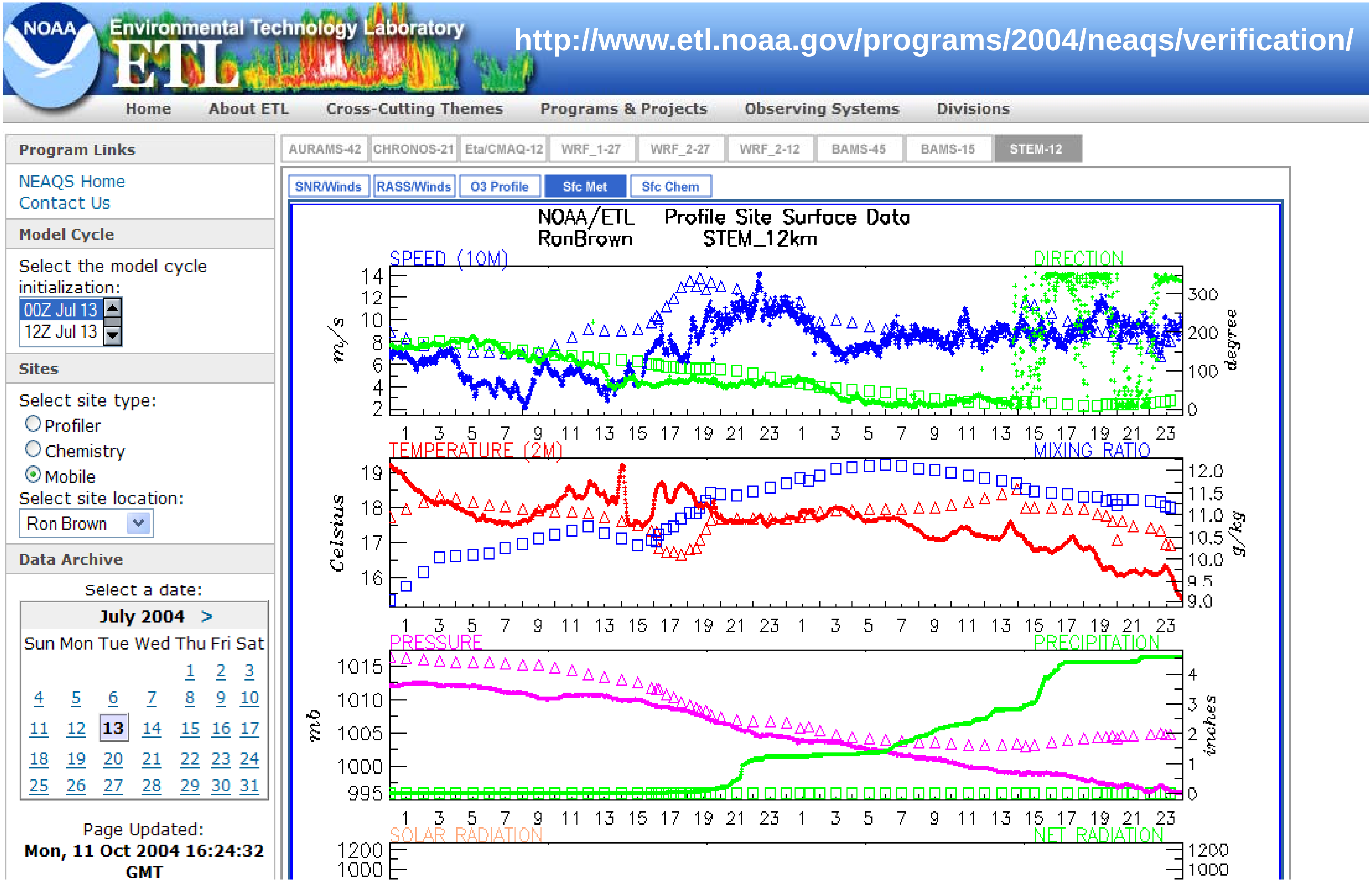


"Lessons" from the ICARTT Experiment

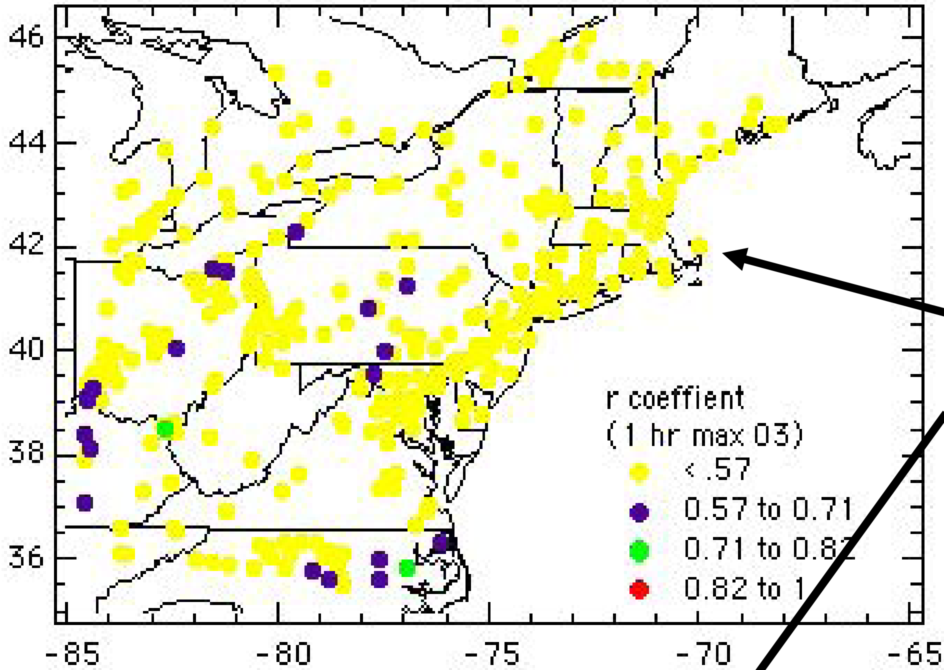
Experiments such as these employ mobile "Super-Sites" and study pollution outflow from source regions



Extensive Real-Time Evaluation of Regional Forecasts – *Stu McKeen*

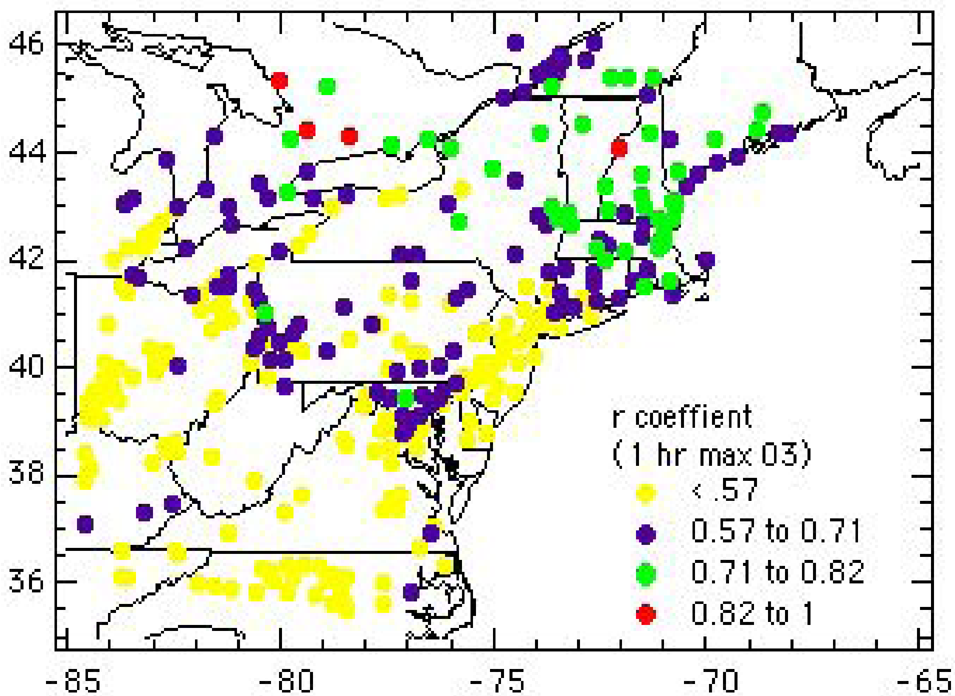


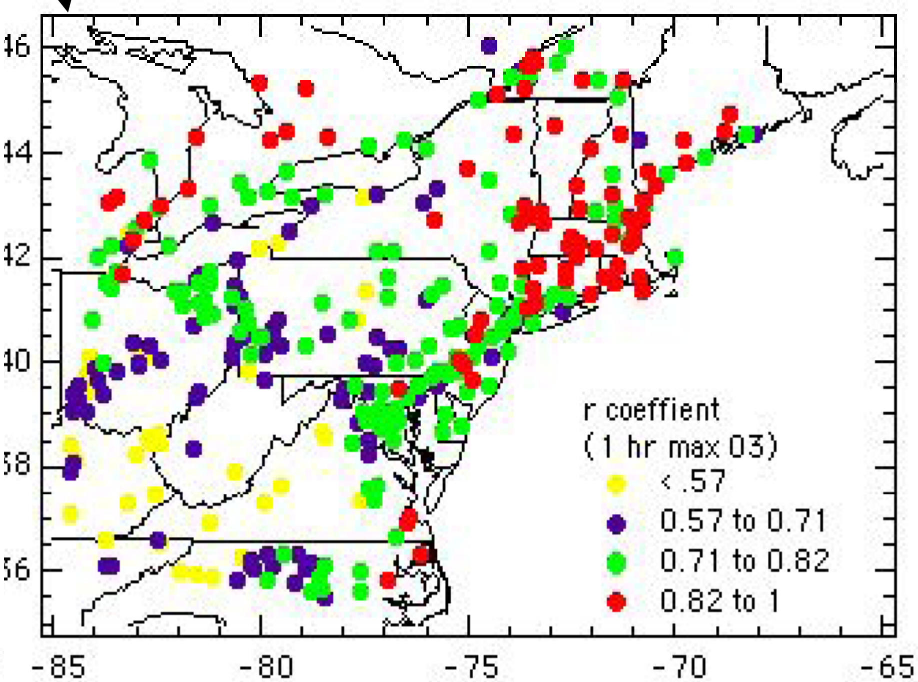
Forecasting Air Quality an Important Activity in Air Quality Management



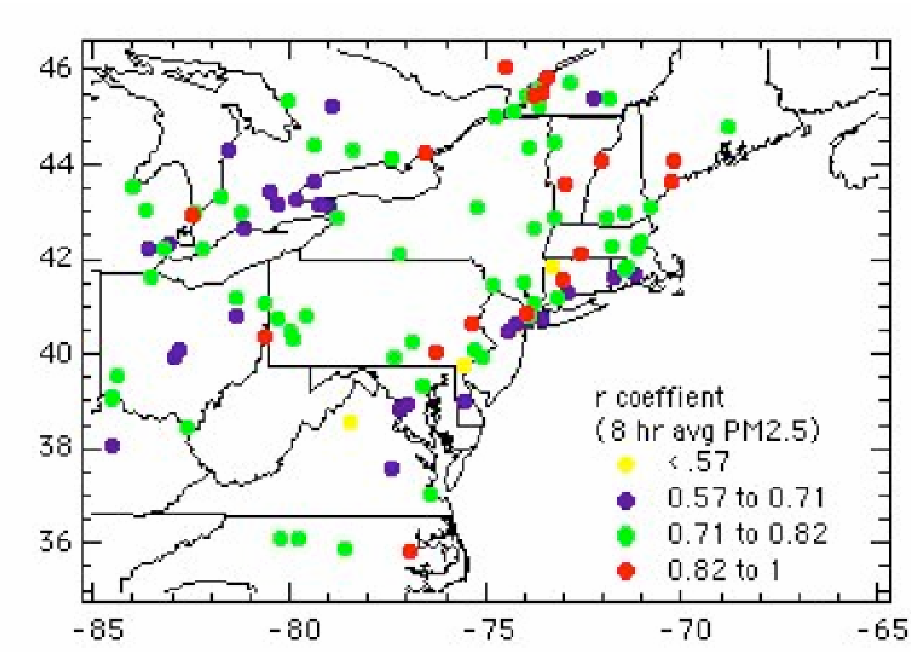
* Persistence
* Single Forward Model w/o assimilation
* Ensemble forecast (8 models) w/o assimilation (*further improvements with bias corrections based on obs*)

McKeen et al., JGR, 2005





Ensemble Methods Also Work for PM2.5 Forecasting



Comparison Statistics for
Geometric Ensemble with
AIRNOW daily 8-hr avg PM2.5
7/14/04 through 8/18/04

	median	average
r coeff.	0.75	0.73
Md/Ob ratio	0.86	1.00
ratio RMSE	1.76	1.90
Sdev	5.55	5.84
Skill factor(%) =		75.42

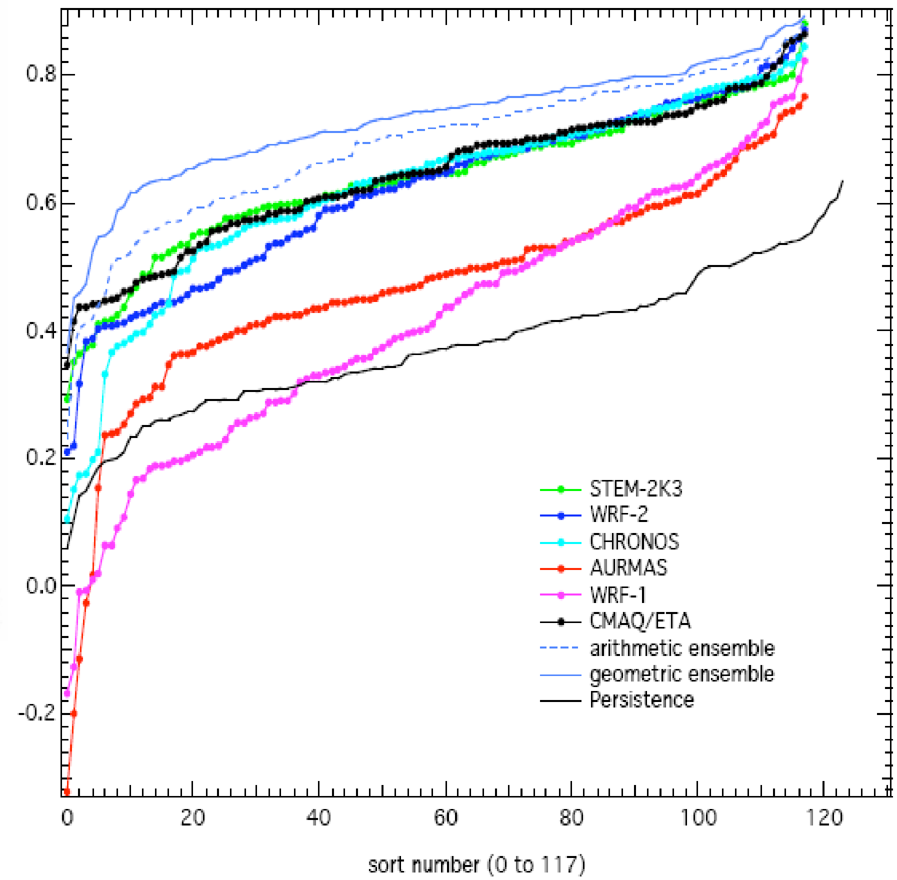
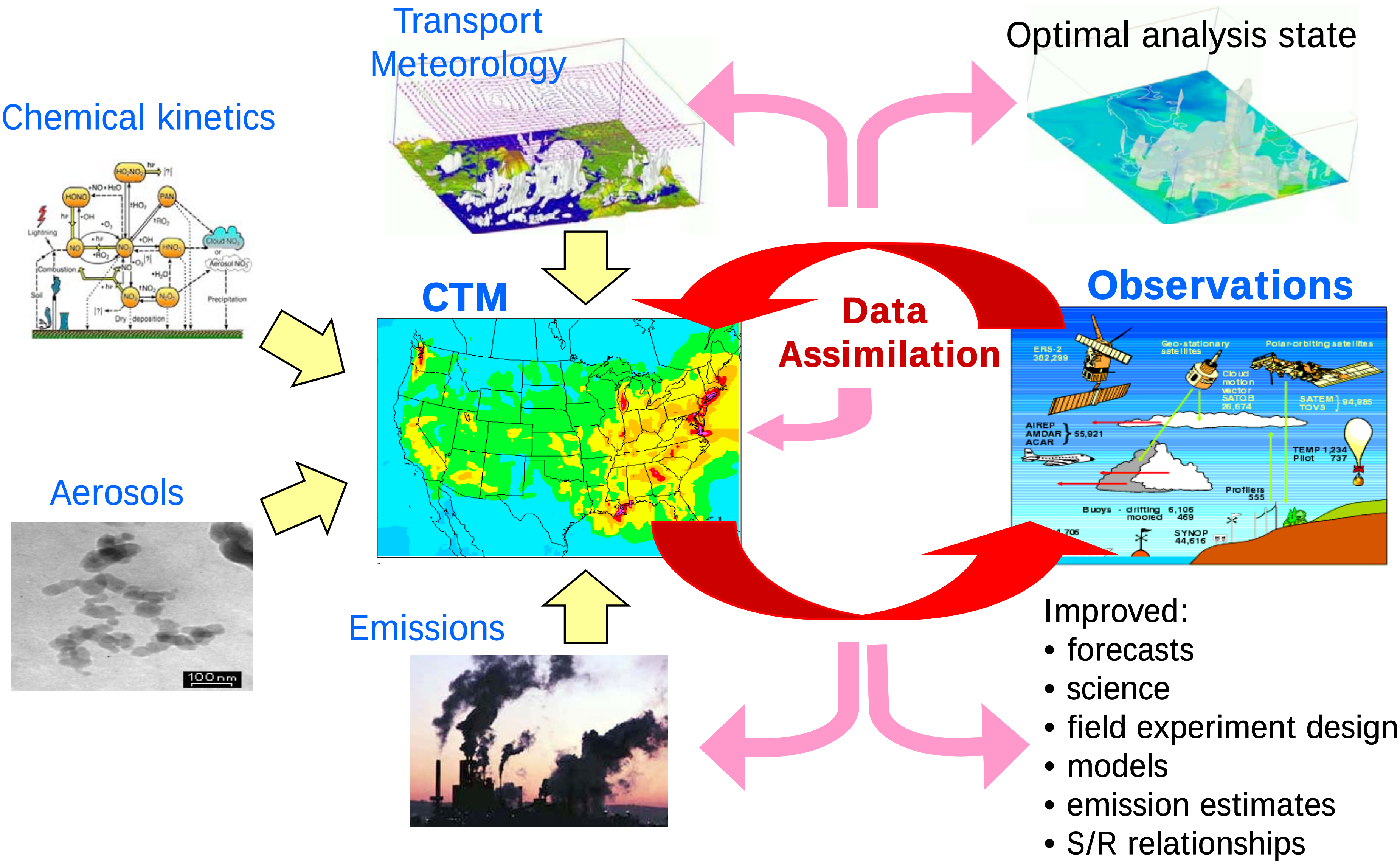


Figure 5. Sorted r-correlation coefficients for the 8 model cases, and persistence.

McKeen et al., JGR, 2007

Regional-Scale Chemical Analysis for Air Quality Modeling:

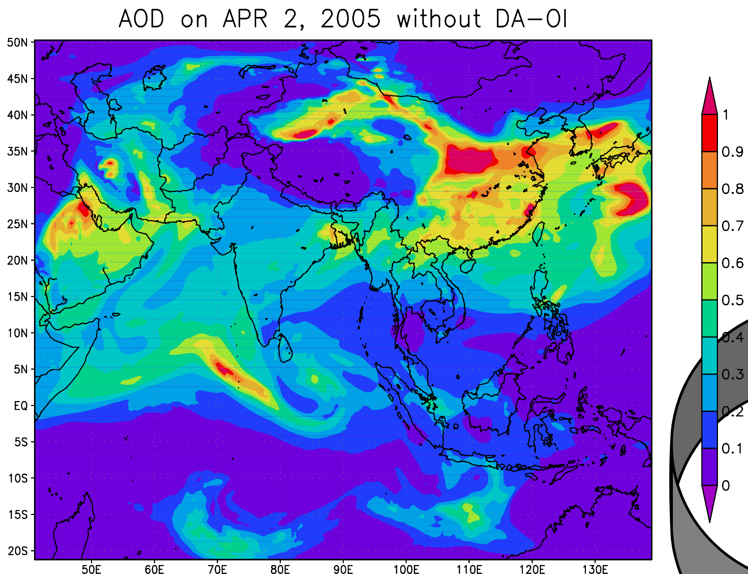
A Closer Integration Of Observations And Models



Data assimilation methods

- "Simple" data assimilation methods
  - Optimal Interpolation (OI)
  - 3-Dimensional Variational data assimilation (3D-Var)
  - Kriging
- Advanced data assimilation methods
  - 4-Dimensional Variational data assimilation (4D-Var)
  - Kalman Filter (KF) - Many variations, e.g. Ensemble Kalman Filter (EnFK)

Assimilation of MODIS AOD to Produce Constrained Fields for Climate Calculations



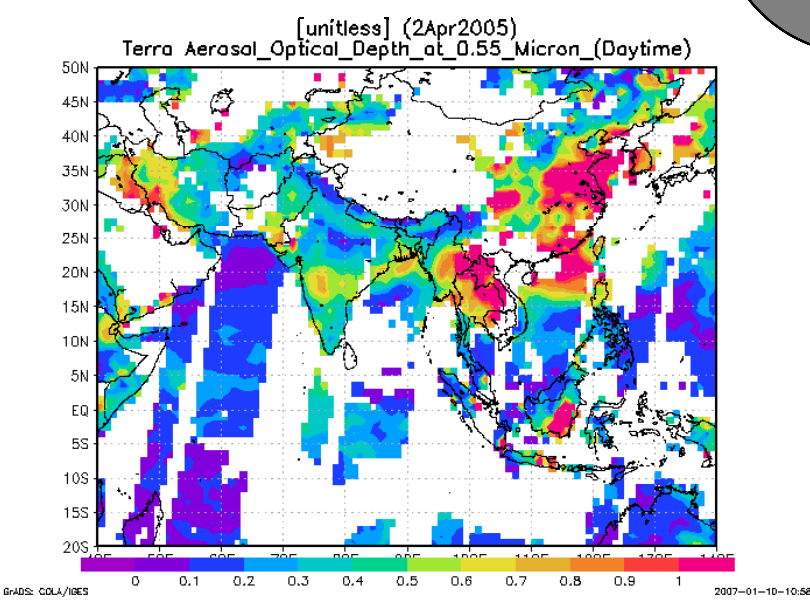
How to optimally adjust individual aerosol quantities given AOD (sulfate, BC, OC, dust, sea salt)?

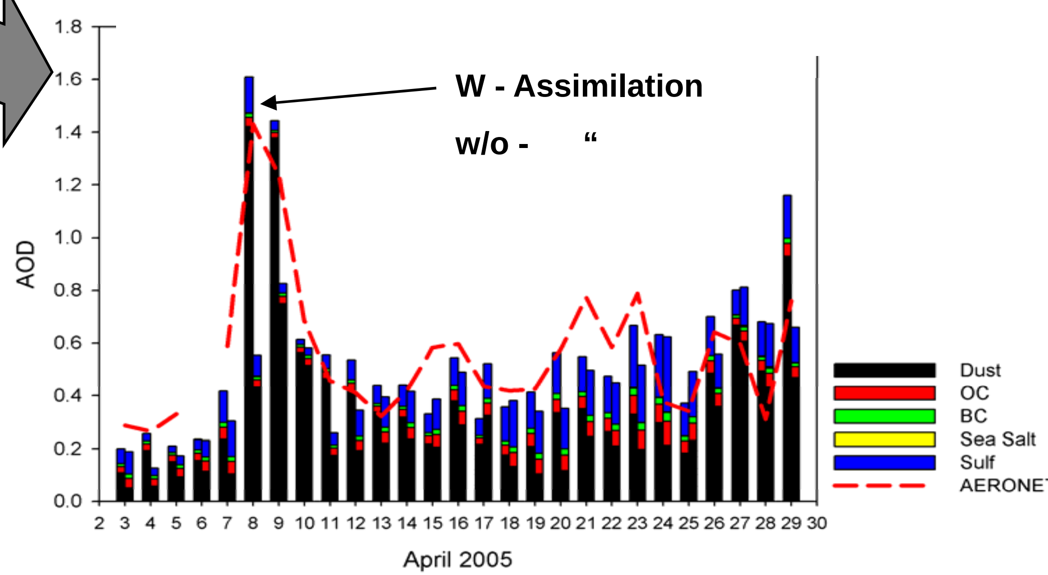
- AOD by itself not unique

- Fine mode fraction helps

- SSA gives info to adjust abs vs scat.

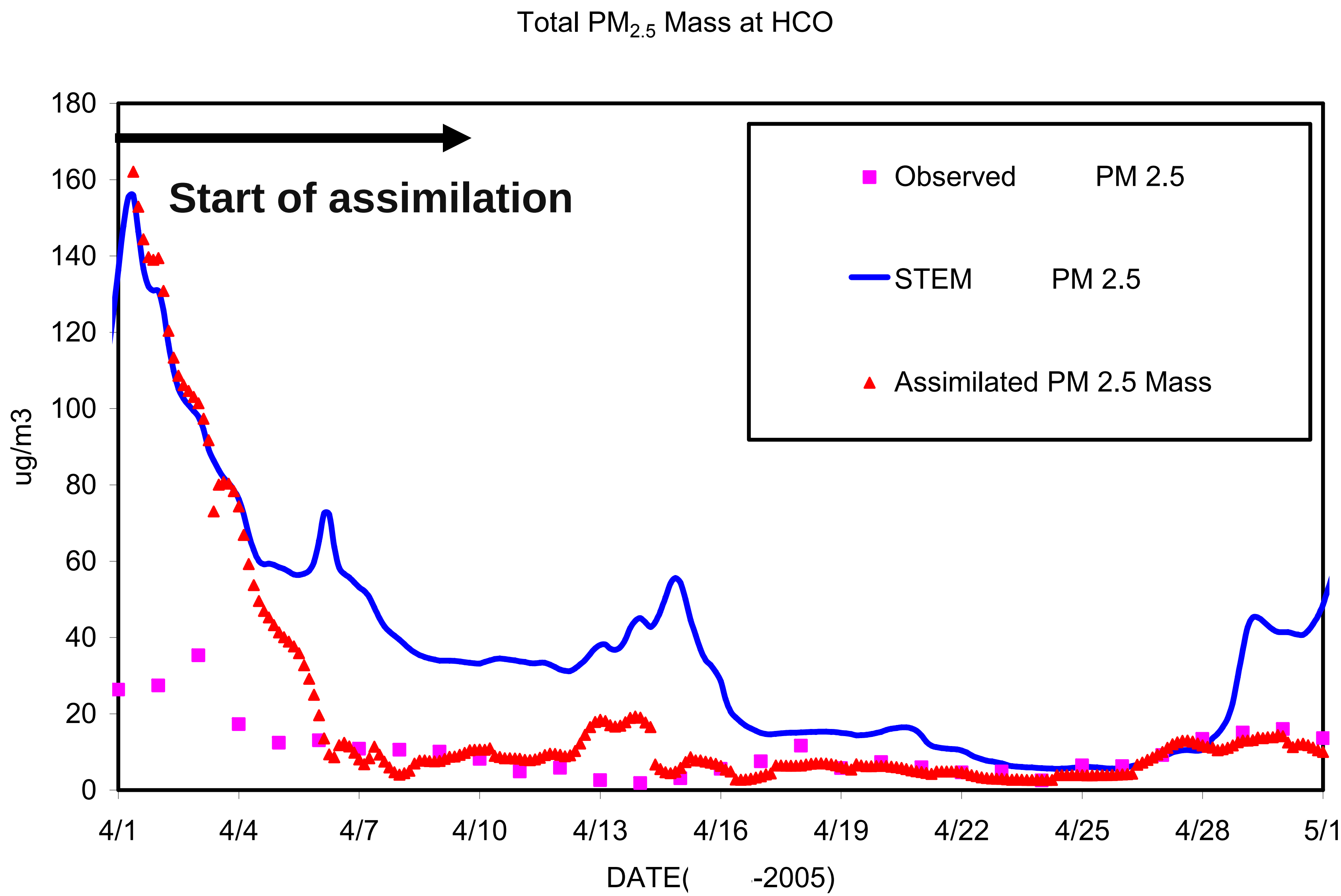
Technique: Collins, W. D., et al. (2001),, *JGR*, *106*, 7313-7336





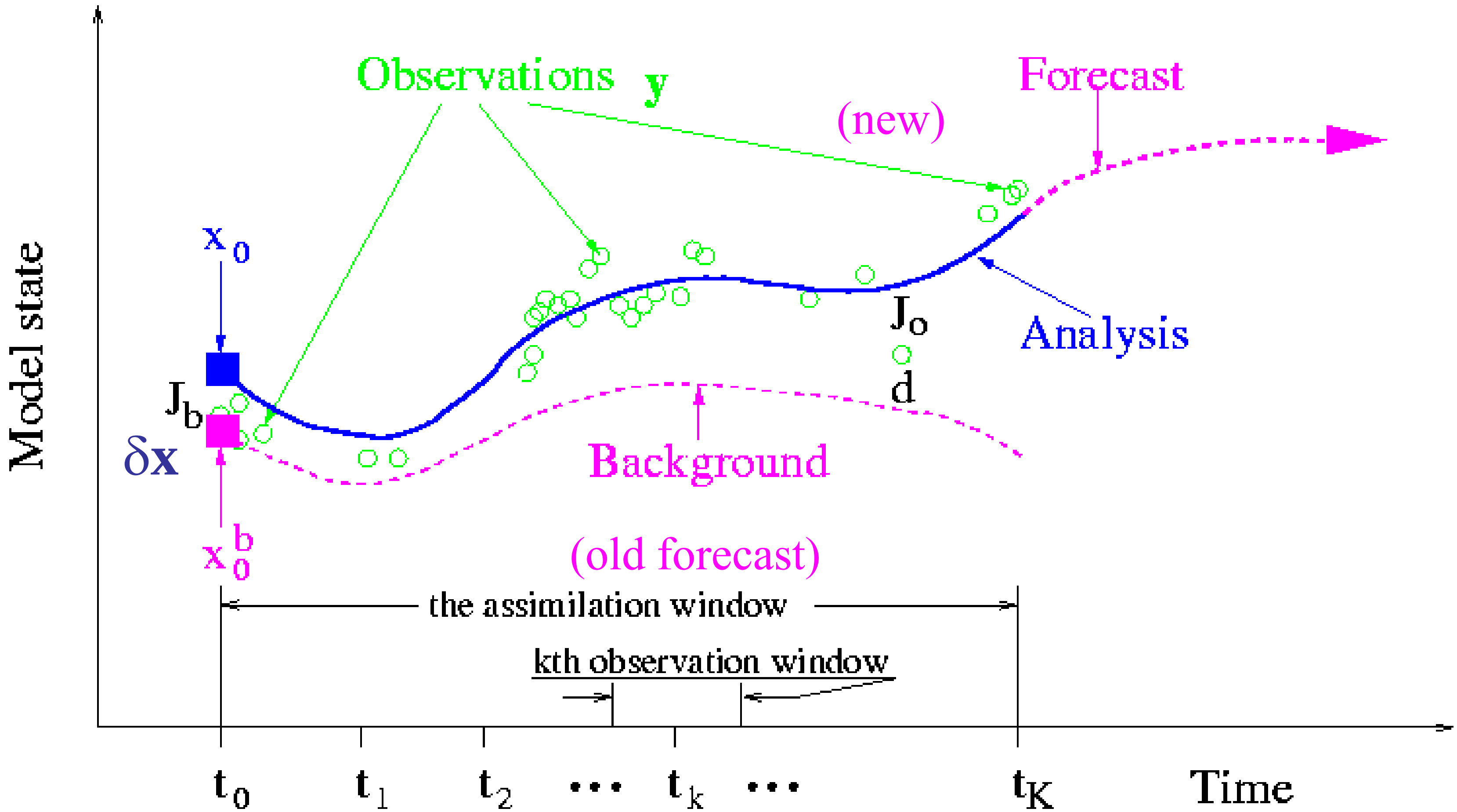
Adhikary et al., 2007,2008

Impact of Daily MODIS Assimilation on Predicted PM 2.5 at HCO



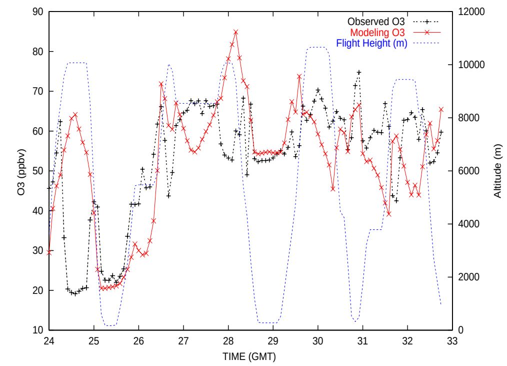
Adhikary et al., 2007,2008

Data assimilation



Advanced Data Assimilation Techniques Provide Data Fusion and Optimal Analysis Frameworks

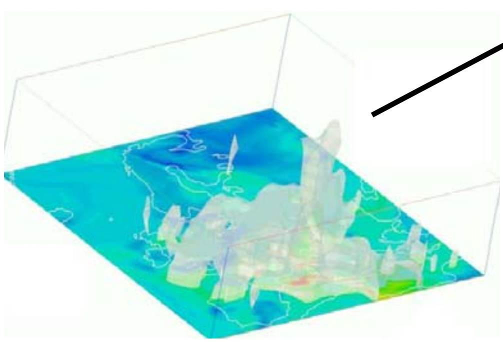
Model ~~vs.~~ + Observations

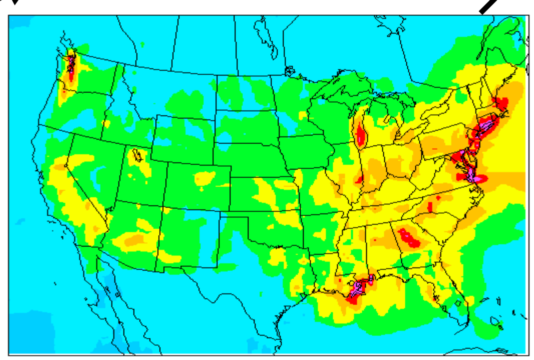


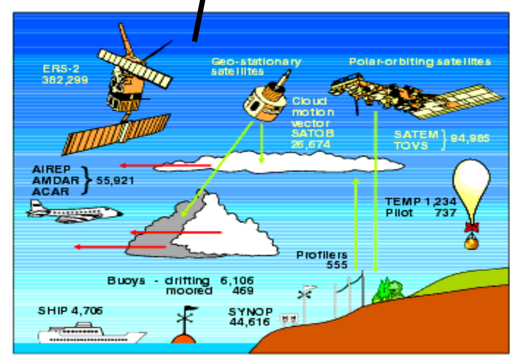
Example 4dVar:

Cost function

$$\min_{\mathbf{y}} \psi(\mathbf{y}) = \left\| \mathbf{y} - \mathbf{y}^{\mathbf{b}} \right\|_{\mathbf{B}^{-1}}^{2} + \left\| \mathbf{H} \cdot M(\mathbf{y}) - \mathbf{o} \right\|_{\mathbf{R}^{-1}}^{2}$$







Current knowledge of the state

Model information consistent with physics/chemistry

Observations information consistent with reality

The system is very under-determined – need to combine heterogeneous data sources with limited spatial/temporal information

Basic idea of 4D-Var

- Define a cost functional

$$J(c^0)=\frac{1}{2}\left(c^0-c^{\mathrm{b}}\right)^T B^{-1}\left(c^0-c^{\mathrm{b}}\right)+\frac{1}{2}\sum_{k=0}^{N}\left(c^k-c^{k,\mathrm{obs}}\right)^T R_k^{-1}\left(c^k-c^{k,\mathrm{obs}}\right)$$

which measures the distance between model output and observations, as well as the deviation of the solution from the background state

- Derive adjoint of tangent linear model

$$\frac{\partial\lambda_i}{\partial t}+\nabla\cdot(u\lambda_i)=-\nabla\cdot\left(\rho K\nabla\frac{\lambda_i}{\rho}\right)-\left(F^T(\rho c)\lambda\right)_i-\varphi_i$$

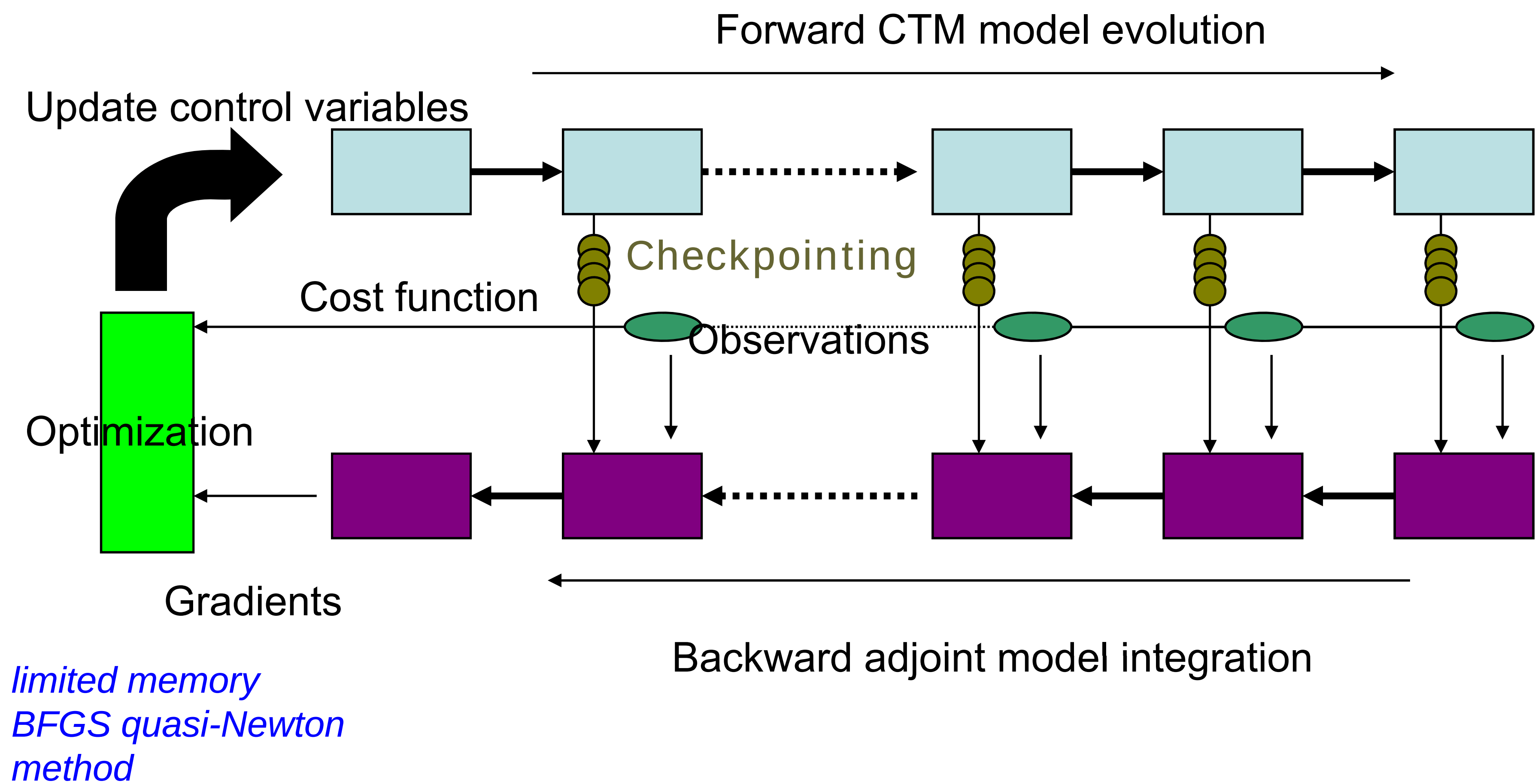
Where φ is the forcing term, which is chosen so that the adjoint variables are the sensitivities of the cost functional with respect to state variables (concentrations), i.e. $\lambda_i=\frac{\partial J}{\partial c_i}$

- Use adjoint variables for sensitivity analysis, as well as data assimilation

Challenges in chemical data assimilation

- A large amount of variables (~300 concentrations of various species at each grid points)
  - Memory shortage (check-pointing required)
- Various chemical reactions (>200) coupled together (lifetimes of species vary from seconds to months)
  - Stiff differential equations
- Chemical observations are very limited, compared to meteorological data
  - Information should be maximally used, with least approximation
- Highly uncertain emission inventories
  - Inventories often out-dated, and uncertainty not well-quantified

4D-Var application with CTMs



Computational aspects

- Discrete/continuous adjoint models, and their analysis, for stiff chemical systems
  - integral-partial-differential aerosol dynamic equations
  - upwind and slope/flux limited hyperbolic schemes
  - second order adjoints and optimization algorithms

KPP

- **KPP: Kinetic PreProcessor**
  - reads ASCII input and generates code for chemistry integration
- e.g.:
  - Damian et al., 2002
  - Sandu et al., 2003
  - Sandu and Sander, 2006
- http://people.cs.vt.edu/~asandu/Software/Kpp/

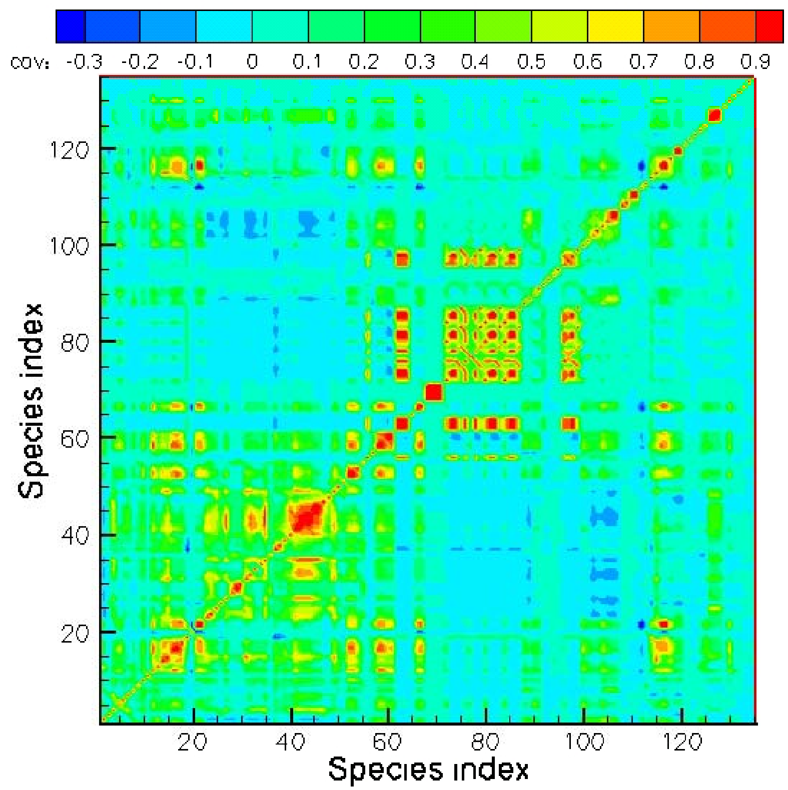
Estimation of B and O critical

NMC method (B)

- Substitute model background errors with the differences between 24hr, 48 hr, 72 hr forecasts verifying at the same time

- Calculate the model background error statistics in three directions separately

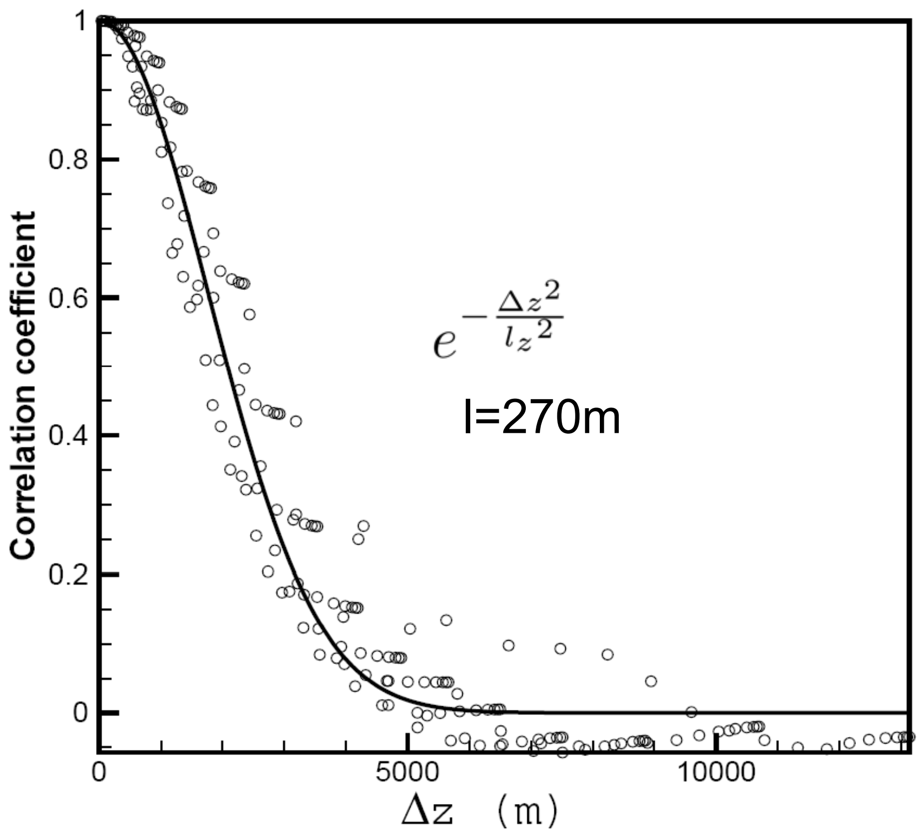
$$CORR(O_3, CO) = \frac{\overline{\epsilon_{O_3} \cdot \epsilon_{CO}}}{\sqrt{\overline{\epsilon_{O_3} \cdot \epsilon_{O_3}}} \cdot \sqrt{\overline{\epsilon_{CO} \cdot \epsilon_{CO}}}}$$



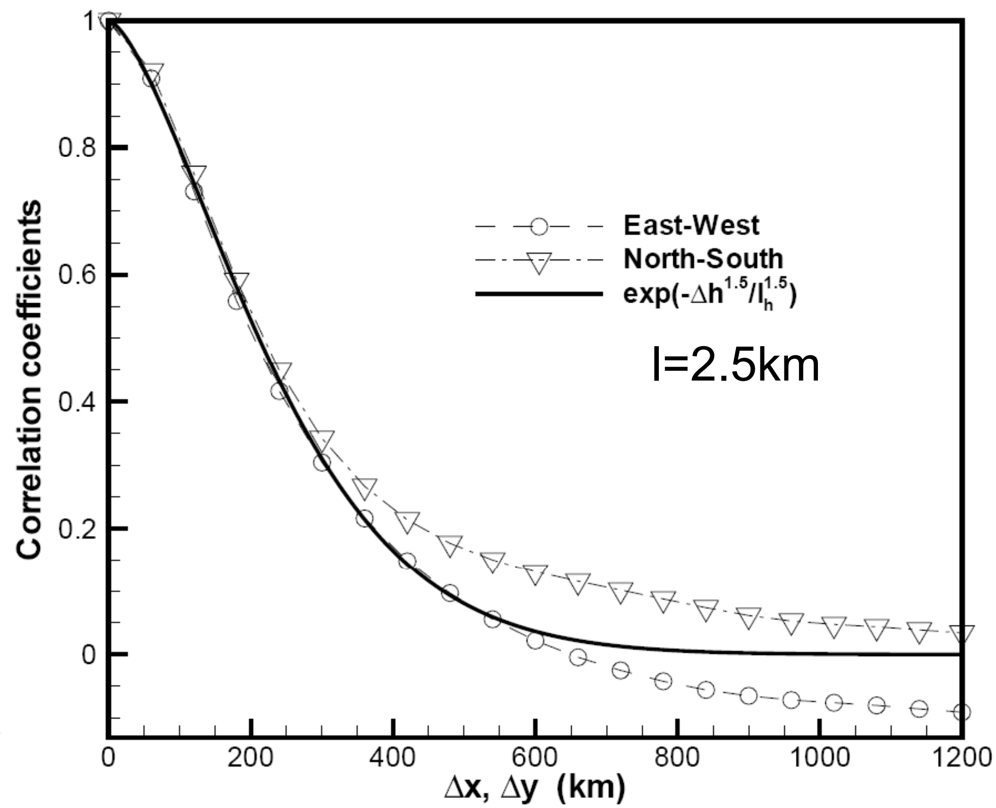
- Equivalent sample number: 811,890

NMC method results

Vertical correlation

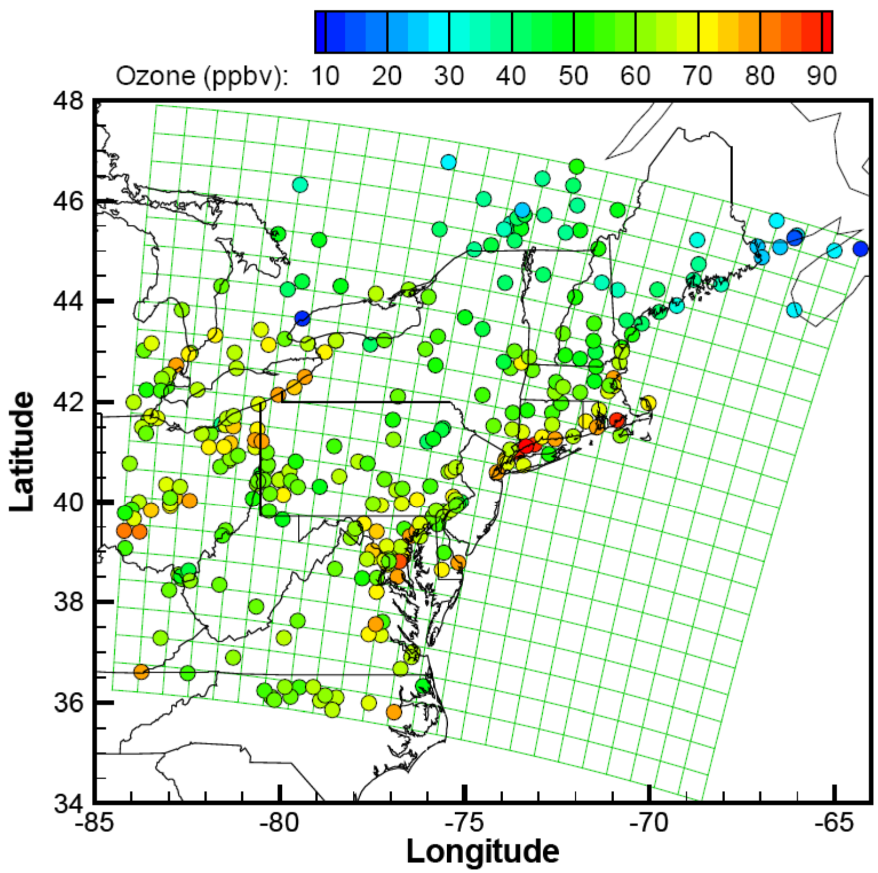


Horizontal correlation



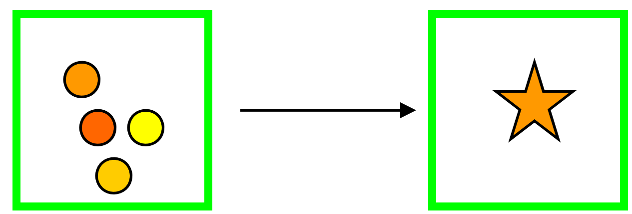
Observational error

$$J = \frac{1}{2}\left[c_0 - c_b\right]^T B^{-1}\left[c_0 - c_b\right] + \frac{1}{2}\left[y - h(c)\right]^T O^{-1}\left[y - h(c)\right]$$



Observational Error:

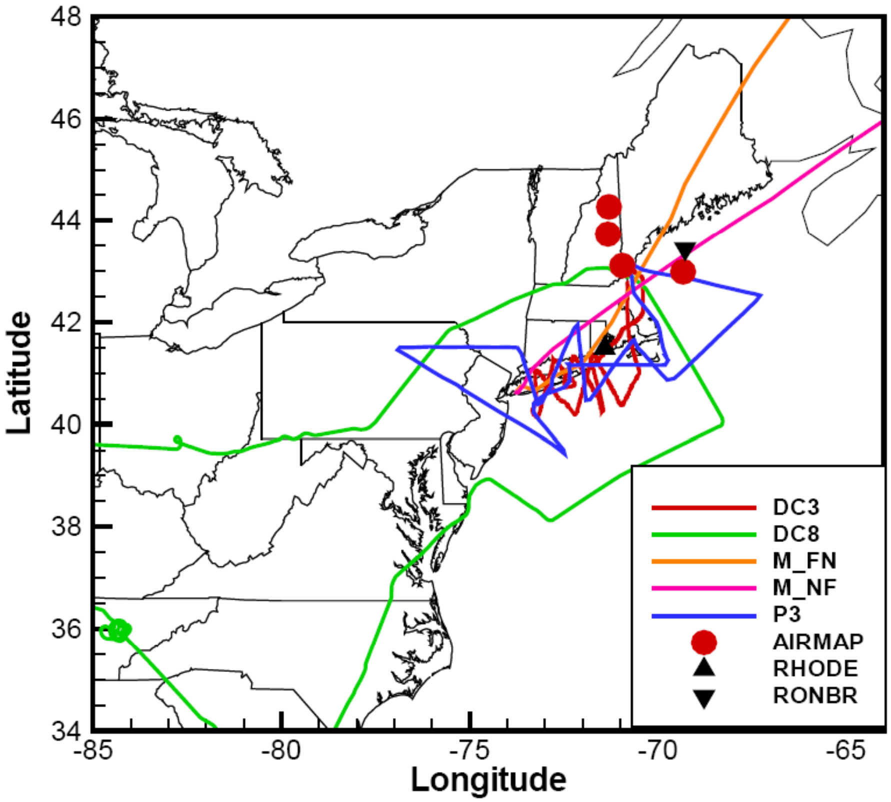
- Representative error
- Measurement error



Observation Inputs

- Averaging inside 4-D grid cells
- Uniform error (8 ppbv)

Assimilation of ICARTT Ozone Observations



Observations	Description
AIRNOW	EPA surface stations, hourly averaged data used
DC3	Vertical profile of ozone mixing ratio from lidar
MOZ-FN	MOZAIC, Frankfurt-New York flight
MOZ-NF	MOZAIC, New York-Frankfurt flight
P3	NOAA P3-B measurement
AIRMAP	UV SPECTROSCOPY measurement at 4 sites
DC8-In	NASA In Situ Ozone via Nitric Oxide Chemiluminescence
DC8-Li	DC-8 Composite Tropospheric Ozone Cross-Sections
RHODE	Ozonesonde/Radiosonde data from Narragansett, RI
RONBR	Ozonesonde/Radiosonde data from the R/V Ronald H. Brown

Intensive Field Experiments (e.g., ICARTT) Provide Our Best Efforts to Comprehensively Observe a Region

O_3 Observations

Table 3.1 Ozone observations on July 20, 2004

Observations	Description
AIRNOW	EPA surface stations, hourly averaged data used
DC3	Vertical profile of ozone mixing ratio from lidar
MOZ-FN	MOZAIC, Frankfurt-New York flight
MOZ-NF	MOZAIC, New York-Frankfurt flight
P3	NOAA P3-B measurement
AIRMAP	UV SPECTROSCOPY measurement at 4 sites
DC8-In	NASA In Situ Ozone via Nitric Oxide Chemiluminescence
DC8-Li	DC-8 Composite Tropospheric Ozone Cross-Sections
RHODE	Ozonesonde/Radiosonde data from Narragansett, RI
RONBR	Ozonesonde/Radiosonde data from the R/V Ronald H. Brown

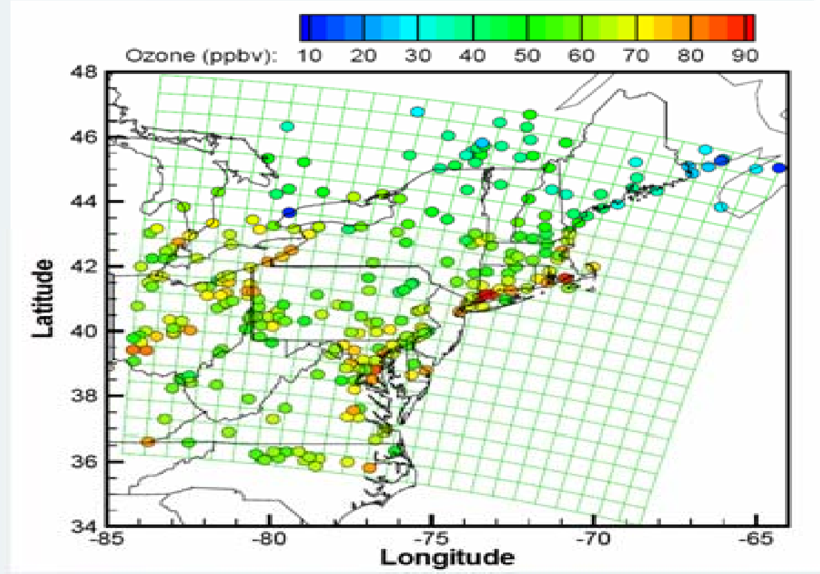


Figure 3.1 Computational grid and AIRNOW stations (color coded by ozone measurements at 1900 UT on July 20, 2004)

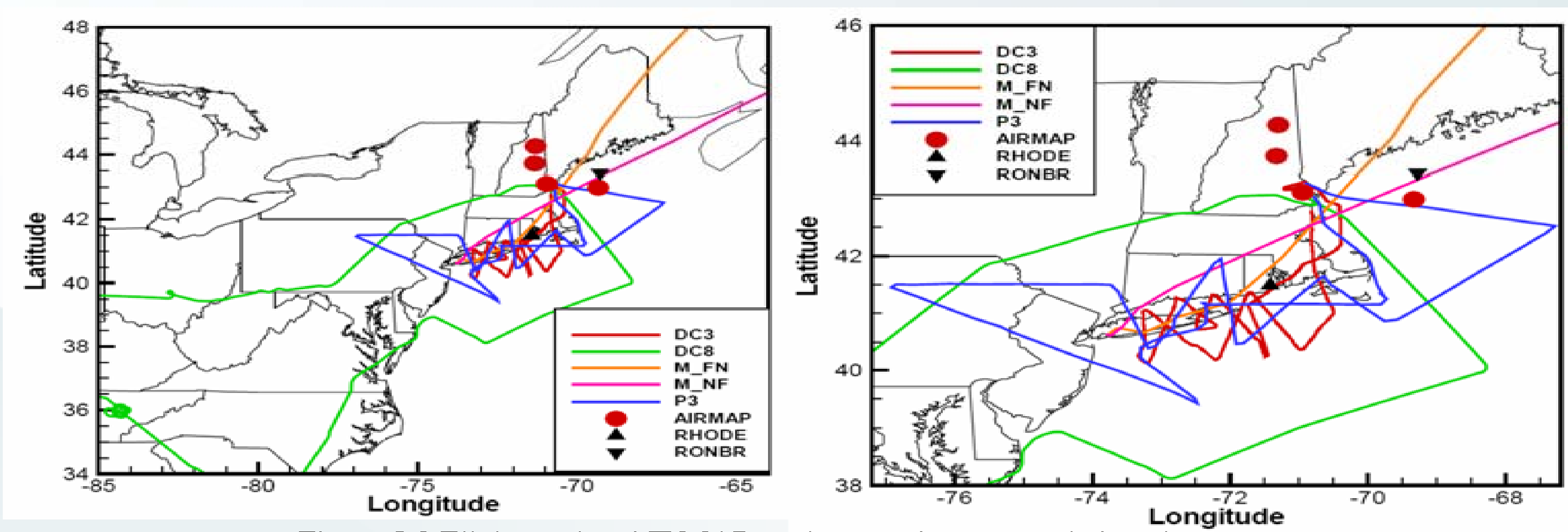
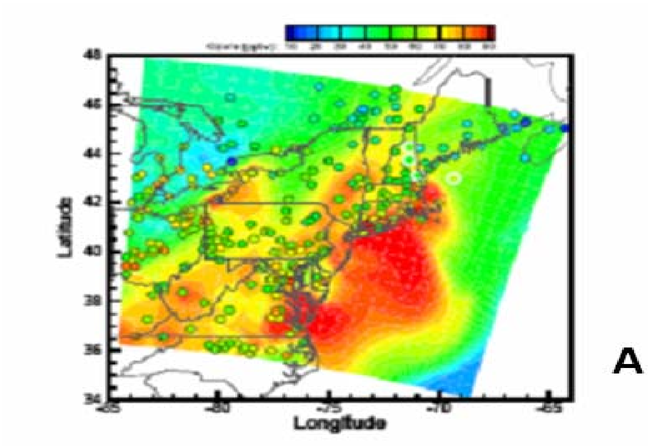


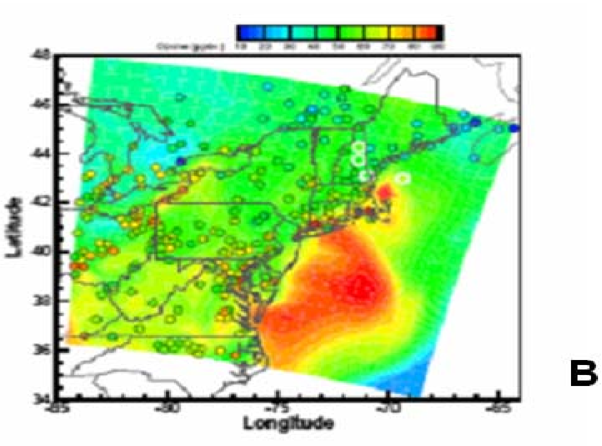
Figure 3.2 Flight tracks, AIRMAP stations, and ozonesonde locations.

Assimilation Produces An Optimal State Space

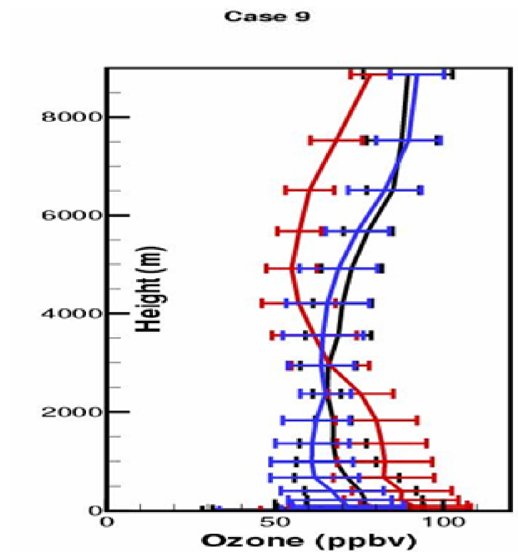
w/o assimilation | **with assimilation**

Ozone predictions





All Data Used



Example July 20, 2004

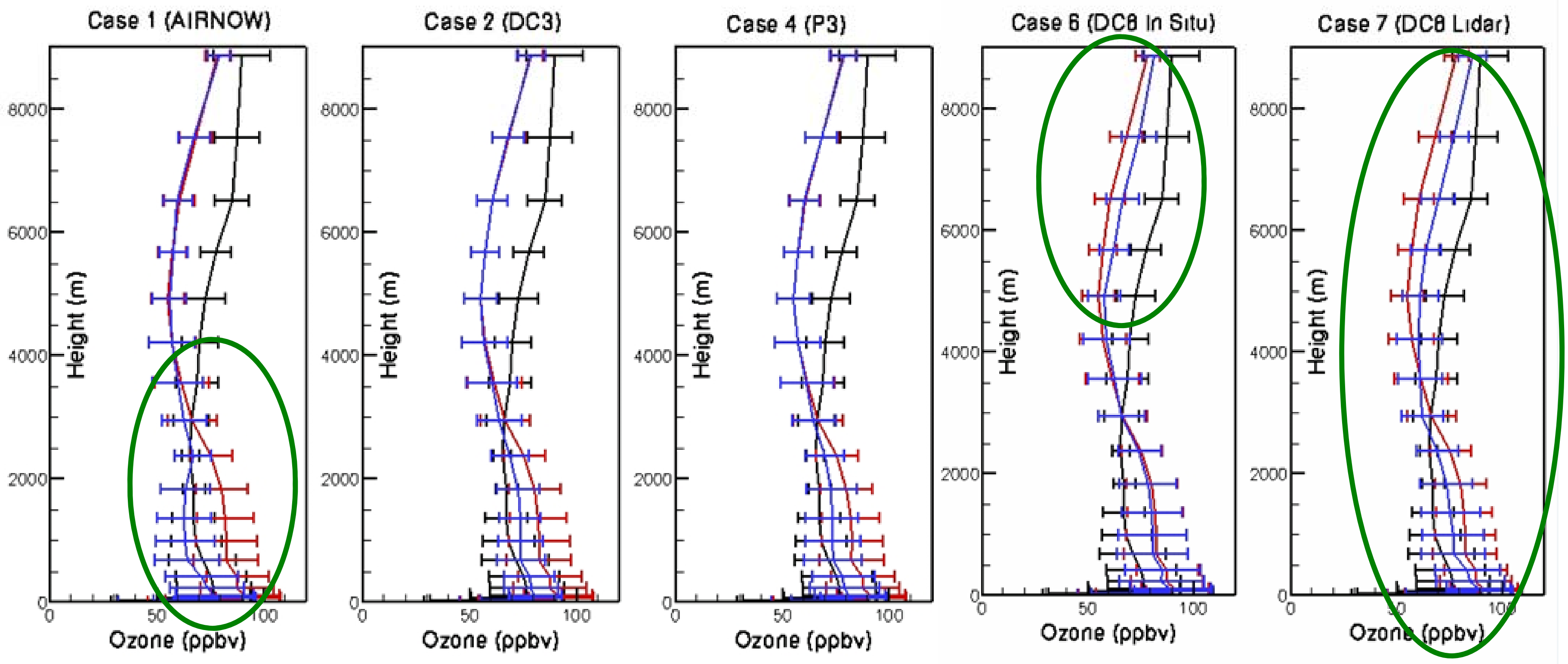
Chai et al., JGR 2007

Region-mean profile

Case	Assimilated Observations	Time	Number
1	AIRNOW	1300–2400 UT, hourly	2075
2	DC3	1852–2356 UT	412
3	MOZ-FN, MOZ-NF	1947–2007 UT, 2238–2252 UT	38
4	P3	1412–2207 UT	208
5	AIRMAP	1215–2400 UT	192
6	DC8-In	1416–2207 UT	138
7	DC8-Li	1429–2137 UT	465
8	RHODE, RONBR	1810–1822 UT, 1900–1921 UT	35
9	All above	1200–2400 UT	3563

Information content of various observations evaluated by different combinations of data sets assimilated –

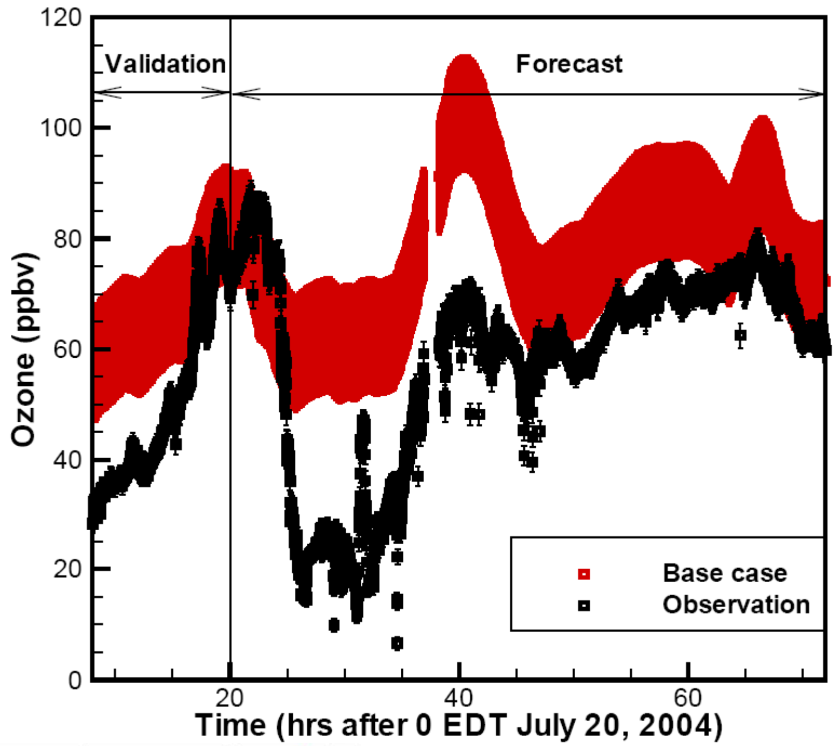
the importance of measurements above the surface.

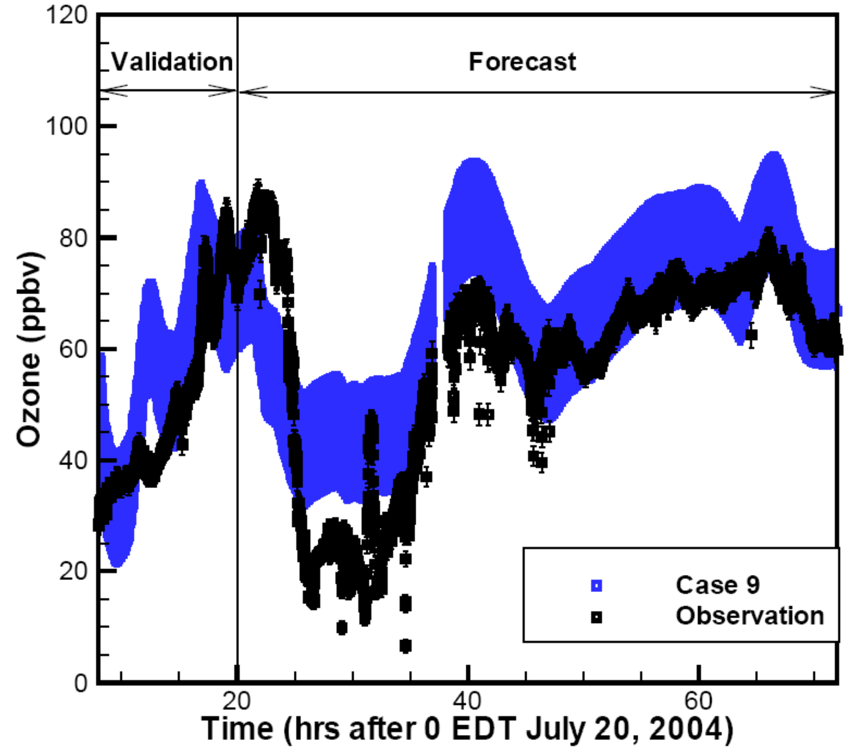


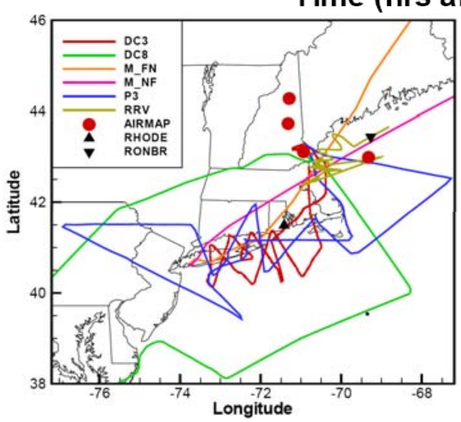
Surface-only

Lidar-DC8

Verification: Ron Brown Observations Independent Data



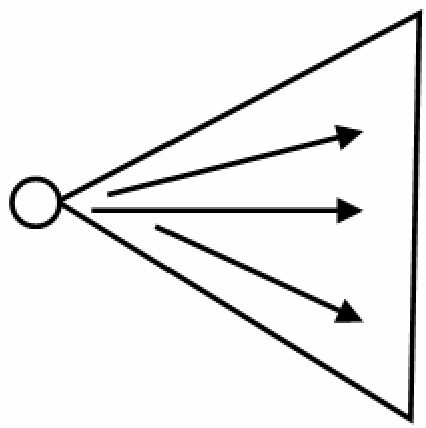




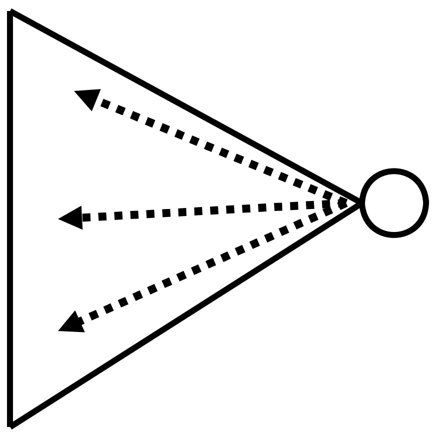
Predicted uncertainties estimated from background (B) error estimates

Source/Receptor Calculations: Perturbation approaches

Source-oriented approach - Direct sensitivity analysis.



Receptor/target-oriented approach - Adjoint sensitivity analysis.

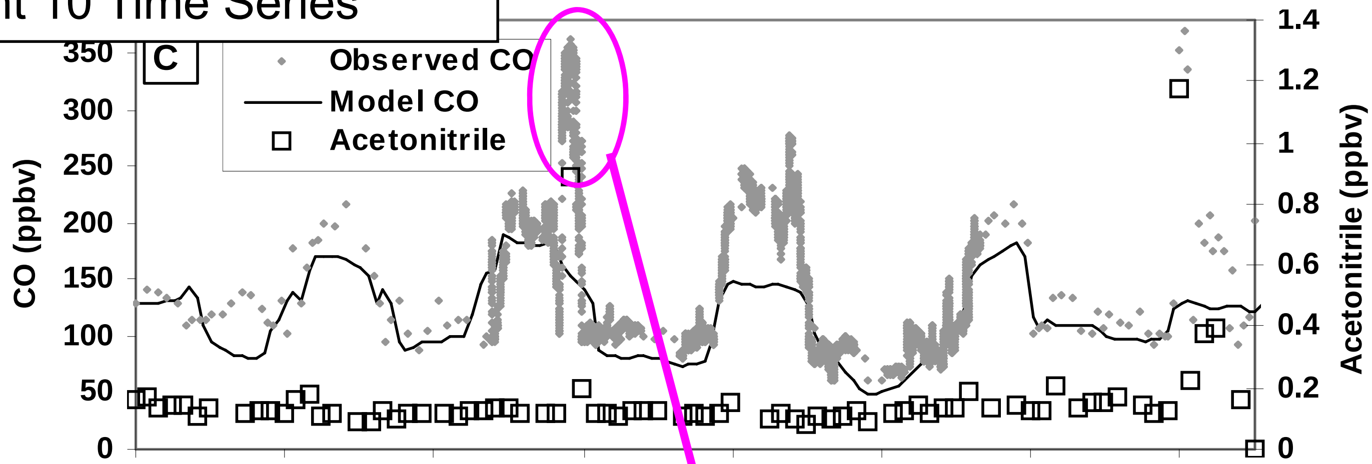


The Adjoints Are Themselves Very Valuable

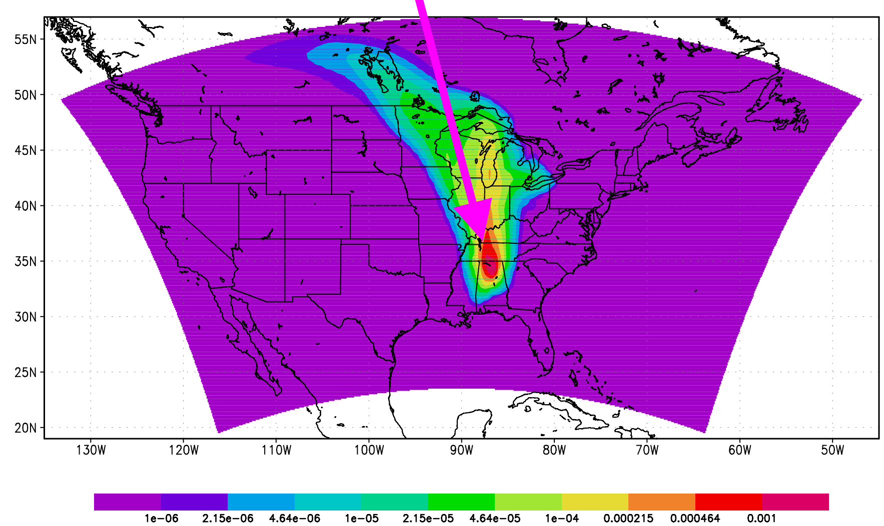
CO as a tracer of fossil fuel CO_2…

Caveat: Fire, chemistry, LPS (Campbel et al, In Press)

Flight 10 Time Series



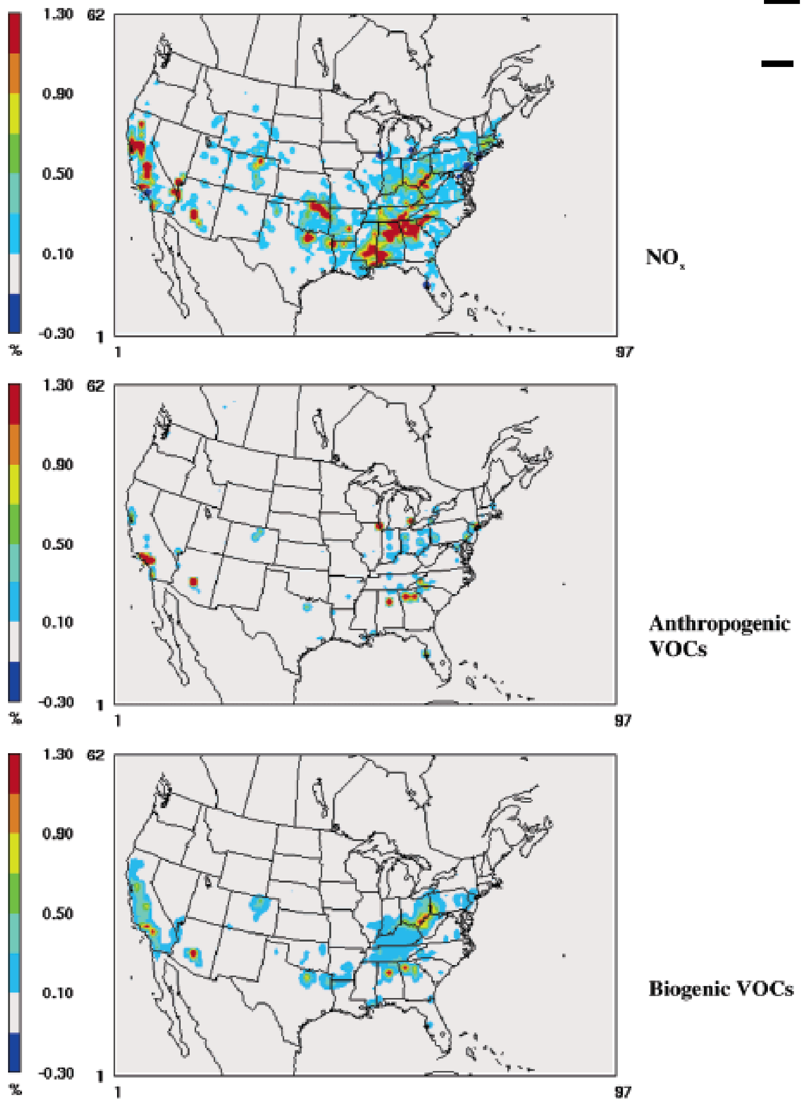
Adjoint Sensitivity

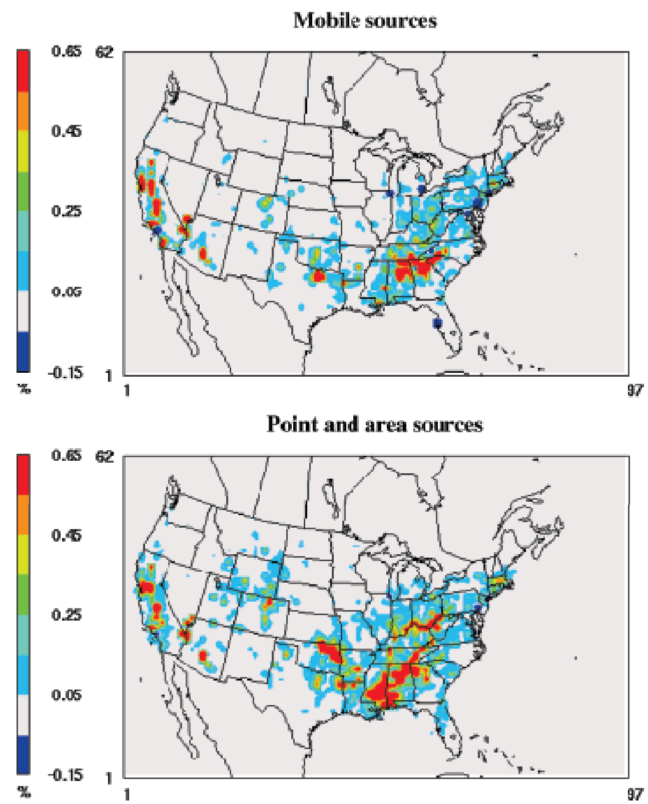


Provide insights into the footprint of an observation

Sensitivity of ozone violations wrt emissions

Adjoint Analysis of the Contribution of Different Emissions to Ozone Violations – July & August 2004





Hakami et al., ES&T 2006

The Ensemble Kalman Filter (EnKF)

For the linear, Gaussian case Bayes formula gives (Kalman)

$$\mathbf{y}_f^k = M\left(t^{k-1}, \mathbf{y}_a^{k-1}\right), \quad \mathbf{P}_f^k = \frac{\partial M}{\partial \mathbf{y}}\left(\mathbf{y}_a^{k-1}\right) \cdot \mathbf{P}_a^{k-1} \cdot \left(\frac{\partial M}{\partial \mathbf{y}}\left(\mathbf{y}_a^{k-1}\right)\right)^T$$

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

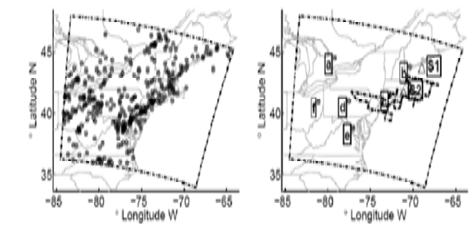
- All sources of information used (model, background, observations)
- Observations incorporated one batch at a time
- Propagation of covariances very expensive
- In EnKF covariances are approximated by an ensemble of runs
- No need for adjoint model (ease of use)
- Can propagate uncertainty through nonlinear models

Ensemble-based chemical data assimilation techniques can complement the variational tools

- Motivation:
  - Ensemble-based d.a. generate a statistical sample of analyses
  - Optimal state estimation applied to each member
  - Can deal effectively with nonlinear dynamics
  - Explicitly propagate (approximations of) the error statistics
  - 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

Challenges for reanalysis and forecasting appear to be different …. 4D-var and EnKF show promise for reanalysis

Simulation and data assimilation method	R^2 (RMS) analysis	
Best guess solution, no assimilation	0.24 (22.1)	
EnKF (50 members) "noiseless application"	0.38 (18.2)	
EnKF (200 members) "noiseless application"	0.49 (16.3)	
EnKF (50 members) adaptive multiplicative inflation	0.67 (12.7)	
EnKF (200 members) adaptive multiplicative inflation	0.82 (9.36)	
LEnKF (50 members), "noiseless application"	0.81 (9.79)	
LEnKF (50 members) adaptive multiplicative inflation	0.82 (9.52)	
LEnKF (50 members), "noiseless". Joint assimilation of state, emissions, and lateral boundary conditions	0.88 (7.75)	
LEnKF (50 members) adaptive multiplicative inflation. Joint assimilation of state, emissions, and lateral boundary conditions	0.91 (6.52)	



(a) ICARTT ground stations (b) selected stations

gure 1: Ground measuring stations (a) in support of the ICARTT campaign (340 in total), d (b) selected stations (#a–#f), two ozonesondes (S1, S2) and the flight path of a P3 plane at will be used for the numerical results/validation illustration.

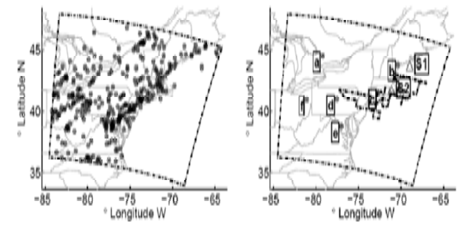
TABLE 2. Model-observations agreement (R^2 and RMS [ppbv]) for the EnKF data assimilation of only the state and of the joint state (ST), emissions (EM) and lateral boundary conditions (BC) parameters. Visible improvements in both the analysis and the forecast are obtained by adjusting the emissions and lateral boundary conditions.

Sandu et al., Quart. J. Roy. Met. Soc, 2007

Challenges for reanalysis and forecasting appear to be different …. 4D-var and EnKF show promise for reanalysis but more work is needed to impact forecasts

Simulation and data assimilation method	**R^2 (RMS) analysis**	**R^2 (RMS) forecast**
Best guess solution, no assimilation	0.24 (22.1)	0.28 (23.5)
4D-Var 50 iterations w/ AR background	0.52 (16.0)	0.29 (22.4)
EnKF (50 members) "noiseless application"	0.38 (18.2)	0.30 (23.1)
EnKF (200 members) "noiseless application"	0.49 (16.3)	0.30 (23.7)
EnKF (50 members) adaptive multiplicative inflation	0.67 (12.7)	0.19 (62.0)
EnKF (200 members) adaptive multiplicative inflation	0.82 (9.36)	0.28 (37.6)
LEnKF (50 members), "noiseless application"	0.81 (9.79)	0.34 (22.0)
LEnKF (50 members) adaptive multiplicative inflation	0.82 (9.52)	0.34 (22.0)
LEnKF (50 members), "noiseless". Joint assimilation of state, emissions, and lateral boundary conditions	0.88 (7.75)	0.42 (20.3)
LEnKF (50 members) adaptive multiplicative inflation. Joint assimilation of state, emissions, and lateral boundary conditions	0.91 (6.52)	0.40 (20.5)

TABLE 2. Model-observations agreement (R^2 and RMS [ppbv]) for the EnKF data assimilation of only the state and of the joint state (ST), emissions (EM) and lateral boundary conditions (BC) parameters. Visible improvements in both the analysis and the forecast are obtained by adjusting the emissions and lateral boundary conditions.



(a) ICARTT ground stations (b) selected stations

igure 1: Ground measuring stations (a) in support of the ICARTT campaign (340 in total), nd (b) selected stations (#a–#f), two ozonesondes (S1, S2) and the flight path of a P3 plane hat will be used for the numerical results/validation illustration.

Sandu et al., QJRMS, 2007

What parameters should be target for adjustment? – emissions, initial conditions, boundary conditions? and What Species?

In AQ Predictions Emissions Are A Major Source Of Uncertainty – Data Assimilation Can Produce Optimal Estimates (Inverse Applications)

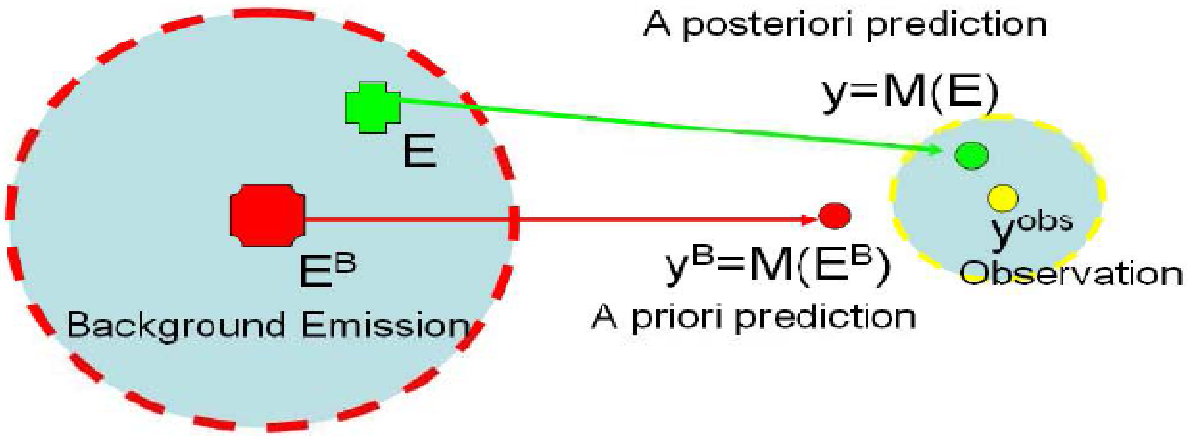


Fig. 14. A-basic methodology of top-down estimates of emissions.

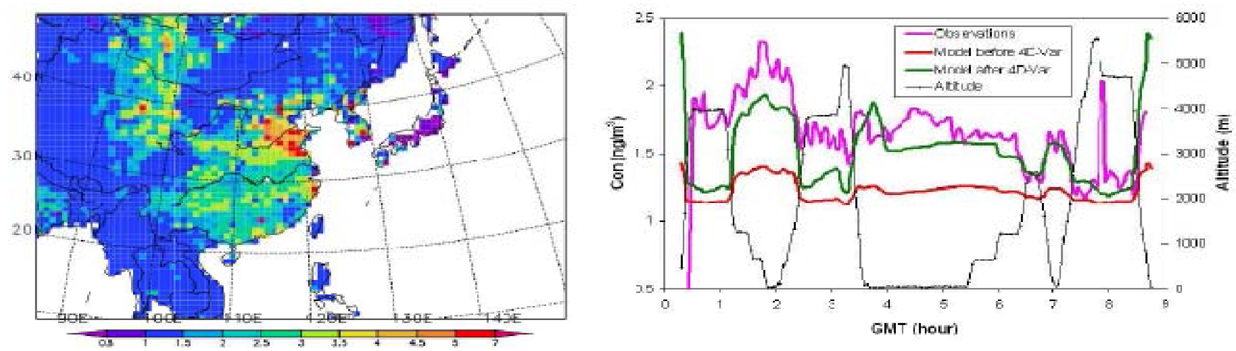
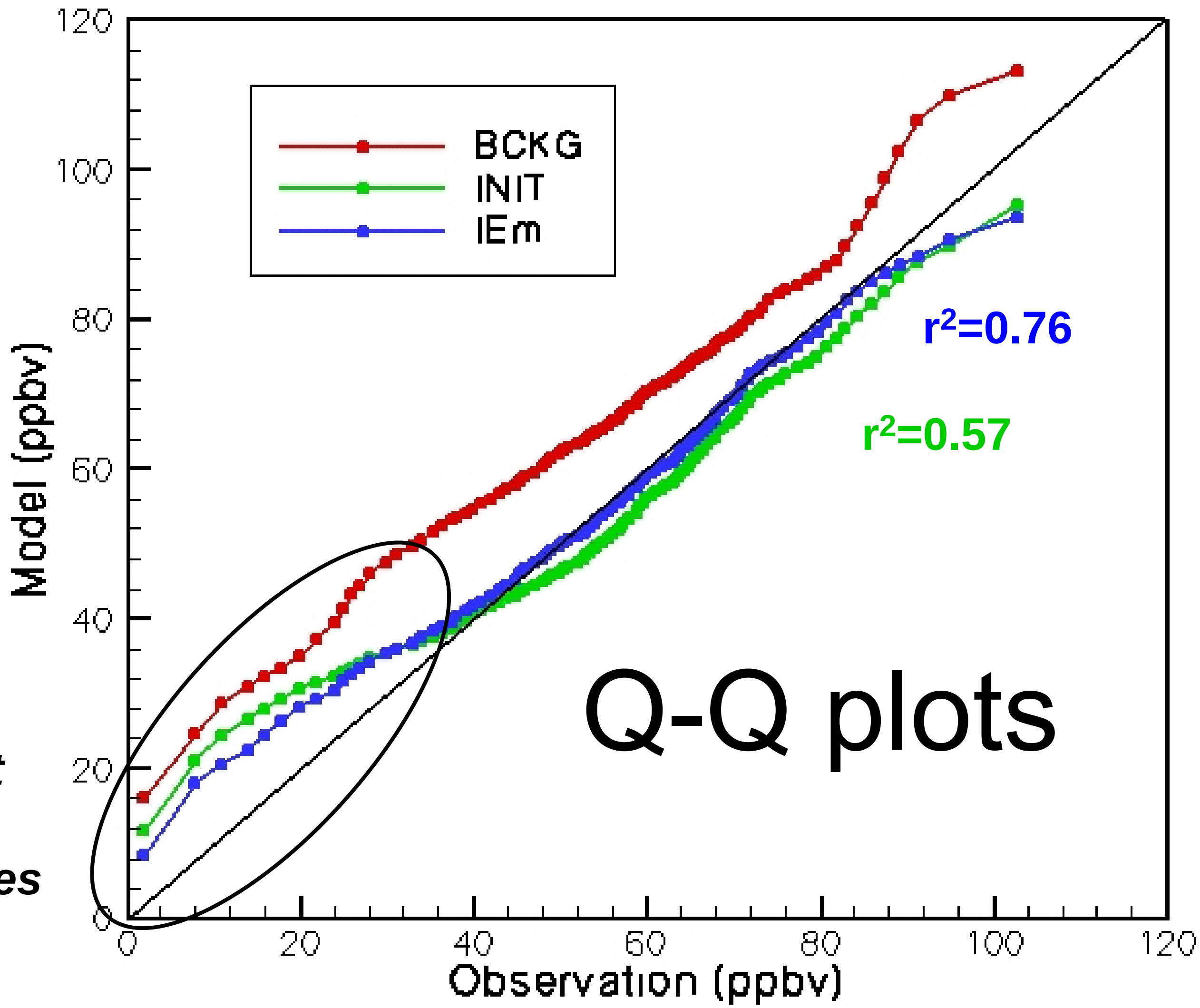


Fig. 15. Optimal mercury emission scaling factors obtained using the 4D-Var approach and the mercury measurements on board the C-130 during the Ace-Asia experiment. Results are for a month-long assimilation window (April 2001).

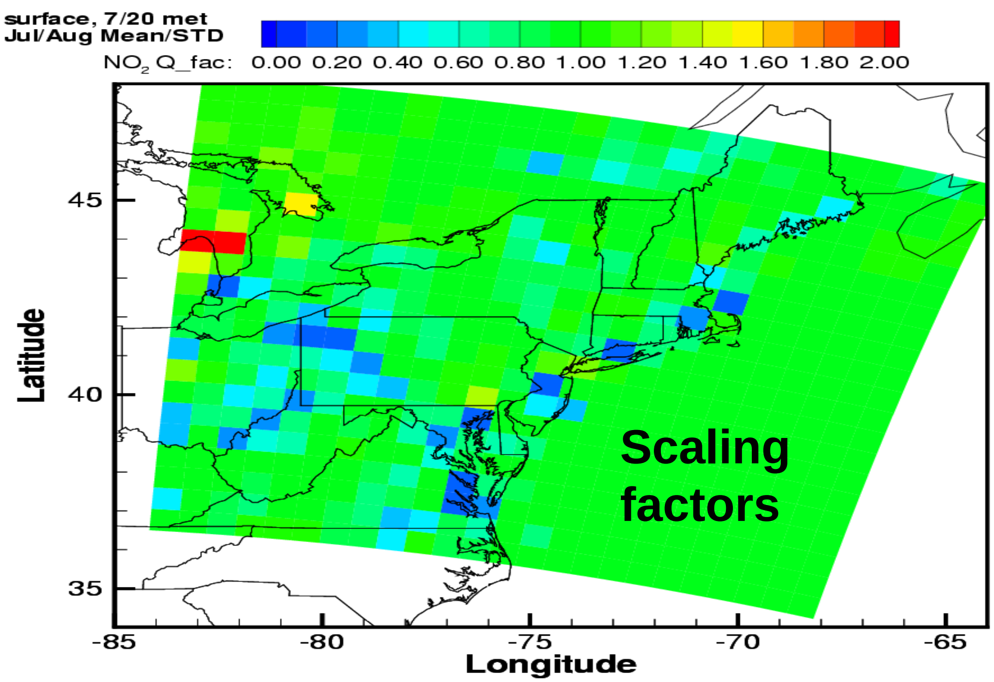
Li et al., Atmos. Env., 2007

Results of Consideration of Emissions only and Emissions and Initial Conditions



Points out model deficiencies

Emission Inversion with Satellite Data



4D-Var setup:

Time window:

1200 UT- 2000 UT

July 20, 2004

Control:

Initial ozone, and NOx emissions

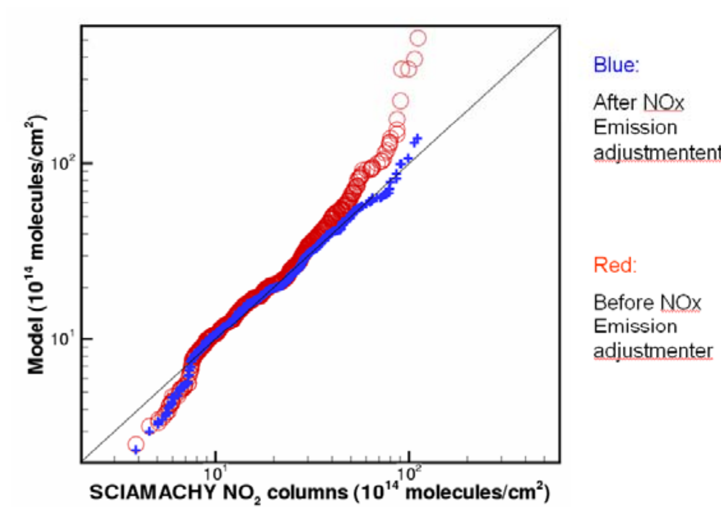
Observations:

Ozone from different platforms, and SCIAMACHY tropospheric NO_2 columns

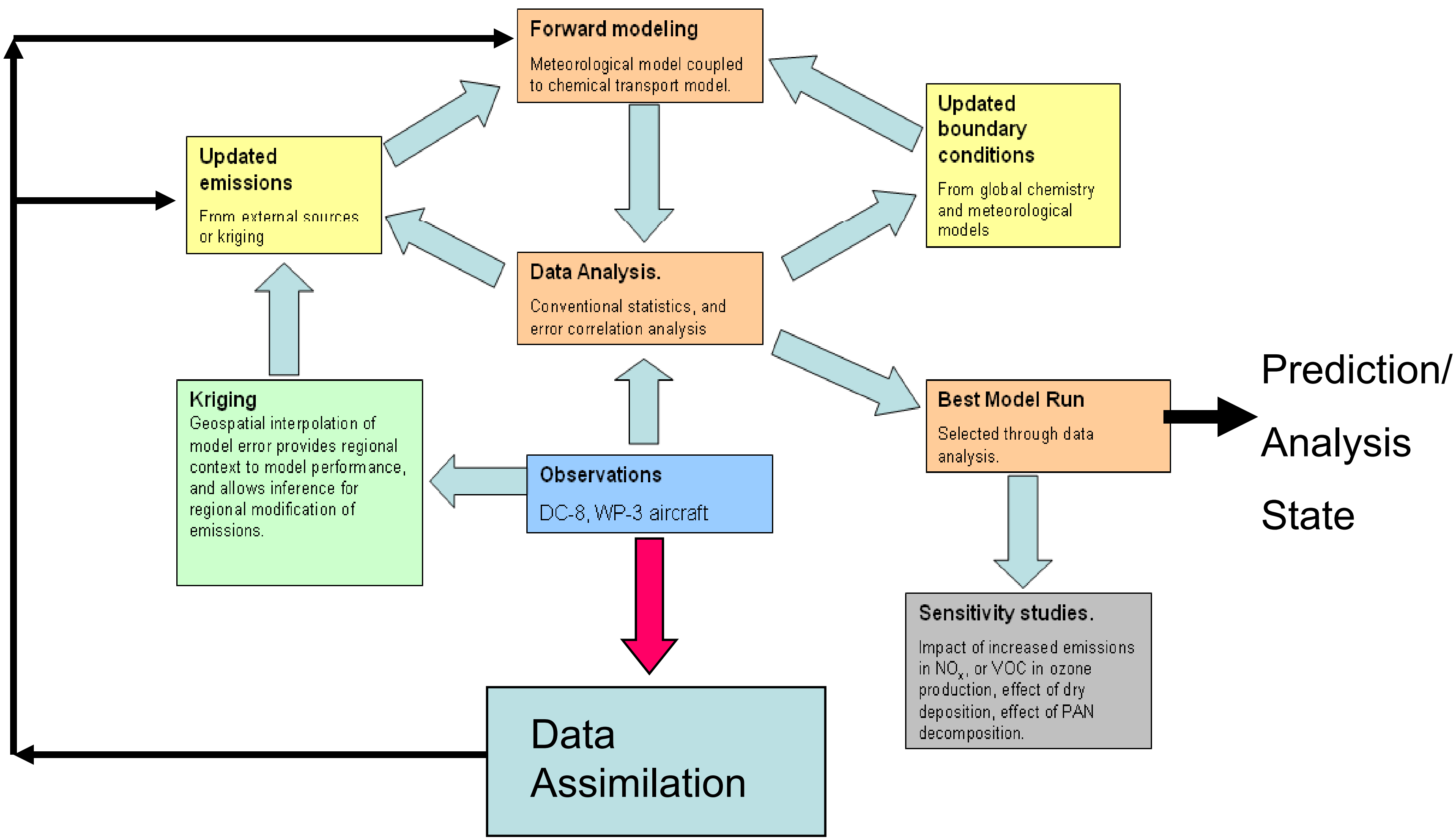
Emission changes over domain
(ratio of new emission over NEI01)

Case	Surface (level 1)	Elevated (2 & above)	Total (all levels)
1 **E only**	0.934	0.849	0.920
2 **E & IC**	0.928	0.881	0.908
"OI"	1.318	1.030	1.246

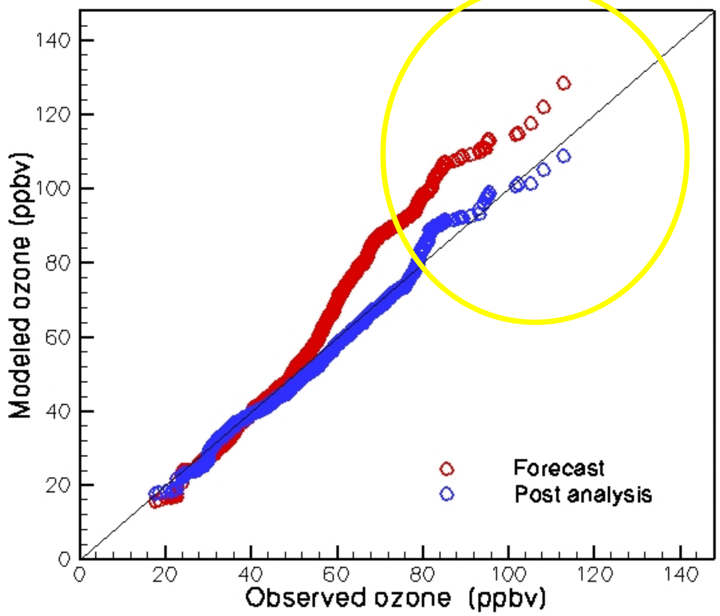
Quantile-quantile plot

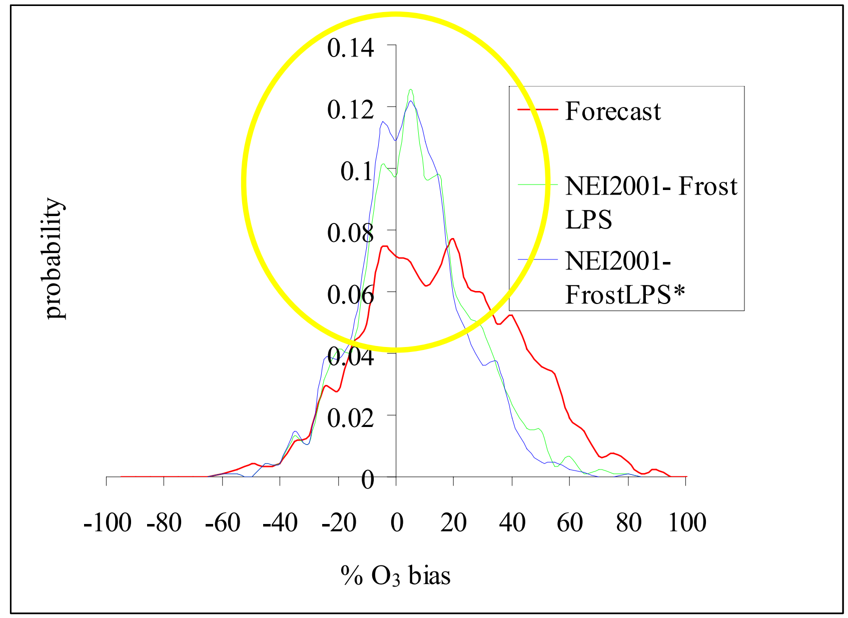


Our Analysis Approach



Documenting improvement (ICART)

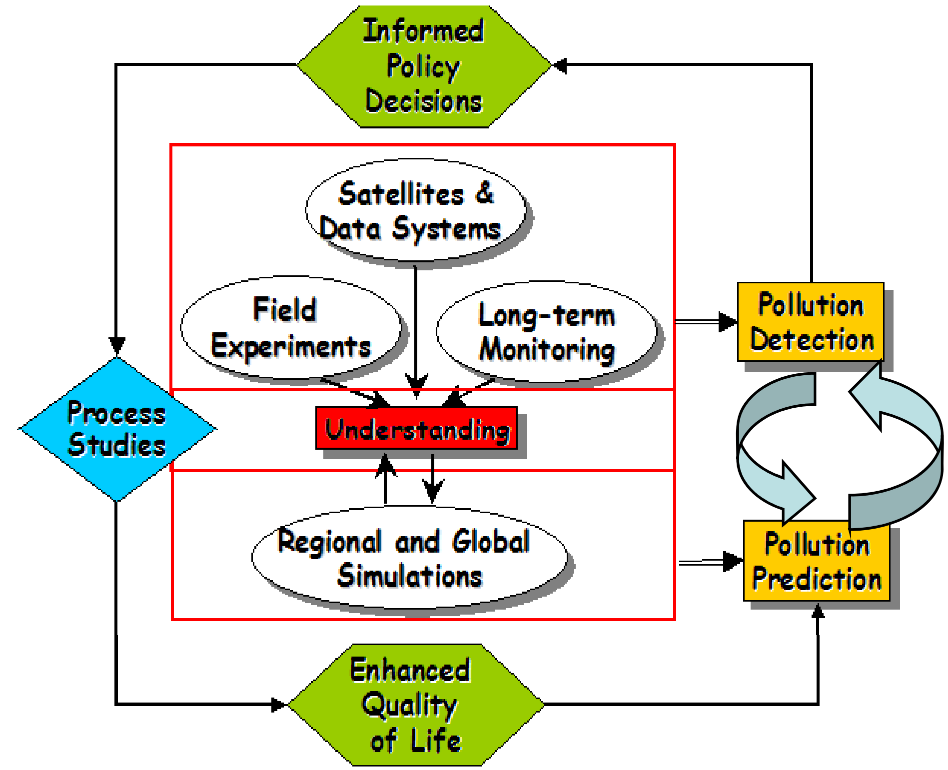


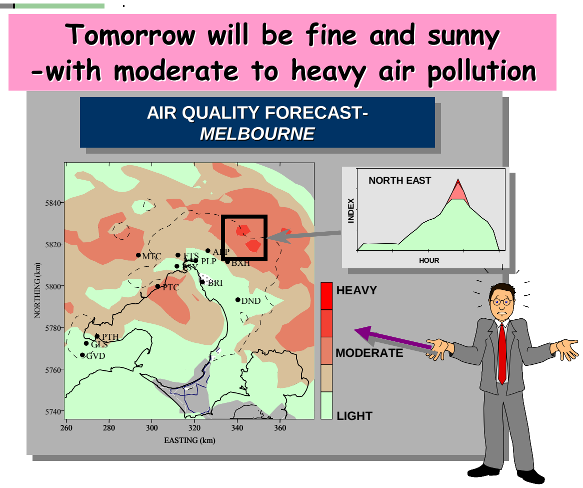


Left: Quantile-quantile plot of modeled ozone with observed ozone for DC-8 platform, data points collected at altitude less than 4000m, STEM-2K3, Forecast: NEI 1999, Post Analysis: NEI2001-Frost LPS*. MOZART-NCAR boundary conditions Right: Probability distribution of % ozone bias for Forecast (NEI 1999) and post analysis runs (NEI2001-FrostLPS and NEI2001-FrostLPS*) for DC-8 measurements under 4000m.

Mena et al., JGR, 2007

FUTURE DIRECTIONS FOR IMPROVING AIR QUALITY PREDICTIONS -- Summary





✓Models & measurements have improved substantially.

✓Further improvements will require reductions in key uncertainties (e.g., emissions, better basic understanding of some processes).

✓Closer integration of observations.

✓ Need to develop better strategies for providing uniqueness to targeted applications (e.g., sources & sectors).