AQ Modeling & Data Assimilation

WMO course on Seamless Prediction of Air Pollution in Africa

Kenyan Meteorological Department

Nairobi, October 2019

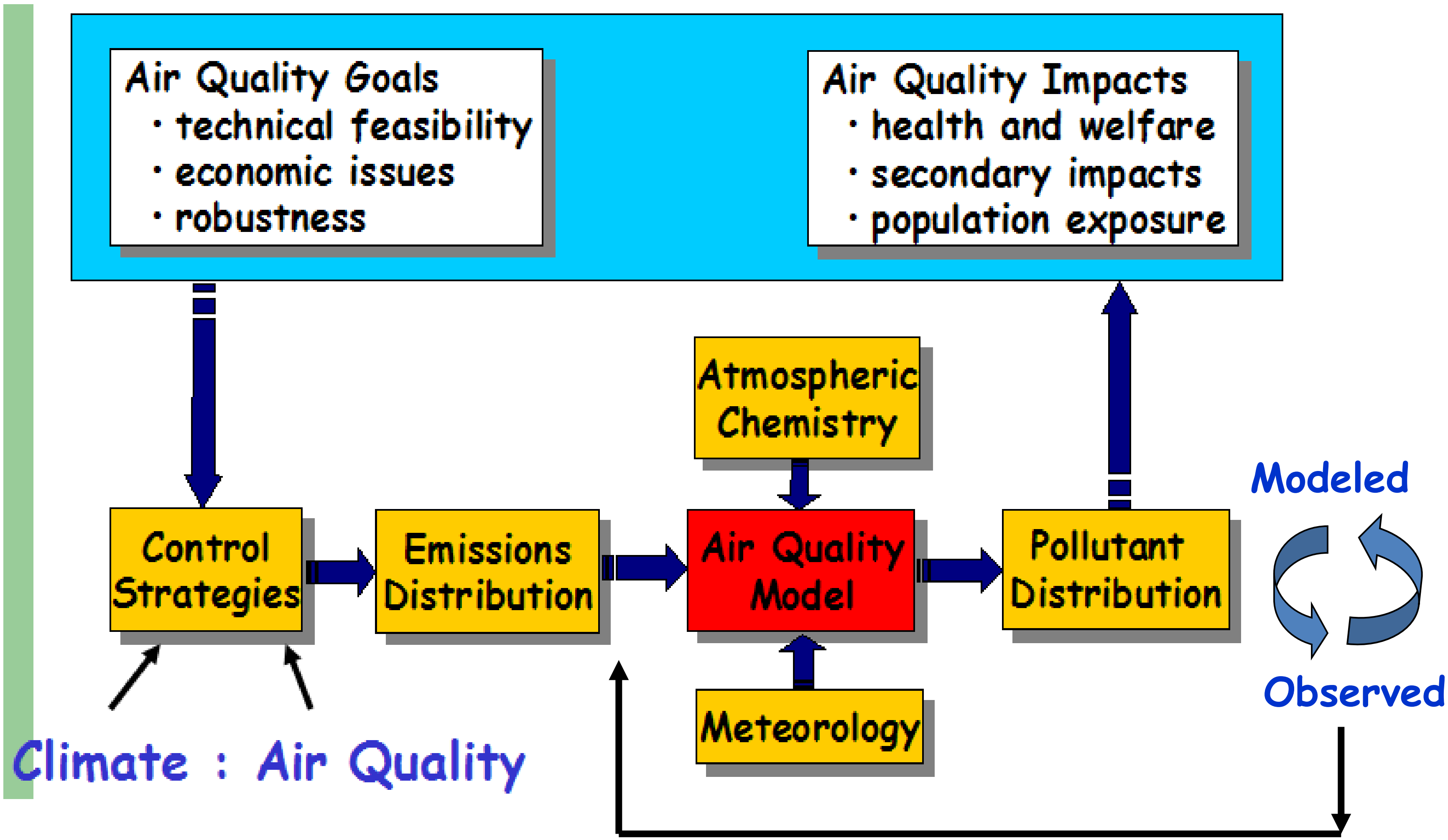
World Meteorological Organisation

GAW

GURME

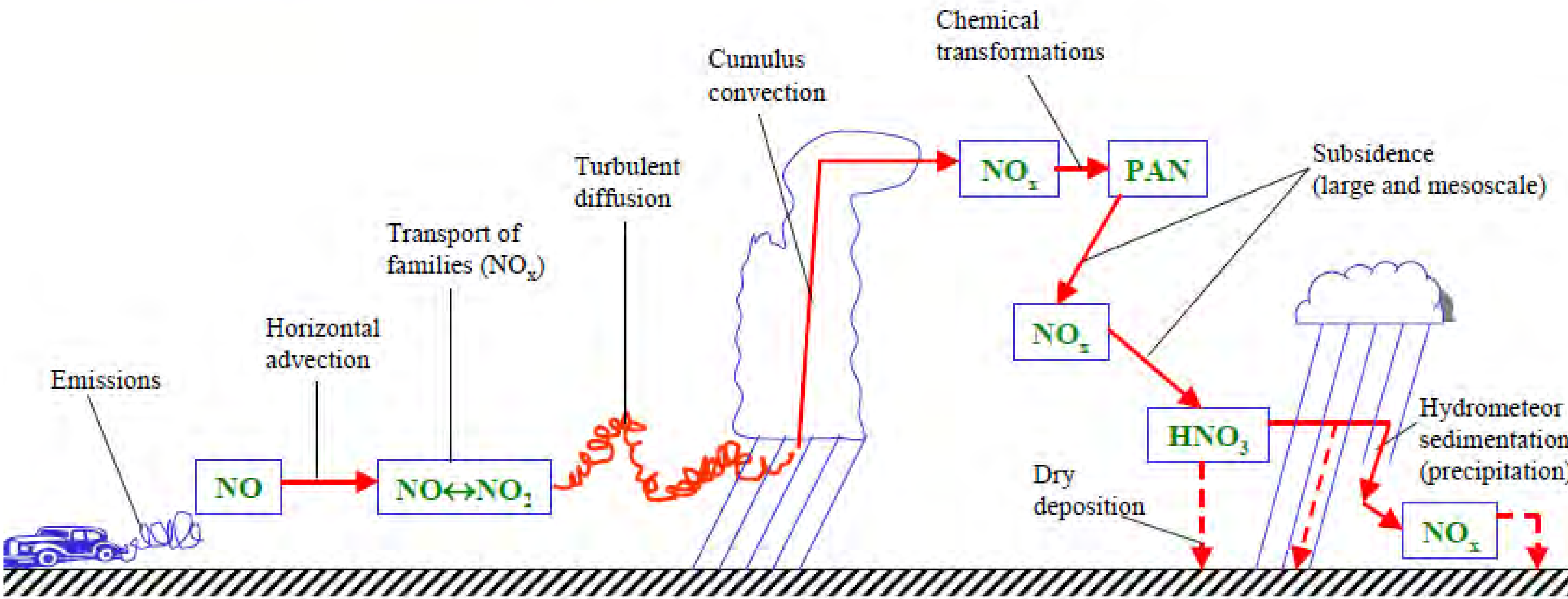
THE UNIVERSITY OF IOWA

Models Play a Critical Role in Linking Emissions to Aerosol and Trace Gas Distributions and Subsequent Effects

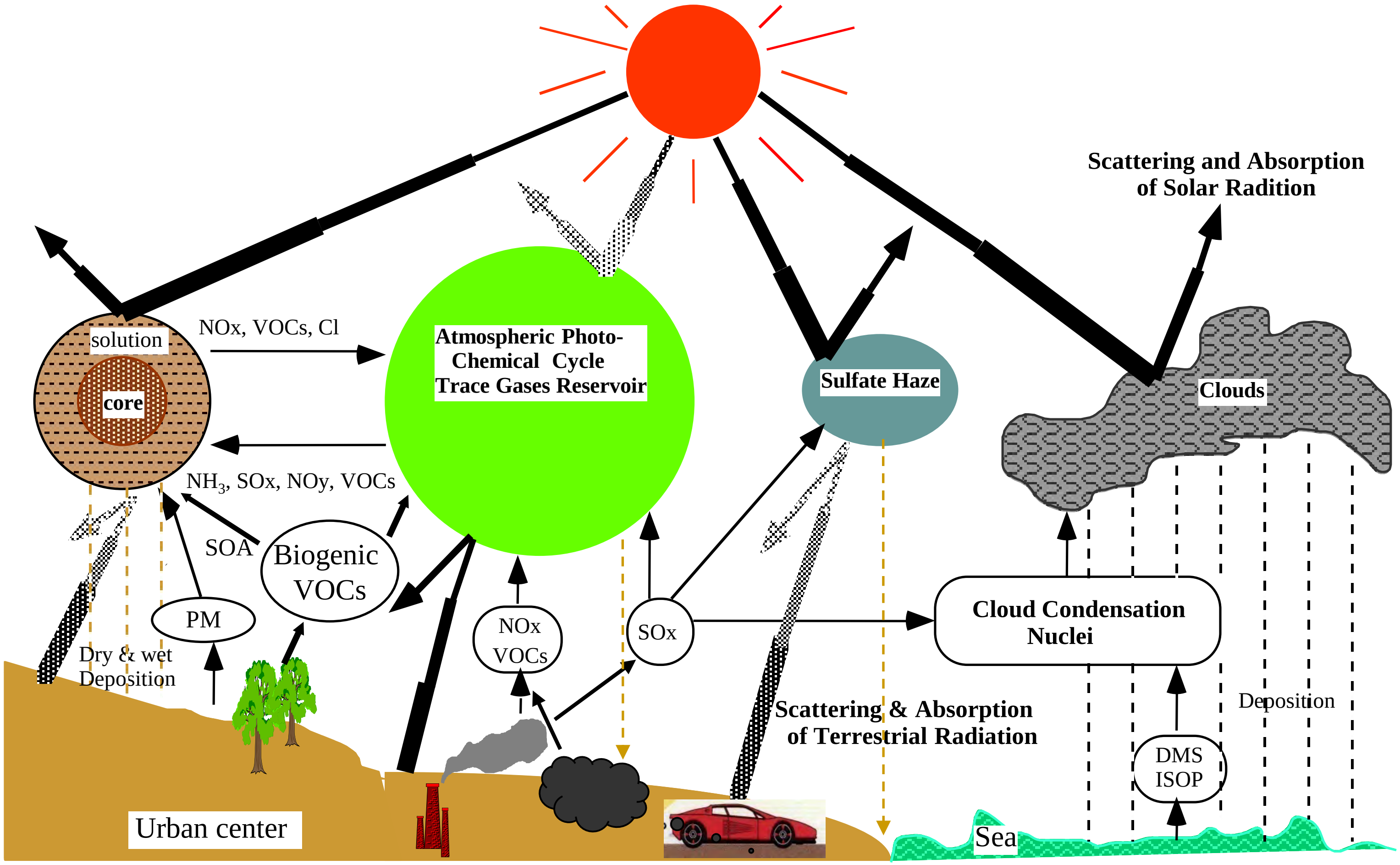


Factors Controlling Tracer Distributions

Example: Reactive Nitrogen



Atmosphere Interactions:
Gases, Aerosols, Chemistry, Transport, Radiation, Climate



After Y. Zhang, Copenhagen, 2007

Temperature → chemistry → concentrations → radiative processes → temperature
Aerosol → radiation → photolysis → chemistry
Temperature gradients → turbulence → surface concentrations, boundary layer outflow/inflow
Aerosol → cloud optical depth through influence of droplet number on mean droplet size → initiation of precipitation
Aerosol absorption of sunlight → cloud liquid water → cloud optical depth

Chemical Transport Model

- 3D atmospheric transport-chemistry model (STEM-III)

$$\frac{\partial c_i}{\partial t} = -u \cdot \nabla c_i + \frac{1}{\rho} \nabla \cdot (\rho K \nabla c_i) + f_i(c) + E_i$$

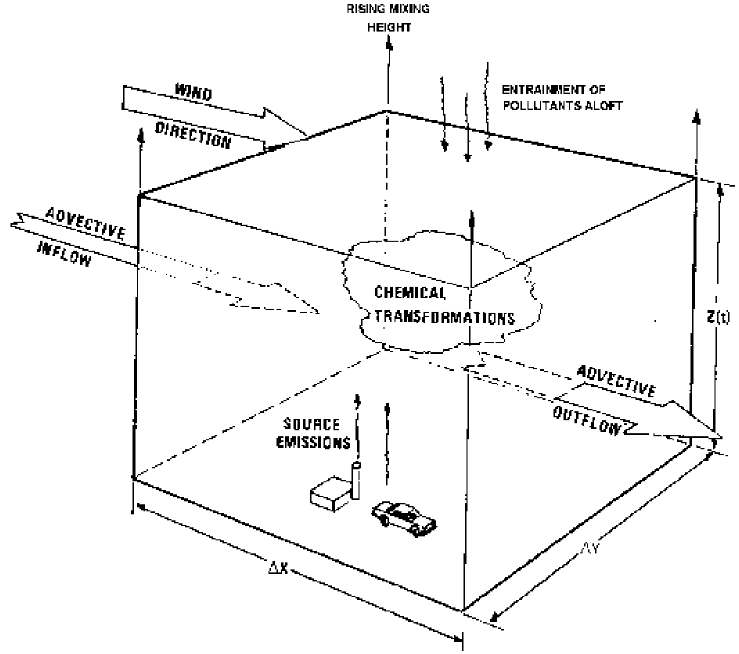
where chemical reactions are modeled by nonlinear stiff terms

$$f_i(c) = P_i(c) - D_i(c) c_i$$

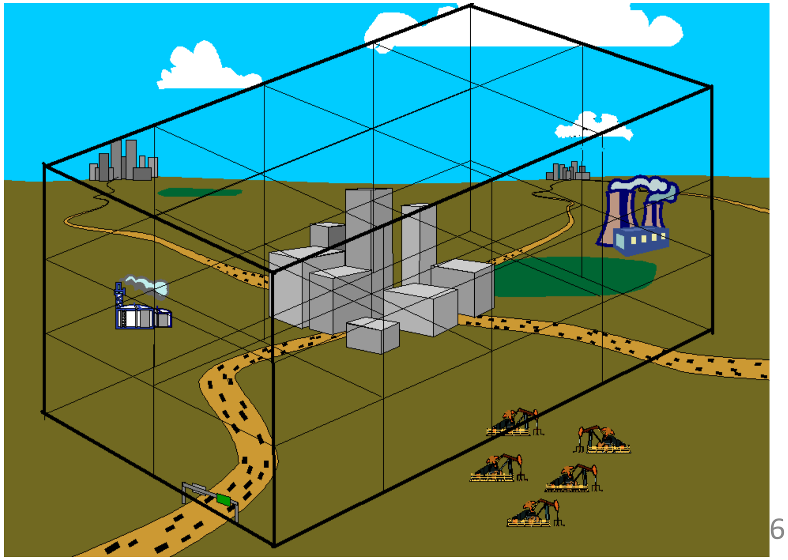
- Use operator splitting to solve CTM

$$\mathrm{M}_{[t, t+\Delta t]} = T_X^{\Delta t/2} \cdot T_Y^{\Delta t/2} \cdot T_Z^{\Delta t/2} \cdot C^{\Delta t} \cdot T_Z^{\Delta t/2} \cdot T_Y^{\Delta t/2} \cdot T_X^{\Delta t/2}$$

Numerical Air Quality Modeling

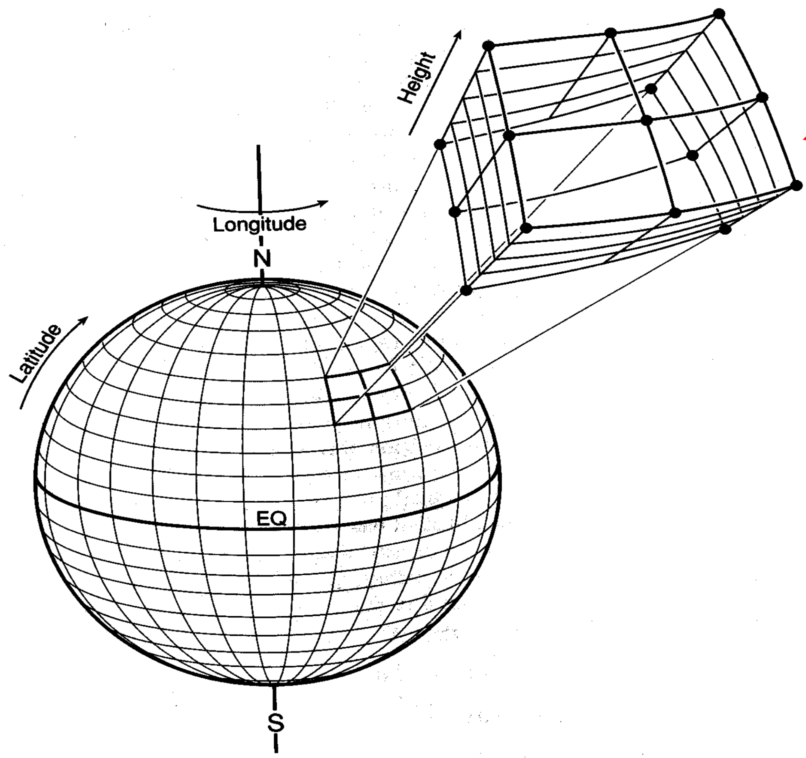


- Mathematically represents the important processes that affect pollution
- Requires a system of models to simulate the emission, transport, diffusion, transformation, and removal of air pollution
  – Meteorological forecast models
  – Emissions models
  – Air quality models



EULERIAN MODELS PARTITION ATMOSPHERIC DOMAIN INTO GRIDBOXES

This discretizes the continuity equation in space



Solve continuity equation for individual gridboxes

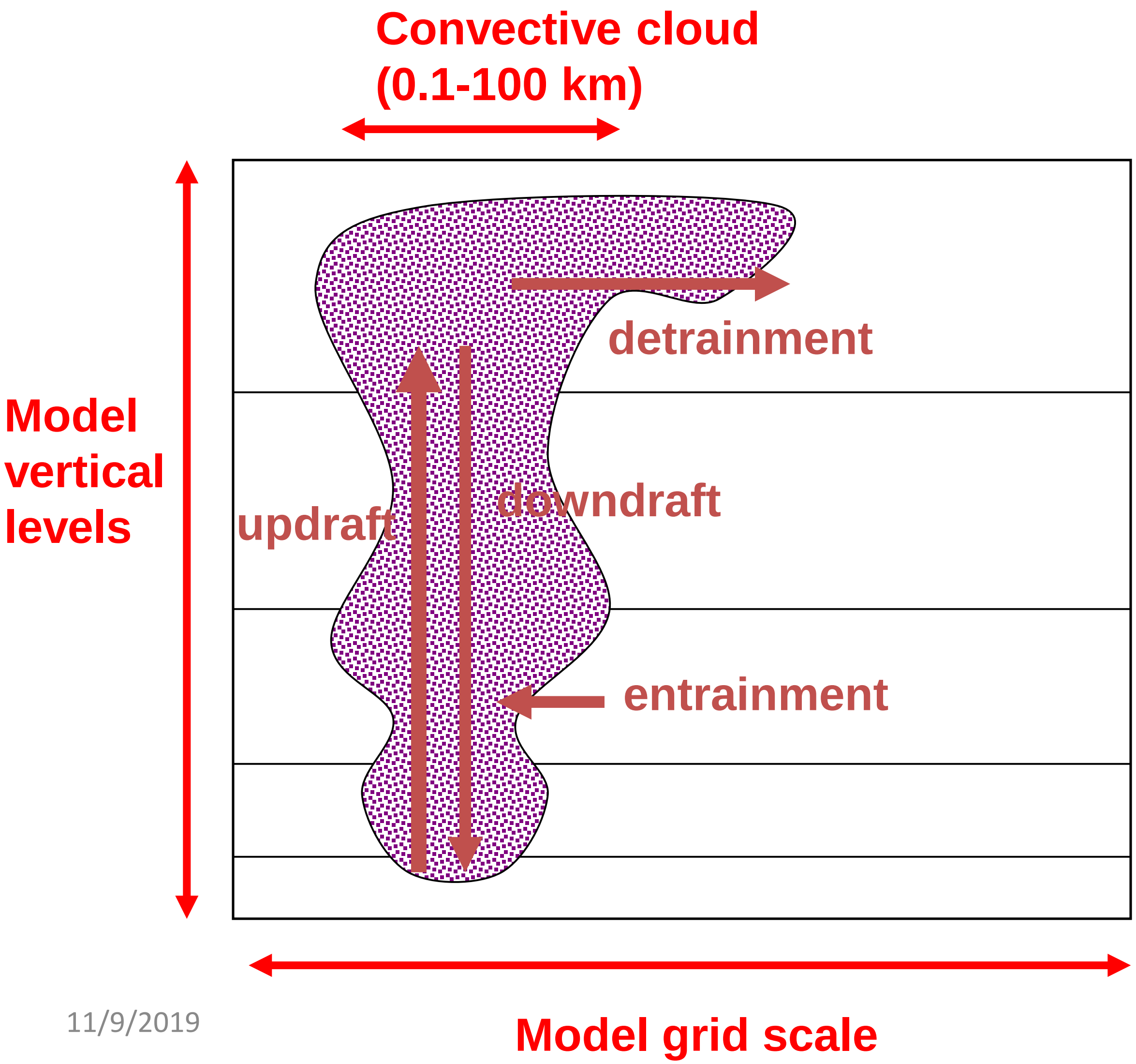
- **Detailed chemical/aerosol models can presently afford -10^6 gridboxes**
- **In global models, this implies a horizontal resolution of ~ 0.5-1^o (~50 to 100 km) in horizontal and ~ 0.5-1 km in vertical**

- **Chemical Transport Models (CTMs) use external meteorological data as input (or run on-line)**
- **General Circulation Models (GCMs) compute their own meteorological fields**

From D. Jacob

VERTICAL TURBULENT TRANSPORT (BUOYANCY)

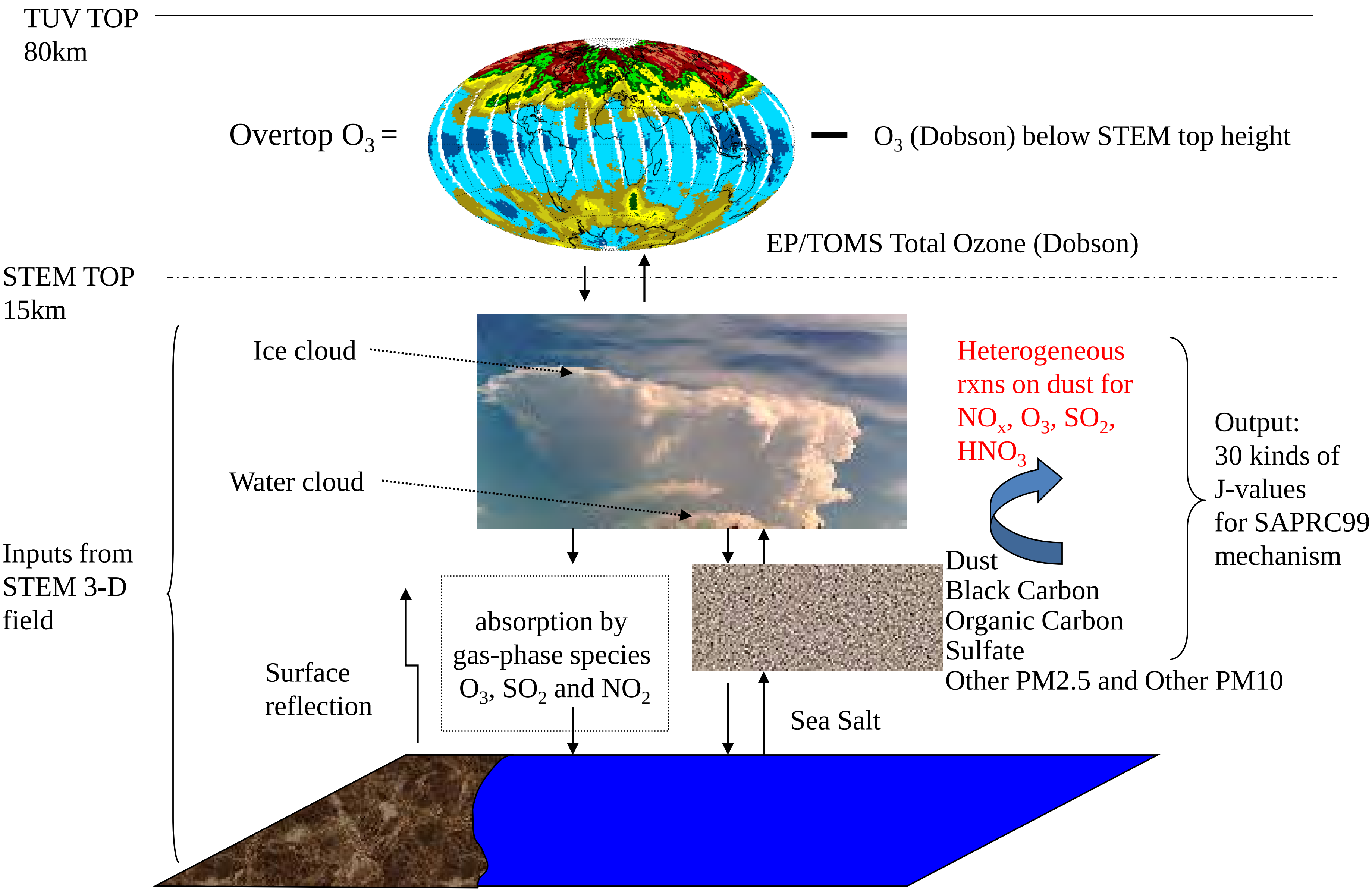
- generally dominates over mean vertical advection
- K-diffusion OK for dry convection in boundary layer (small eddies)
- Deeper (wet) convection requires non-local convective parameterization



Wet convection is subgrid scale in global models and must be treated as a vertical mass exchange separate from transport by grid-scale winds.

Need info on convective mass fluxes from the model meteorological driver.

Aerosols Are a Key Component in Urban Environments -- Impacting Chemistry and Physics

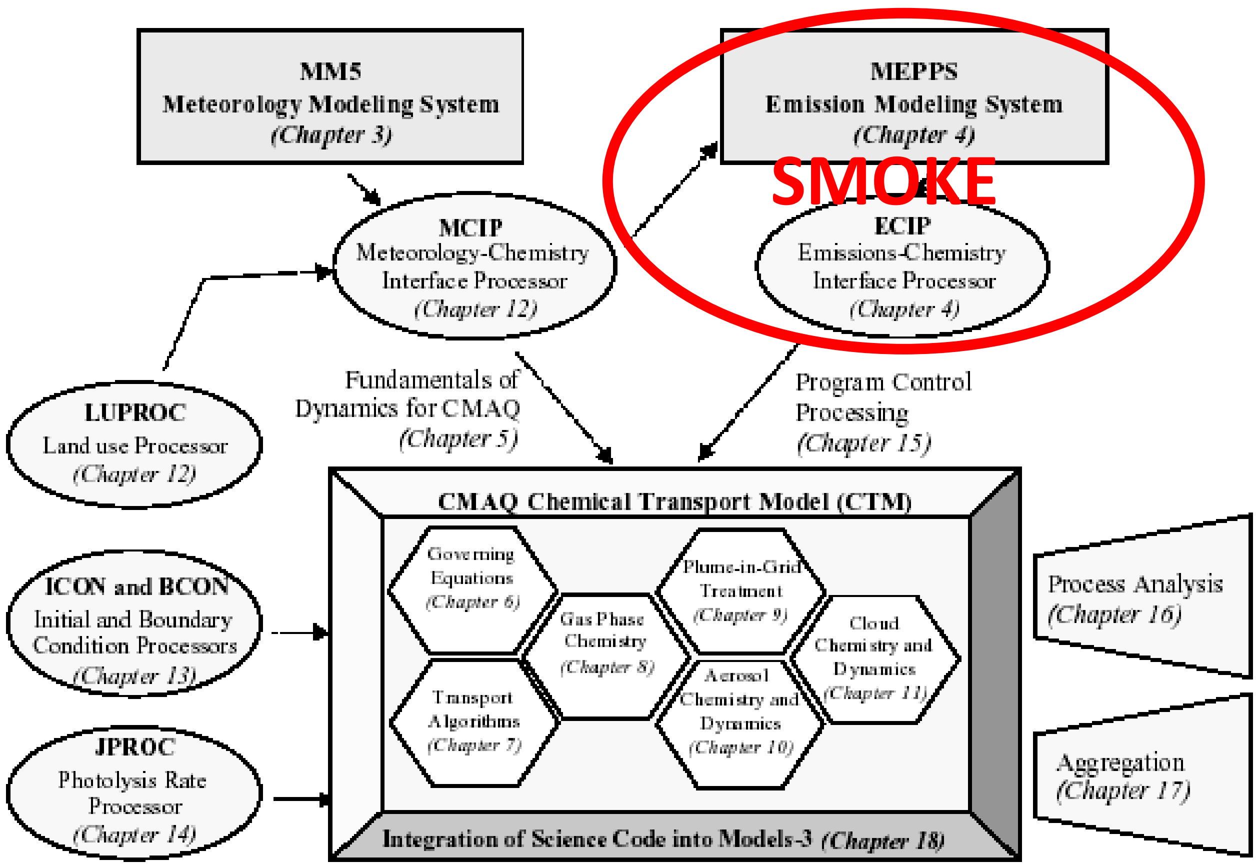


Models are an Integral Part of Air Quality Studies

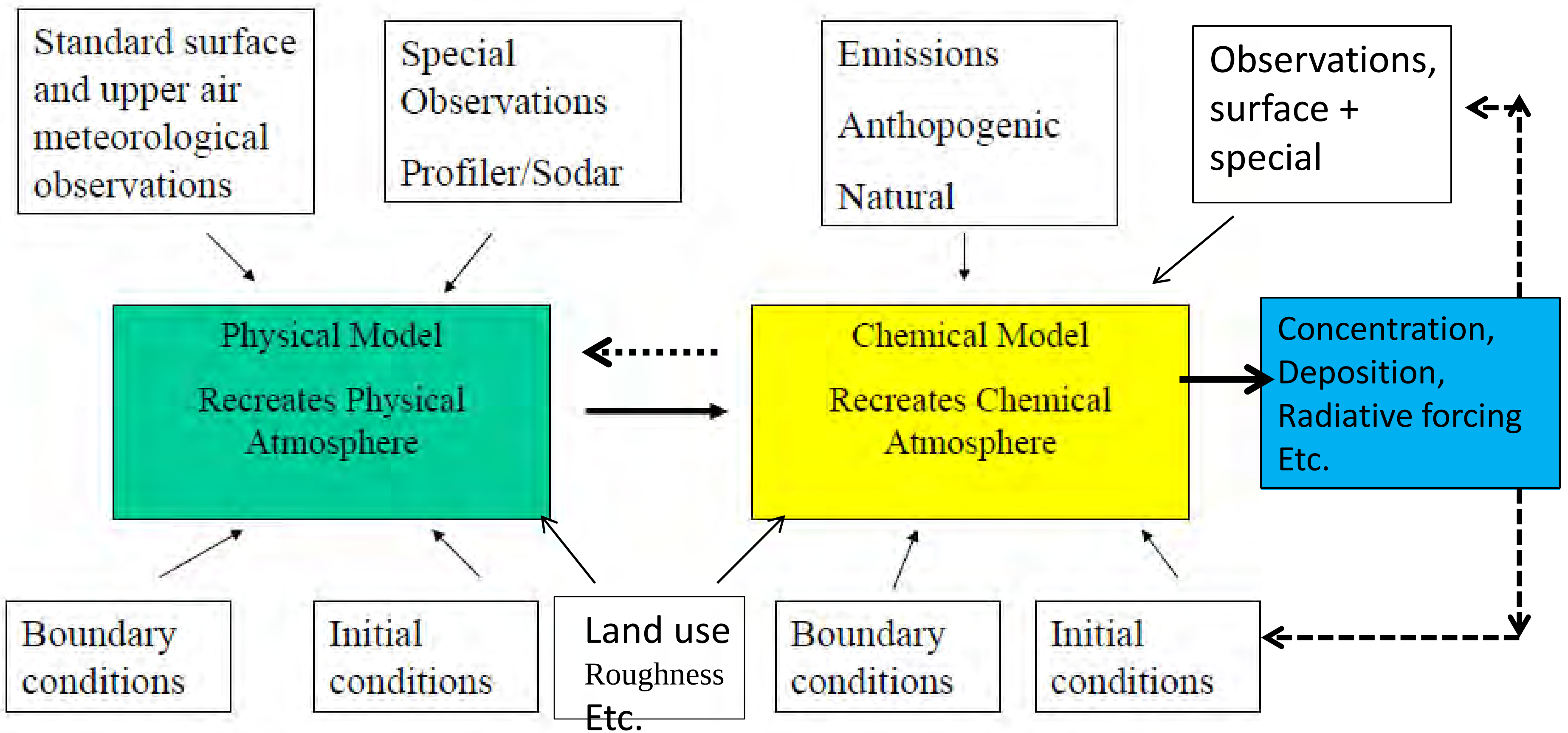
- **Field experiment planning**
- **Provide 4-Dimensional context of the observations**
- **Facilitate the integration of the different measurement platforms**
- **Evaluate processes (e.g., role of biomass burning, heterogeneous chemistry….)**
- **Evaluate emission estimates (bottom-up as well as top-down)**
- **Emission control strategies testing**
- **Air quality forecasting**
- **Measurement site selection**

CMAQ Modeling System

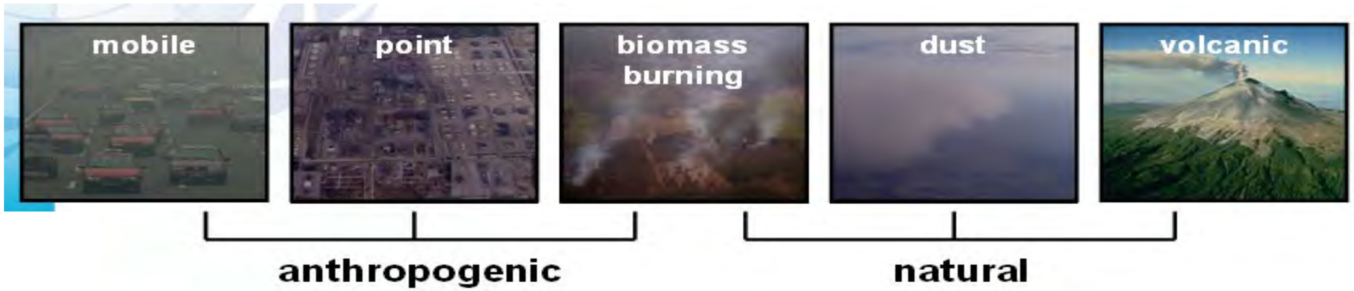
- Model used by US EPA to forecast the AQI
- Other than usual chemicals and aerosols, can predict visibility, acid depositions and toxics
- A WRF-CMAQ online version is under development

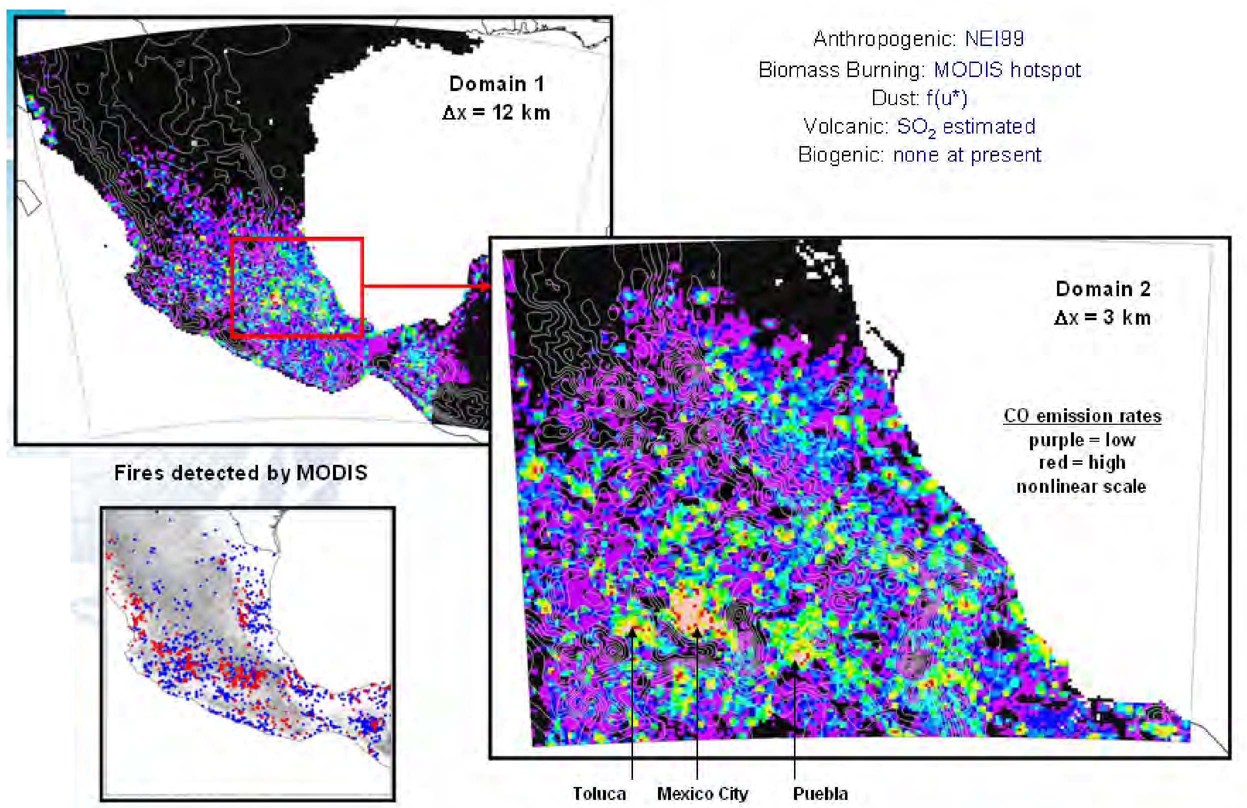


Coupling Weather and Air Quality Prediction



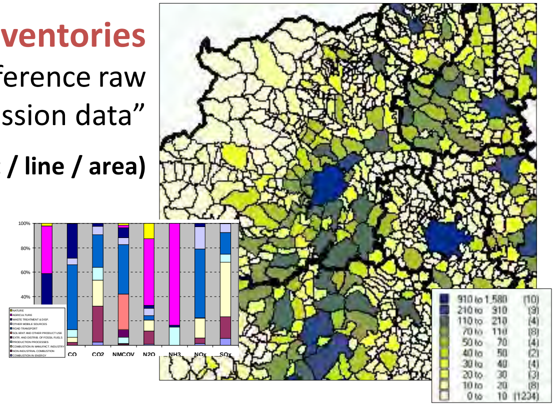
Need to Estimate **Emissions** at Appropriate Scales



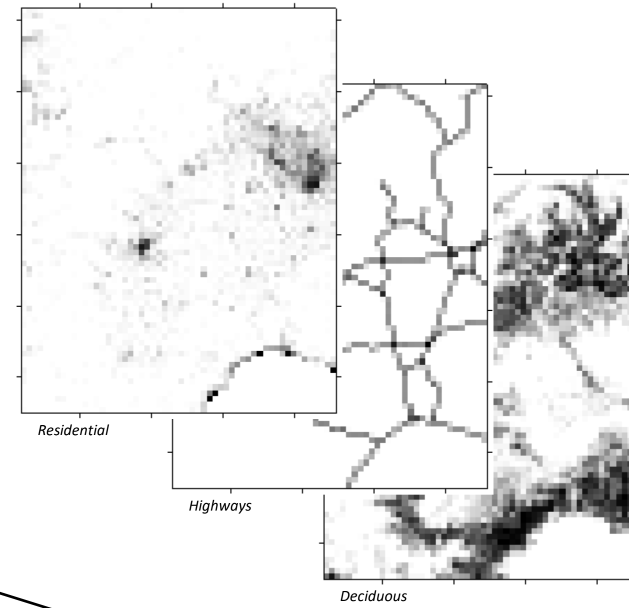


Emissions processing for AQM

Inventories
“reference raw emission data”
(point / line / area)



Thematic data

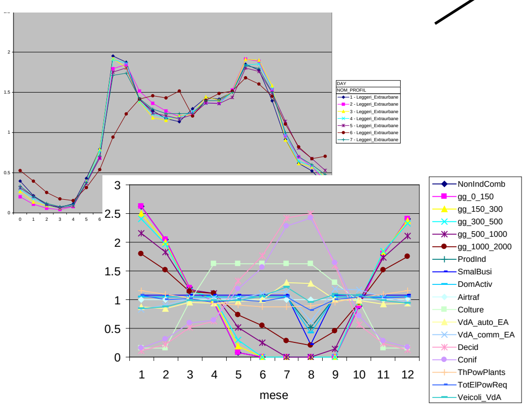


SPACE DISAGGREGATION

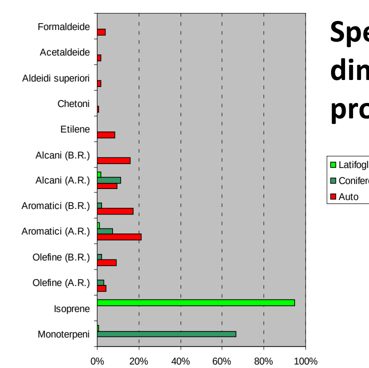
TIME MODULATION

NMVOC & PM SPECIATION & SIZE

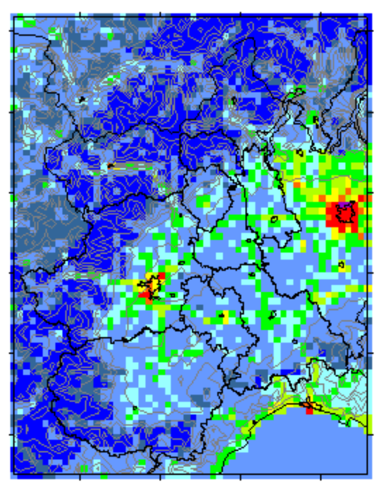
Modulation profiles
(hourly, daily, monthly)



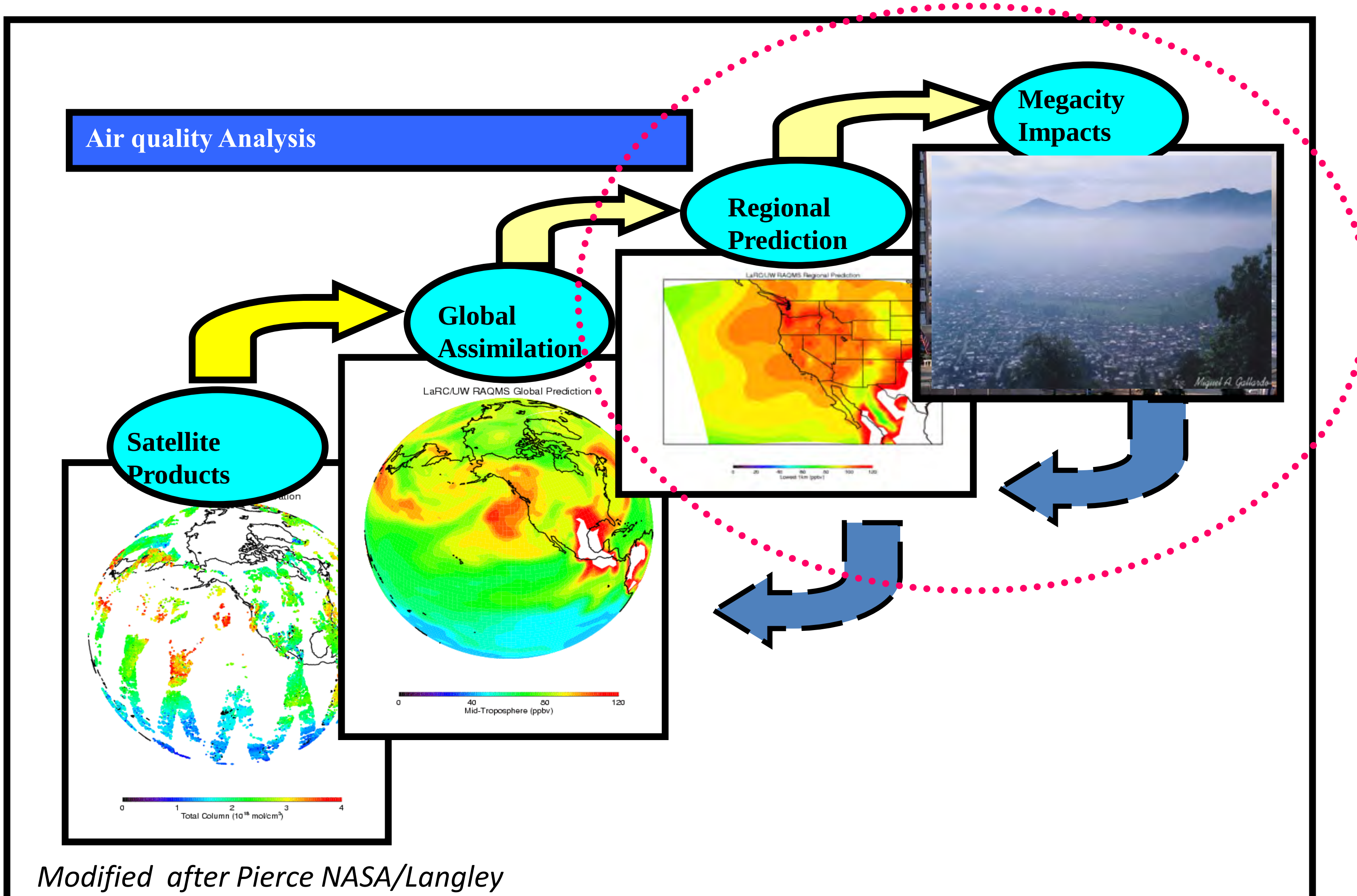
Speciation & dimensional profiles



Model-ready input
(hourly, gridded, speciated emissions)

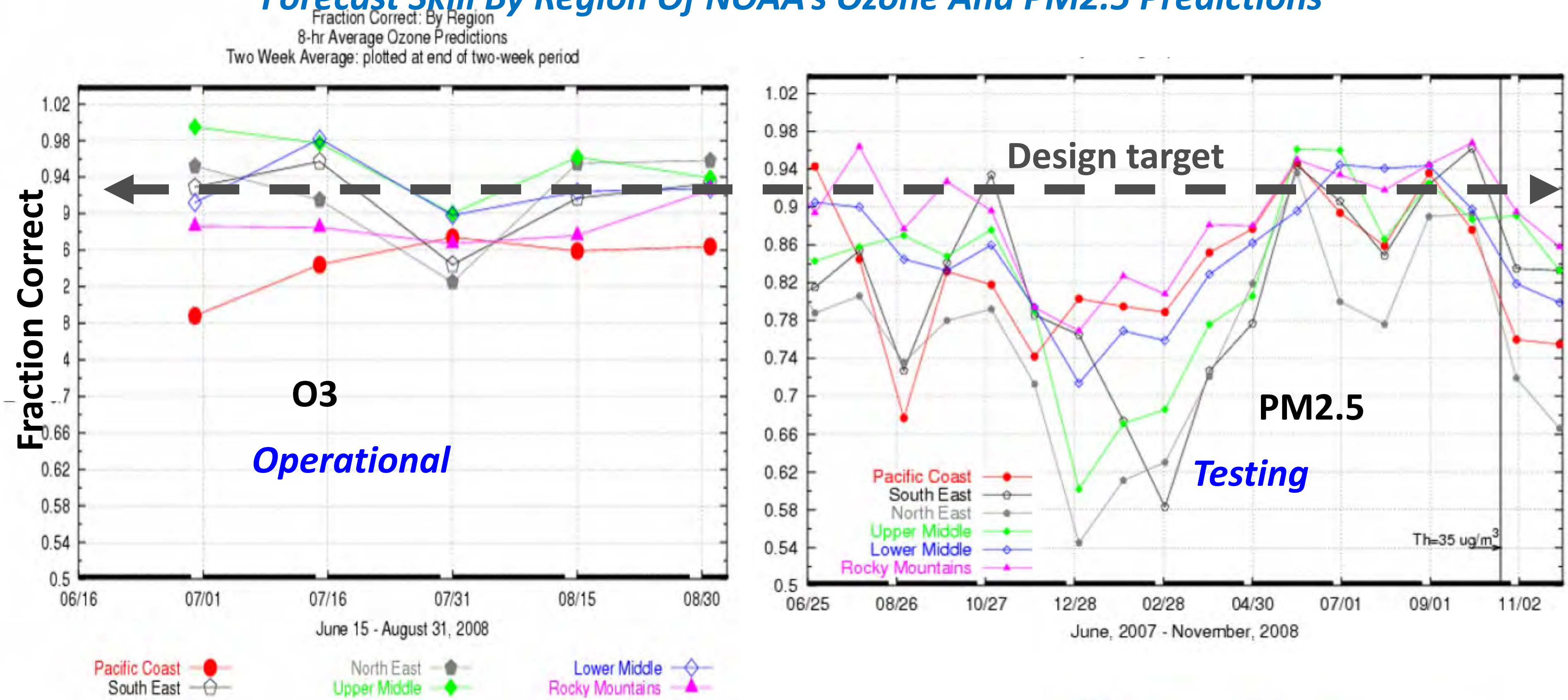


Air Quality Prediction: A Challenge of Scales and Integration



Increasing Needs and Expectations for Chemical Weather Forecasts

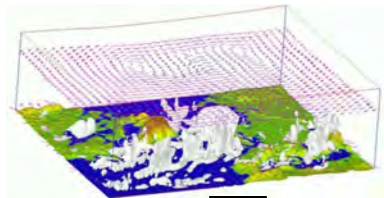
Forecast Skill By Region Of NOAA's Ozone And PM2.5 Predictions



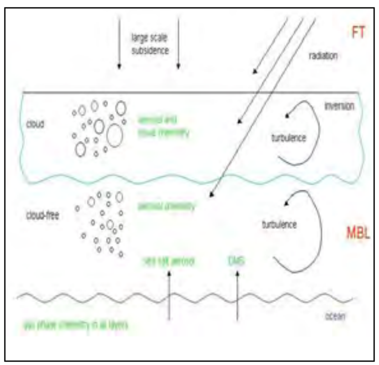
Slide provided by Paula Davidson

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

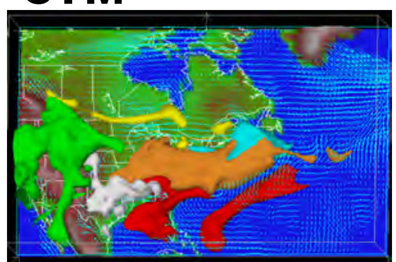
Met model



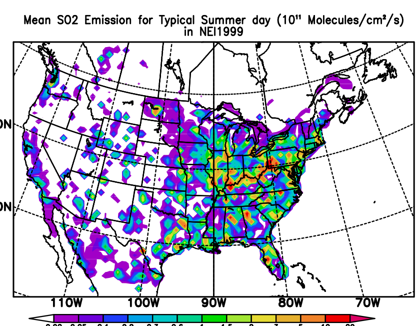
Chemical, Aerosol, Removal modules



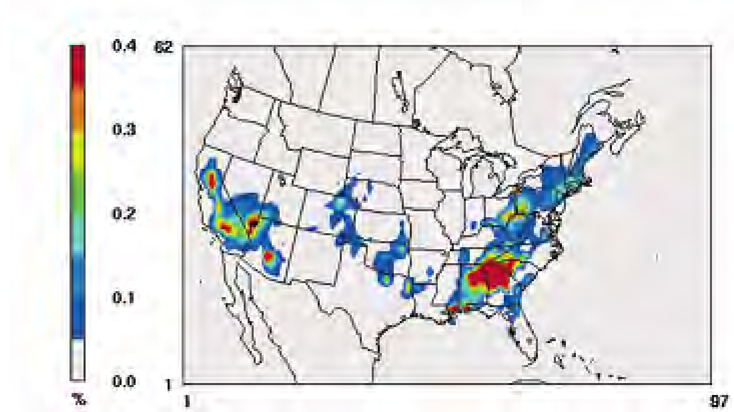
CTM



Emissions

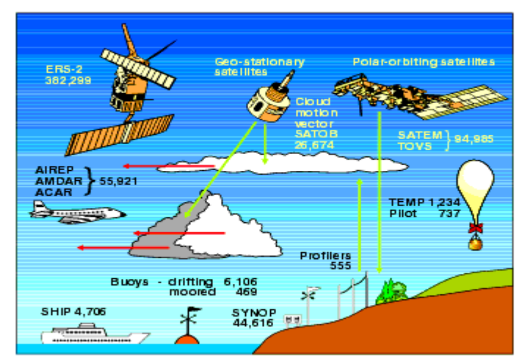


Predicted Quantity: e.g., *ozone AQ violation*



How confident are we in the models & predictions?

Observations



Major Sources of Errors and Uncertainties

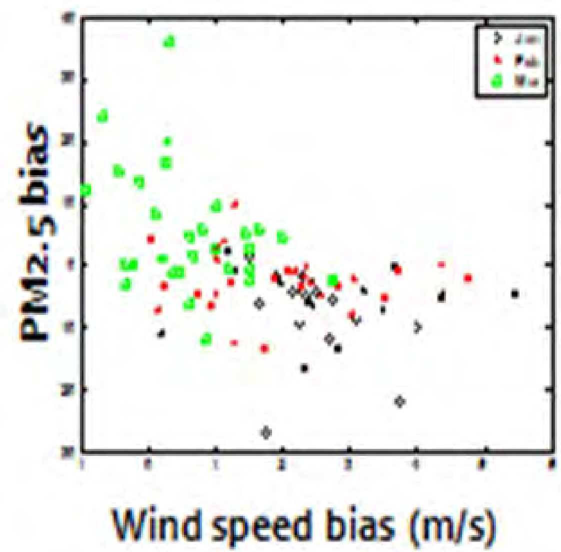
•Emissions

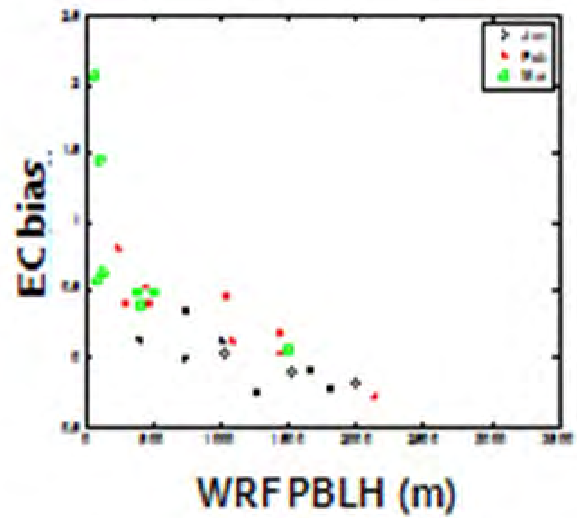
•Key meteorological aspects – pbl, clouds, temperature, water vapor, precip …

•Chemical processes – Secondary Organic & Inorganic Aerosol, nighttime chemistry, …

•Boundary conditions

•Other key model inputs, land use, surface roughness, anthro heat flux….

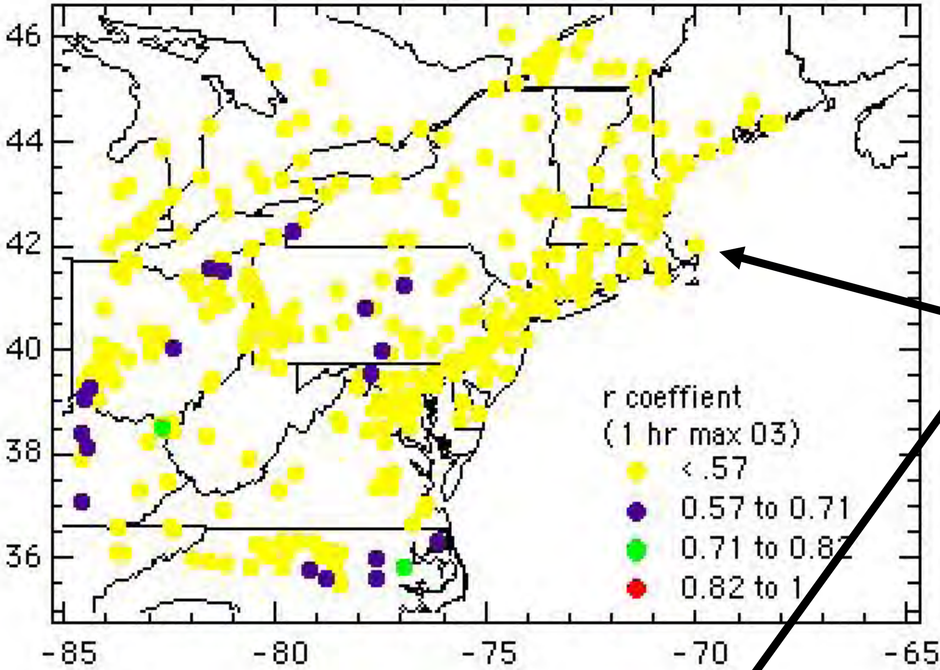




Improving Predictions

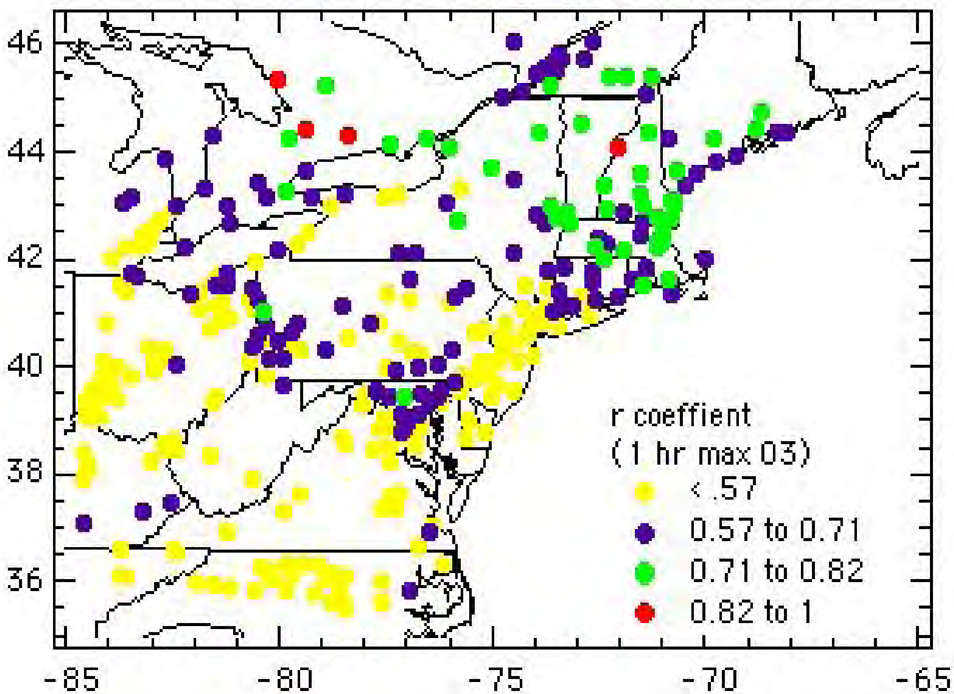
- Bias correction

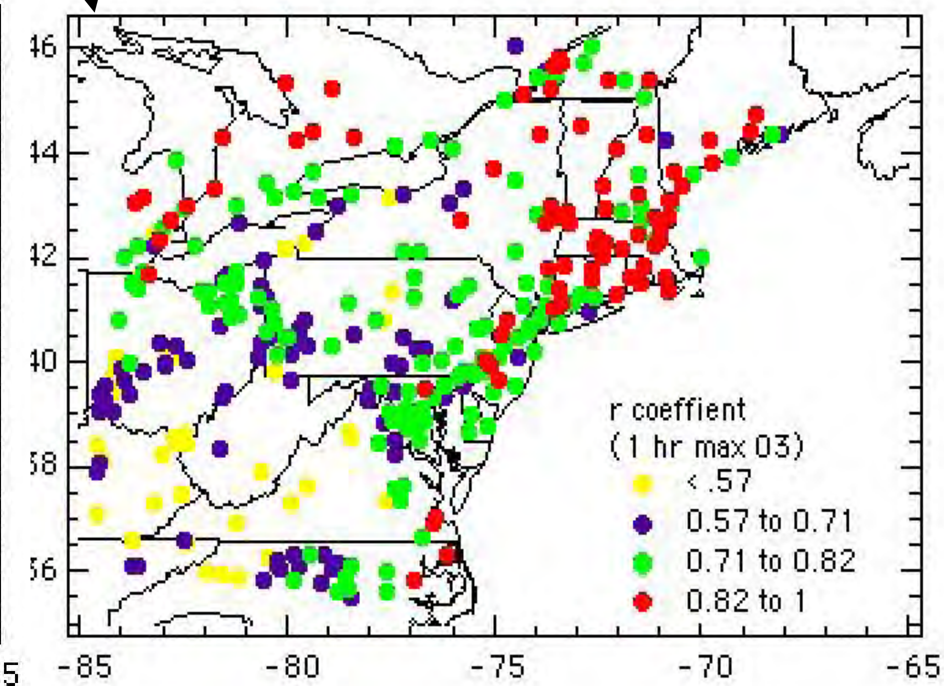
Forecasting Air Quality an Important Activity in Air Quality Management



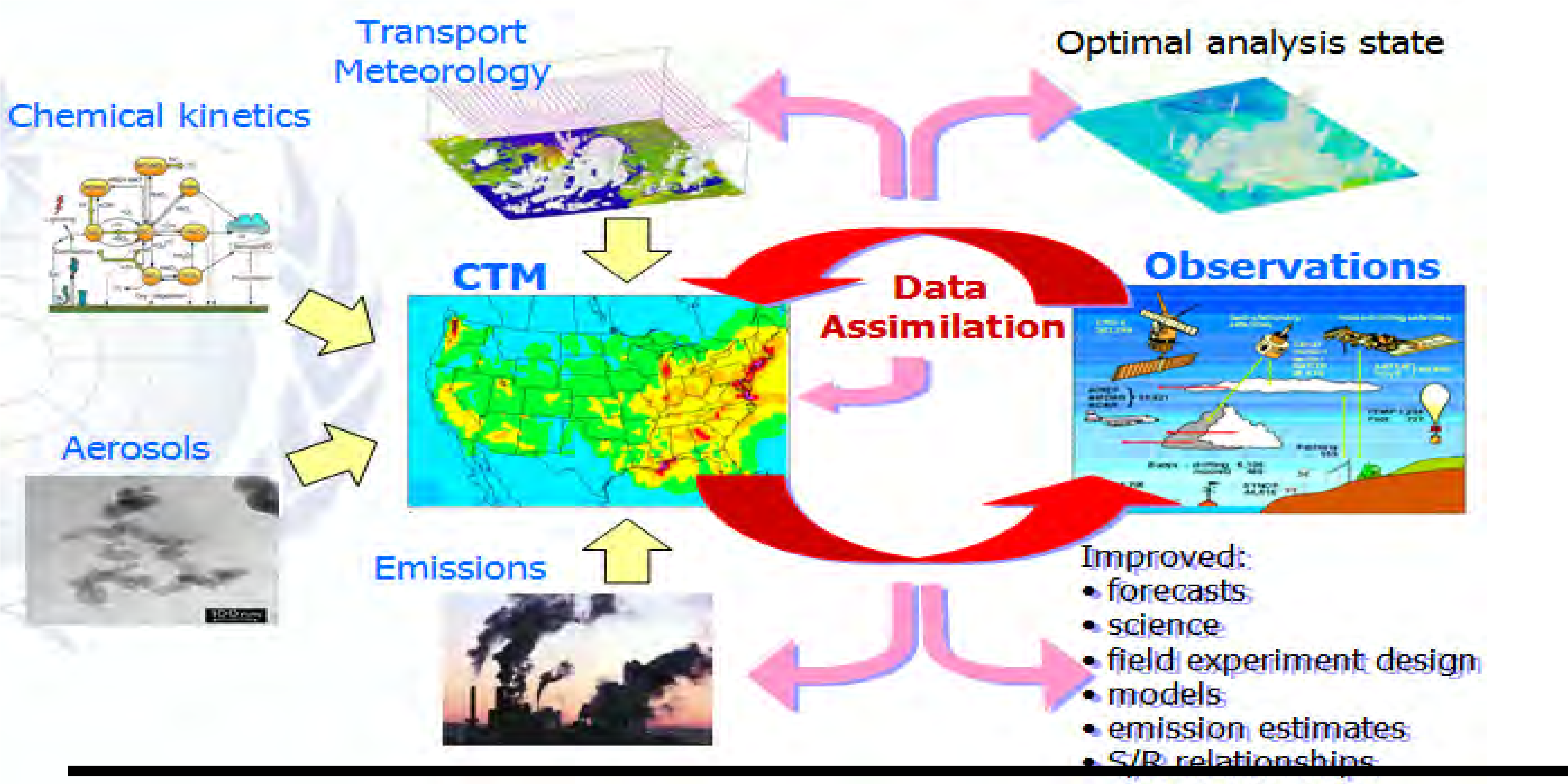
* Persistence
* Single Forward Model w/o assimilation
* Ensemble forecast (8 models) w/o assimilation (*further improvements with bias corrections based on obs*)

McKeen et al., JGR, 2005





Challenge: Achieving A Closer Integration Of Observations And Models



+ Need to Integrated Air Quality & Met. Model assimilation systems

+ New requirements for NRT data, observing systems, and assimilation systems for chemical applications!!

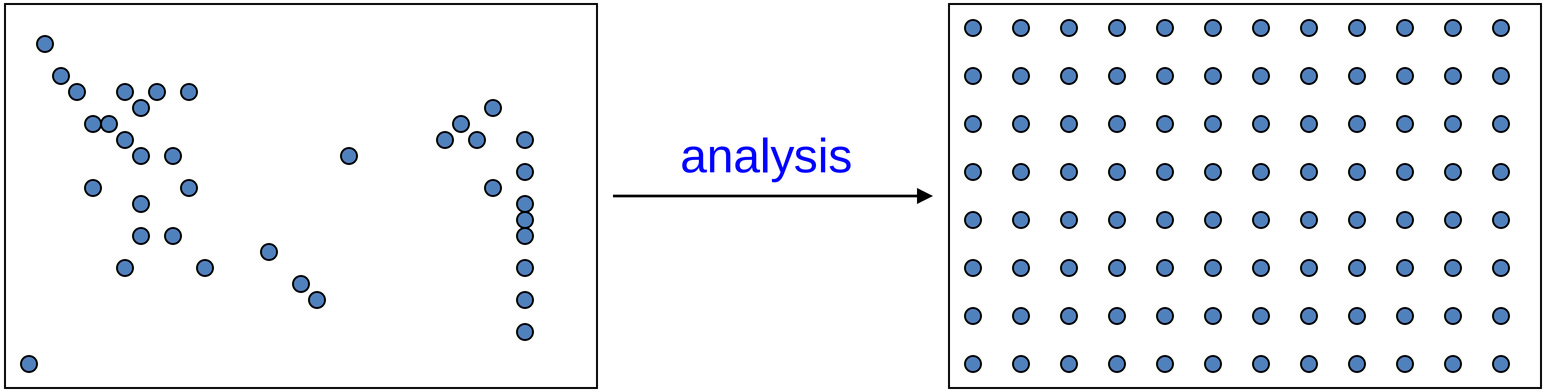
Data assimilation: Model + Observation

To understand and/or forecast air pollution, we need

- Measurements, samplings of the ‘reality’
- Chemical Transport Models (CTMs), describing the physical and chemical processes
- Data assimilation techniques, optimally integrate models and observations
  - Background (a priori) error statistics
  - Observation error statistics
  - Model error information

Atmospheric Data Analysis

Goal: To produce a regular, physically consistent, four-dimensional representation of the state of the atmosphere from a heterogeneous array of in-situ and remote instruments which sample imperfectly and irregularly in space and time. (Daley, 1991)



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
- Strongly source driven and they are highly uncertain
  - Inventories often out-dated, and uncertainty not well-quantified, effect of initial conditions can quickly decrease

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)
  - Hybrid Methods

Optimal Interpolation (OI) Algorithm

- **Assimilation: A methodology to optimize state and evolution of the system using model predictions with observational constraints**

- **Collins et al. (2001)**

$$\tau'_m = \tau_m + K\left(\tau_o - H\tau_m\right)$$

$$K = BH^T\left(HBH^T + O\right)^{-1}$$

$$O = \left(f_o\tau_o + \varepsilon_o\right)^2 I$$

$$B_{ij} = \left(f_m\tau_m + \varepsilon_m\right)^2 \exp\left[-\frac{d_x^2 + d_y^2}{2l_{xy}^2}\right]$$

•K – Kalman gain matrix

•H – Interpolator from model to observation space

•O/B Observed/Background error covariances

•f and ε - fractional error and RMS uncertainty

•L_{xy} horizontal correlation length scale for errors in model fields

•d_x and d_y grid cell spacing

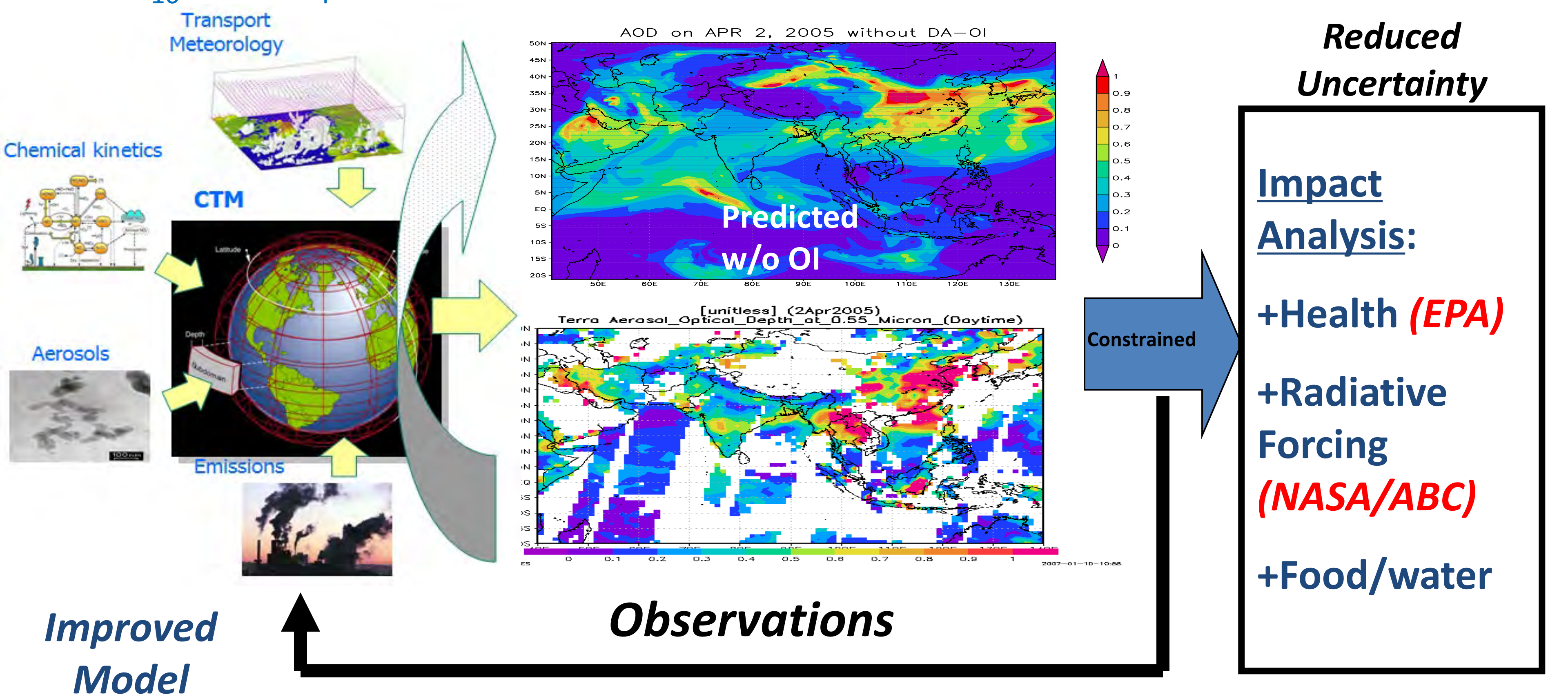
Collins, W. D., P. J. Rasch, B. E. Eaton, B. V. Khattatov, J.-F. Lamarque and C. S. Zender (2001). "Simulating aerosols using a chemical transport model with assimilation of satellite aerosol retrievals: Methodology for INDOEX." Journal of Geophysical Research, [Atmospheres] **106**(D7): 7313-7336.

Models Constrained With Observations Play Increasing Important Roles In Research and Application

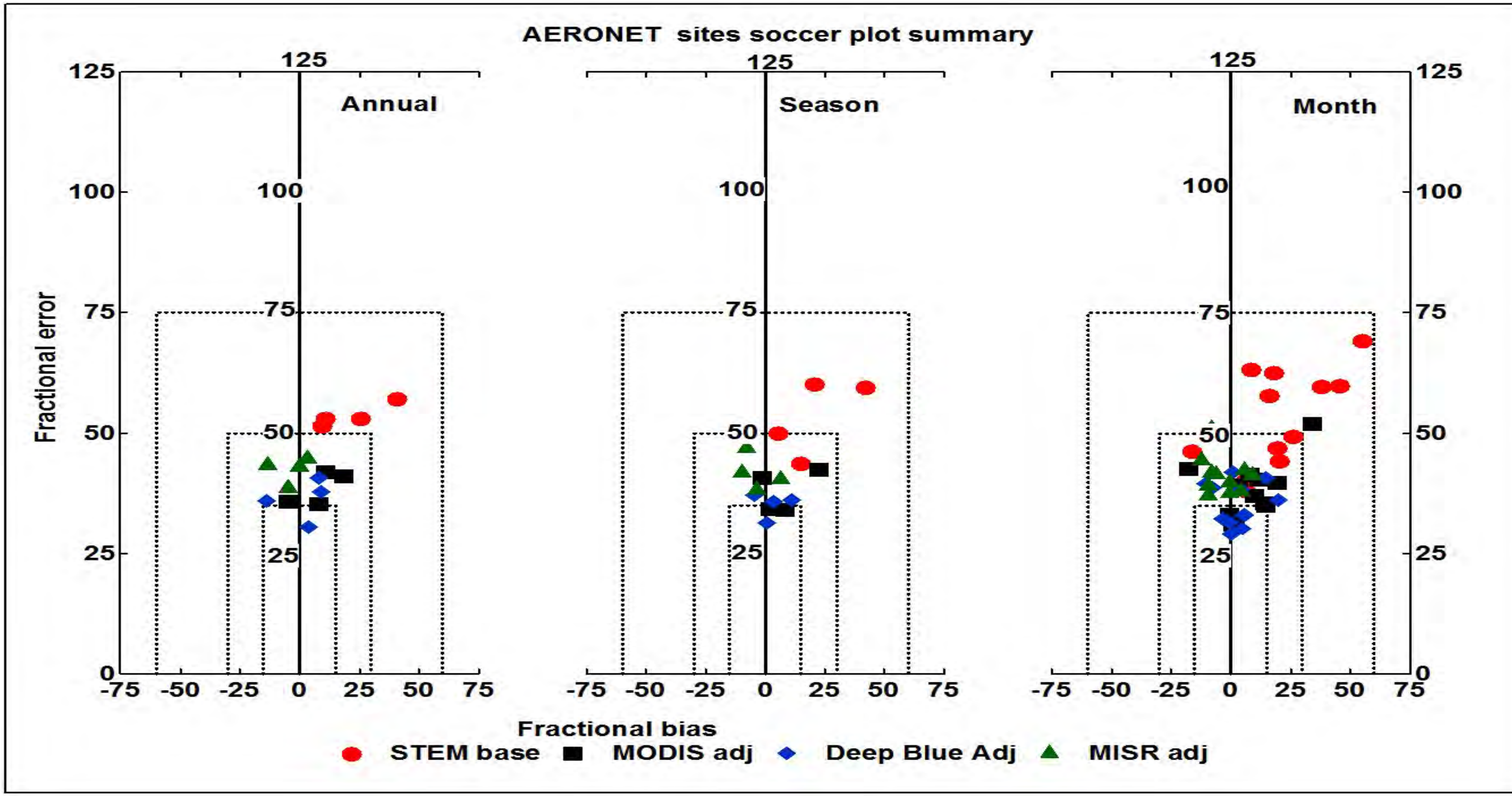
Assimilation method: optimal interpolation *(Adhikary et al., 2008, 2009)*

Data used in assimilation: MODIS, Deep Blue (5.1 L3), and MISR (CGAS L3); Total AOD, fine mode, AAOD

Validation against observations: AOD from AERONET using 2 different retrievals (V2 L2); PM_{10} and SO_4 from EANET

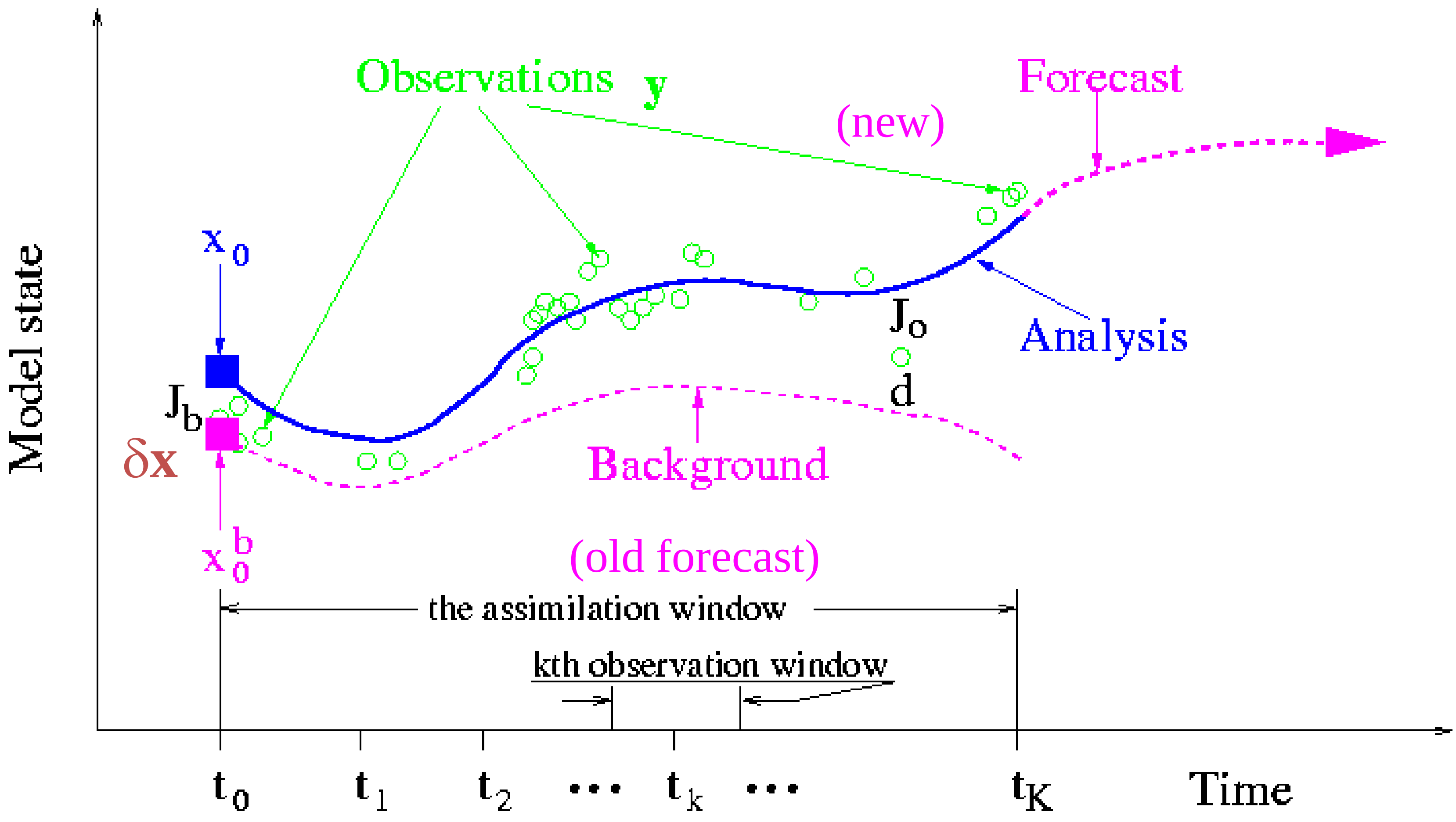


Total AOD - Assimilation Reduces Annual/Seasonal and Monthly Bias and Error



Annual/Seasonal/Monthly Soccer Plots across the domain *(Best configuration case)*

Data assimilation



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 φis 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. $\lambda_i = \frac{\partial J}{\partial c_i}$

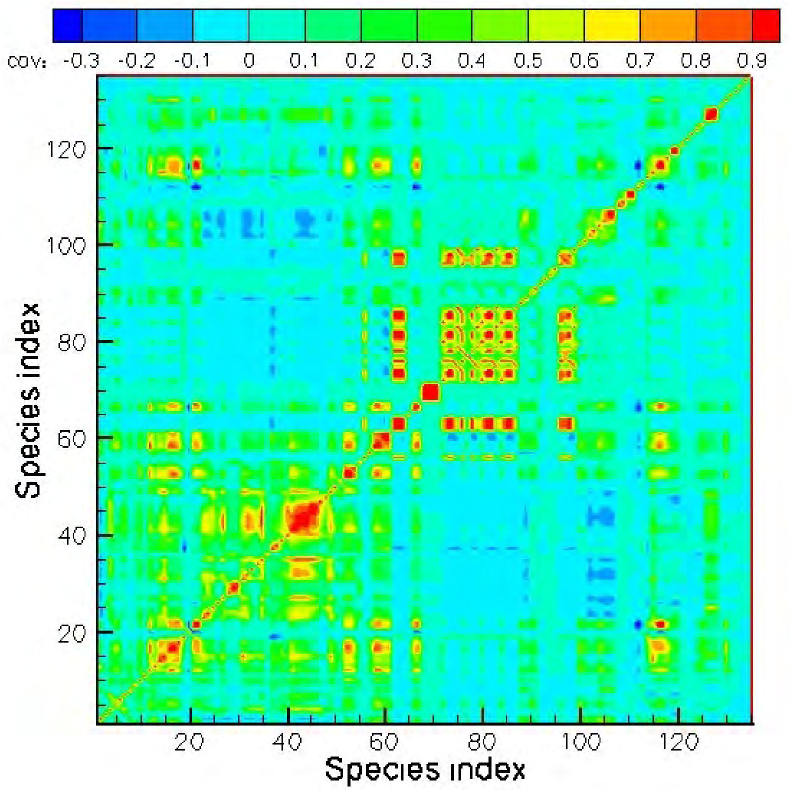
- Use adjoint variables for sensitivity analysis,   as well as data assimilation

Model Background Errors NMC method

- 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

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

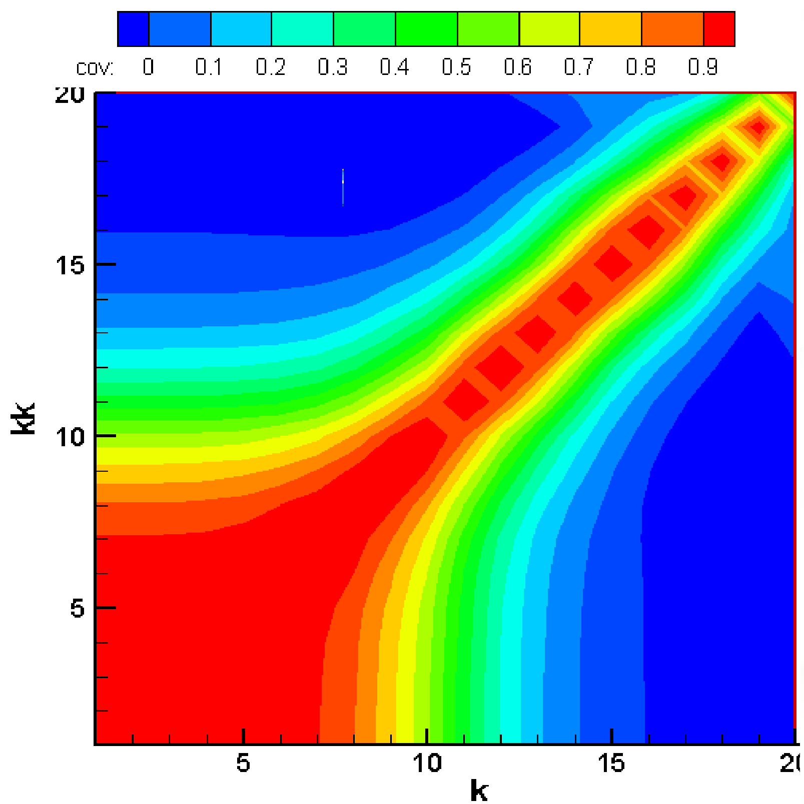


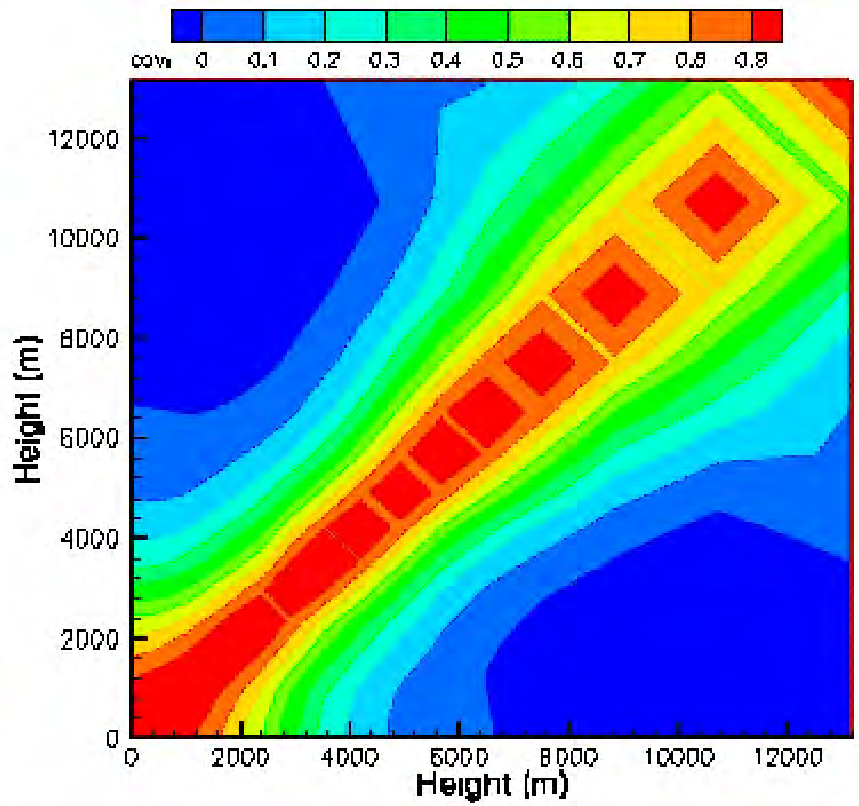
- Equivalent sample number: 811,890

NMC method results

$$CORR(K,KK) = \frac{\overline{\epsilon_K \cdot \epsilon_{KK}}}{\sqrt{\overline{\epsilon_K \cdot \epsilon_K}} \cdot \sqrt{\overline{\epsilon_{KK} \cdot \epsilon_{KK}}}}$$

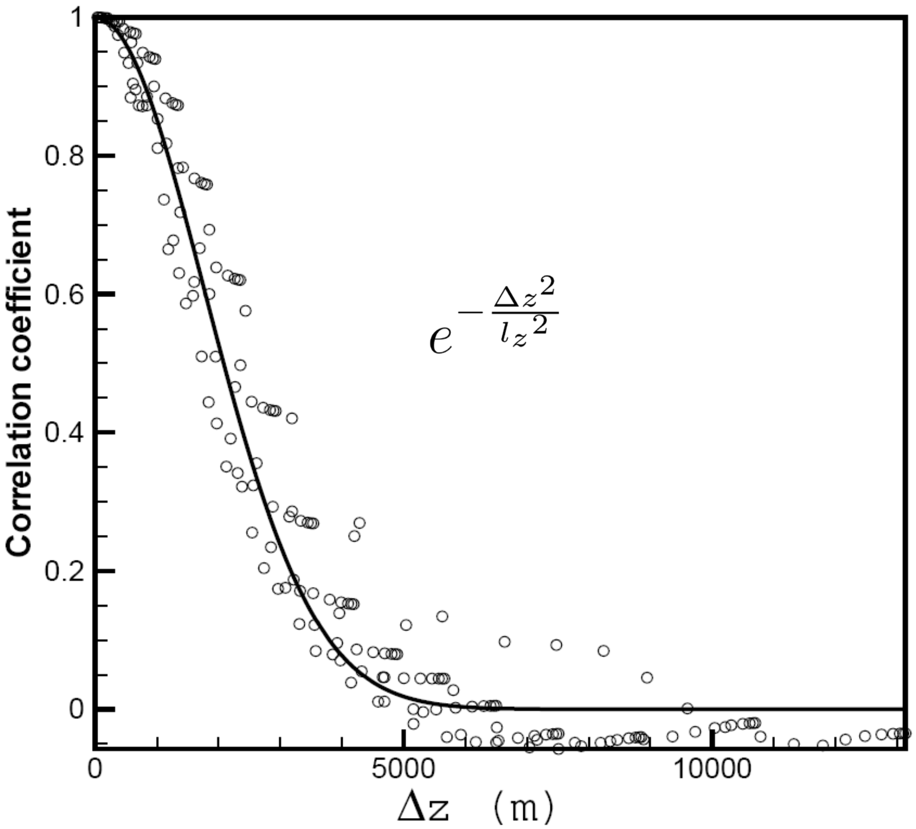
- Equivalent sample number: 360,840



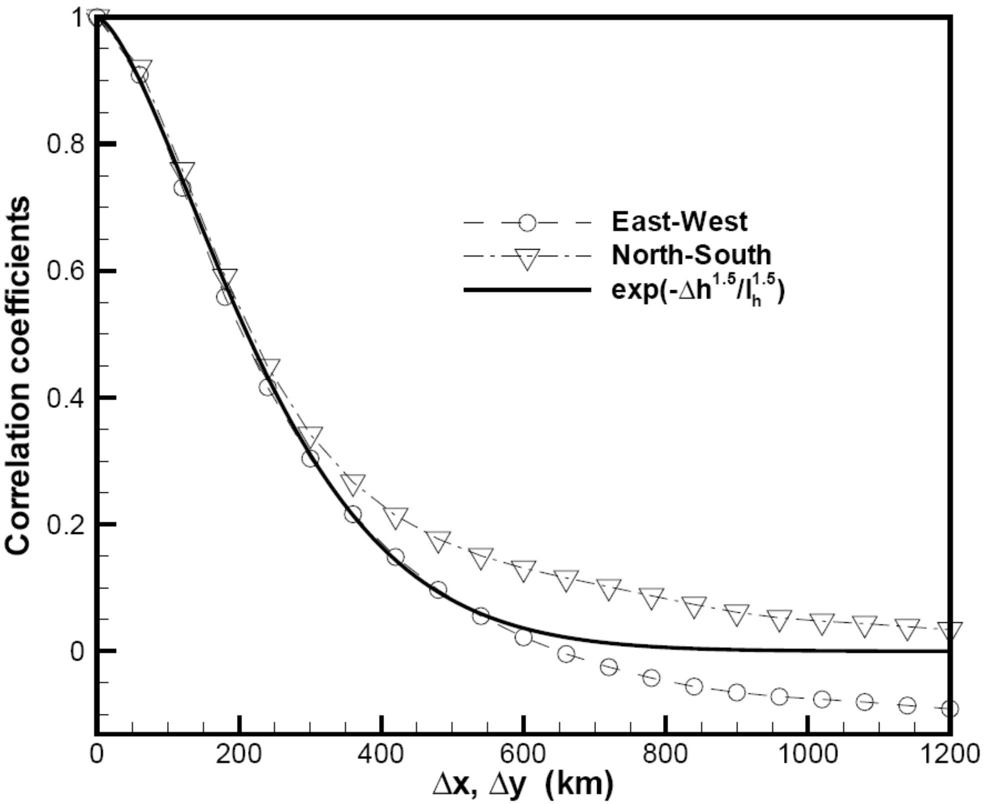


NMC method results

Vertical correlation

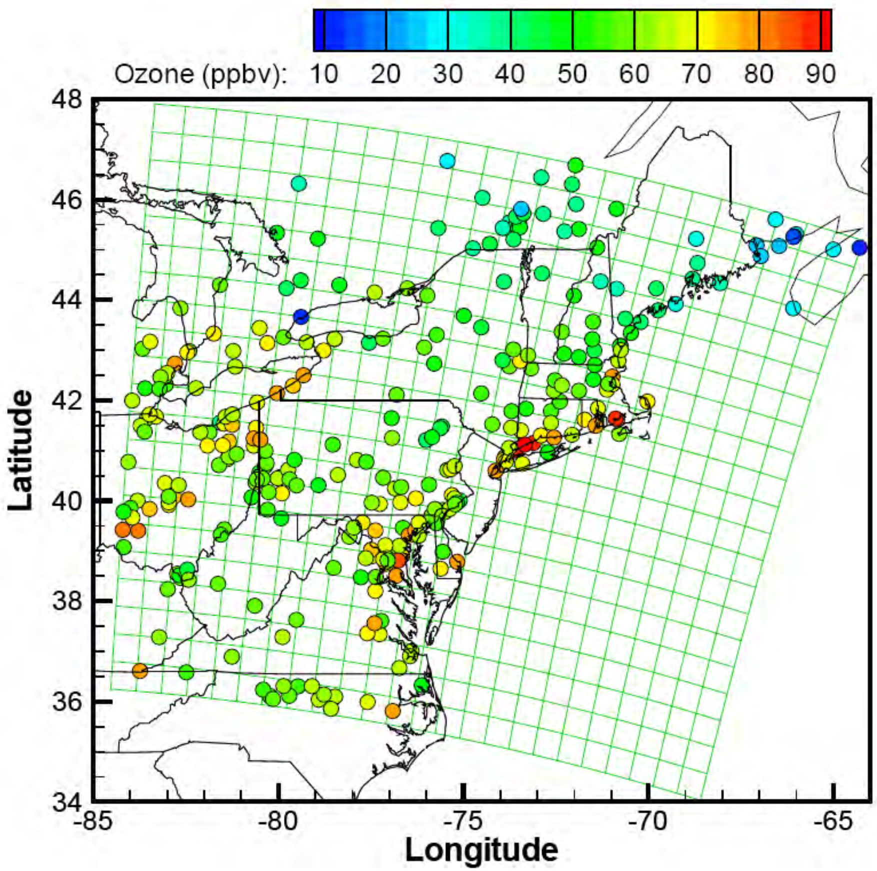


Horizontal correlation



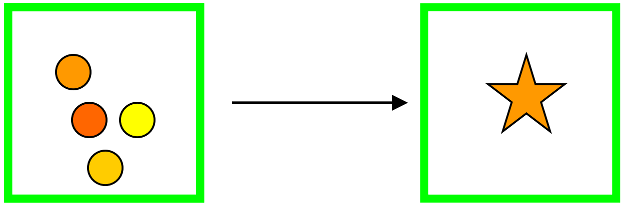
Observational error

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



Observational Error:

- Representative error
- Measurement error

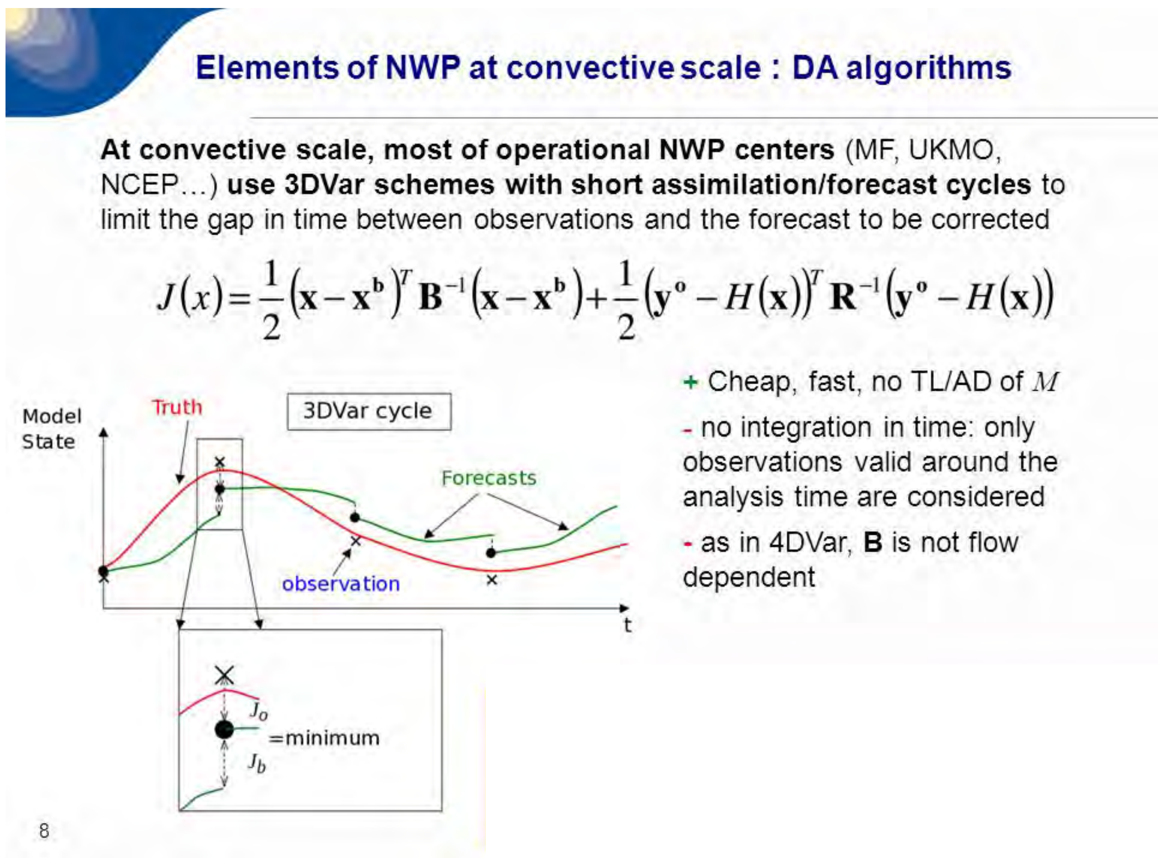


Observation Inputs

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

GSI AOD assimilation

- Gridpoint Statistical Interpolation (GSI) (3dVAR) can perform simultaneous DA of different datasets (e.g. ground PM2.5 and AOD, different AOD retrievals, and number) for WRF-Chem model.



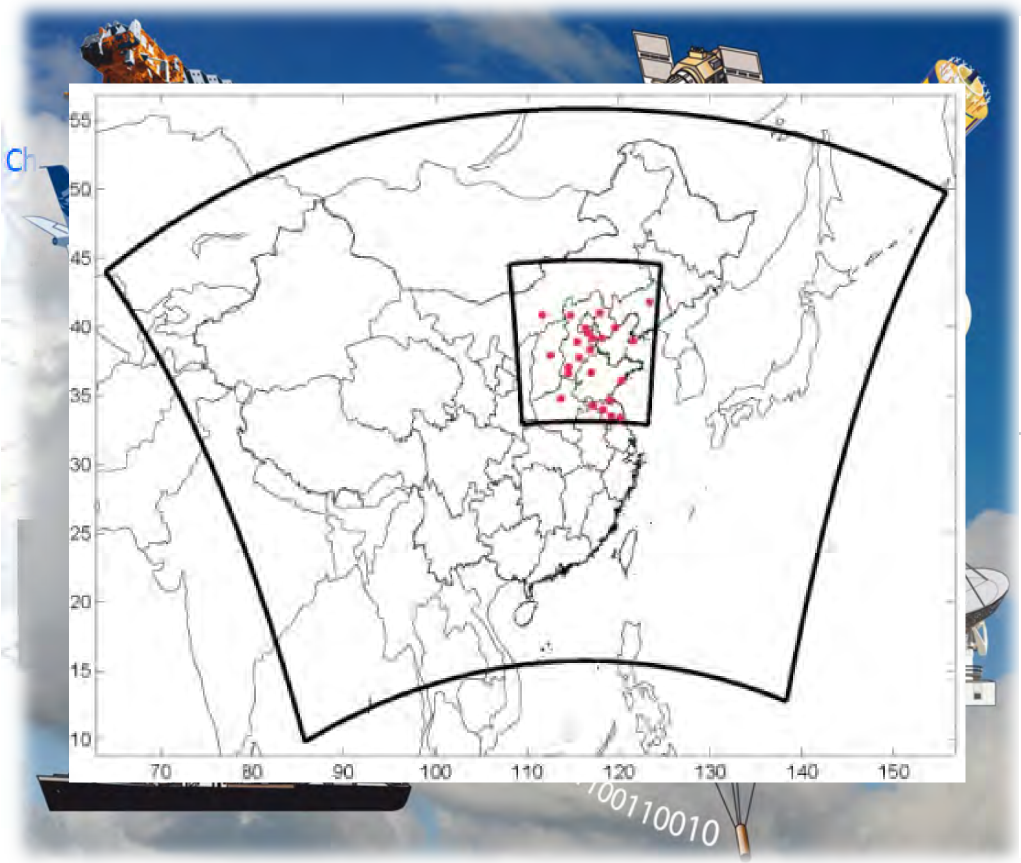
Advancing AOD assimilation

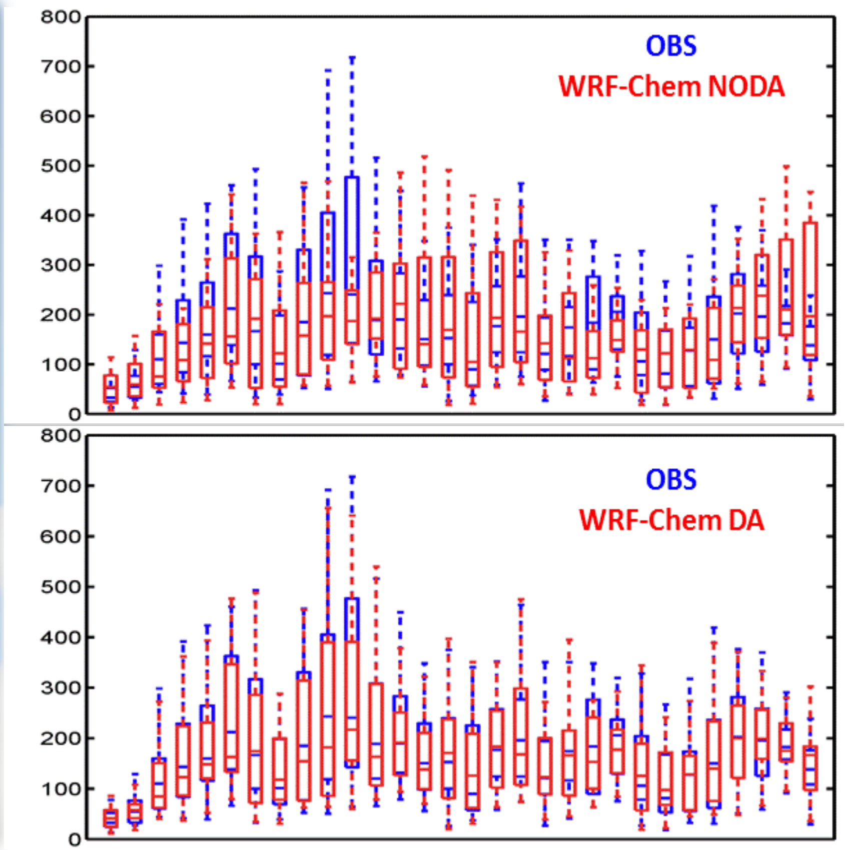
- Using Gridpoint Statistical Interpolation (GSI) (3dVAR) which can perform simultaneous DA of different datasets (e.g. ground PM2.5 and AOD, different AOD retrievals, ….) for a variety of models including WRF-Chem.
- **Our focus:** Building techniques for use with more detailed aerosol treatments, i.e., MOSAIC (sectional) AOD assimilation in GSI:
  - AOD and sensitivities computed with WRF-Chem optical averaging routine (Mie code + Internal Mixture) and its adjoint (TAPENADE)
  - Aerosol water updated as in MOSAIC (use electrolytes)
  - Correlation between bins sizes by using smoothing filters
  - Simultaneous assimilation of AOD at multiple wavelengths
  - Can also assimilate surface PM and LIDAR extinction

AND already extended to the assimilation of cloud optical depths (COD) and emission inversions, as well as all the met related fields

Ref: Saide et al., 2013, Schwartz et al., JGR 2012, Liu et al., JGR 2011, Fast et al., JGR 2006, Saide et al., PNAS 2012

Models Constrained With Observations Play Increasing Important Roles In Research and Applications



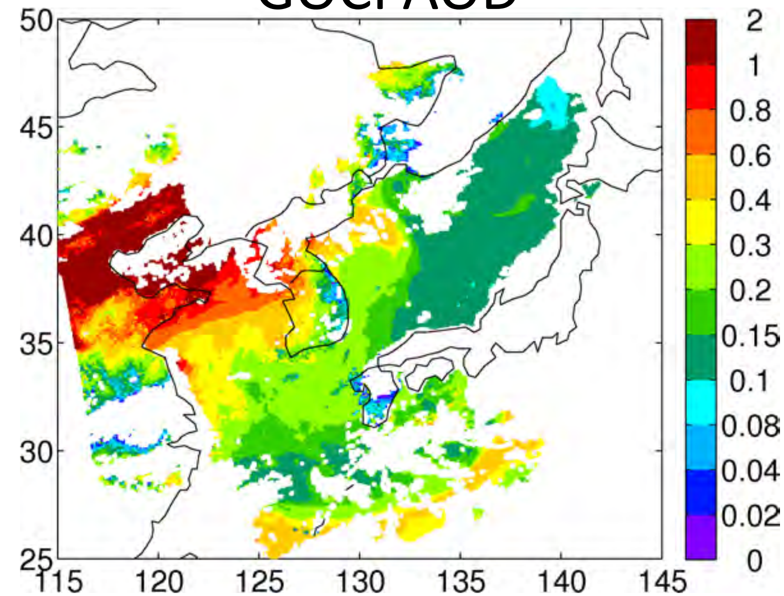


Health impacts from winter haze; Gao et al., Science Tot. Env., (2015)

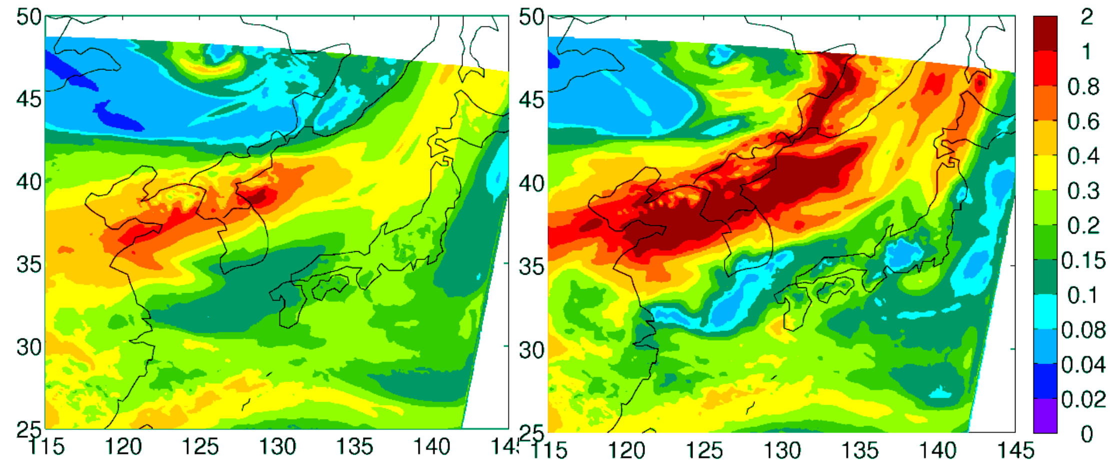
Impacts of Geostationary AOD Assimilation
(Are we "ready " to see an impact?)

- **Objectives:** Assess performance of assimilating Geostationary GOCI AOD into a system already assimilating MODIS AOD
- **System:** WRF-Chem - GSI for MOSAIC sectional aerosol model (Saide et al., ACP 2013) allows assimilation of multiple data
- **Experiments:** GSI AOD assimilation every 3 hours, MODIS only, MODIS+GOCI. (*Only over-sea AOD used)*

GOCI AOD



Aug 27-29, 2012

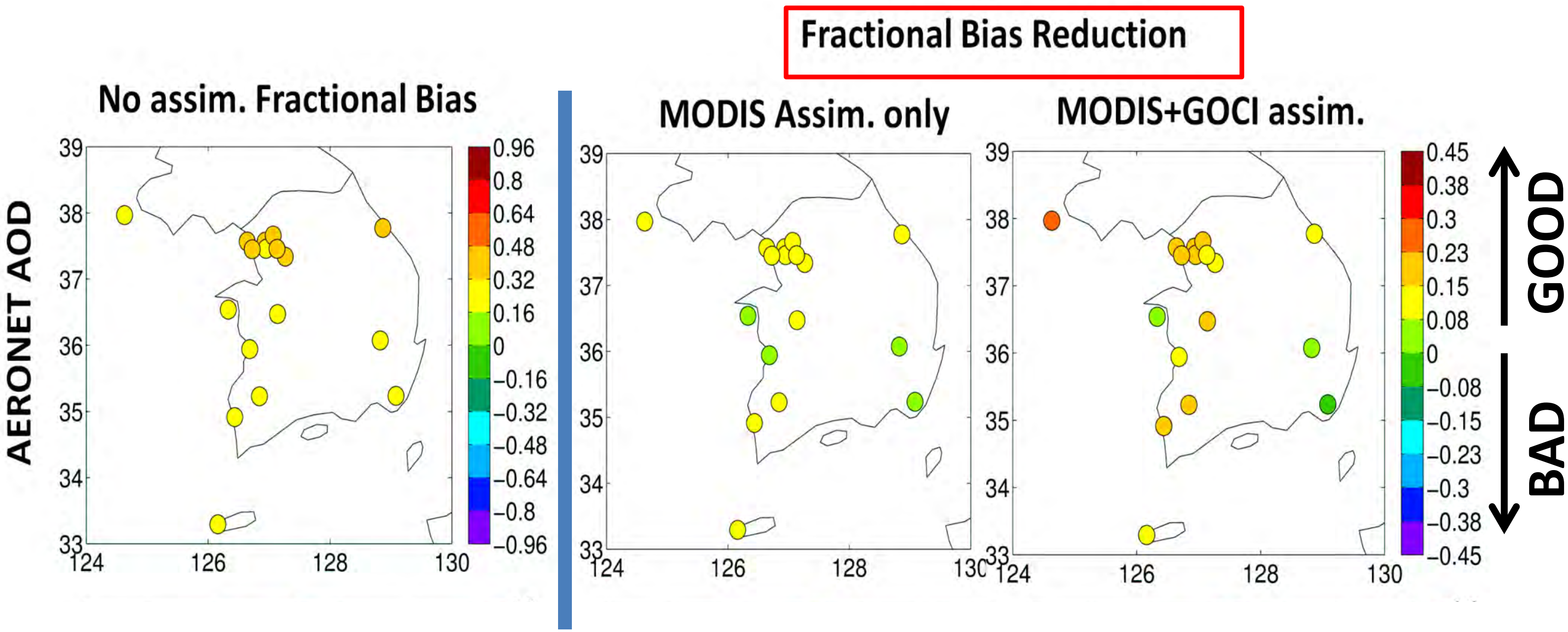


WRF-Chem NO Assim

WRF-Chem MODIS+GOCI Assim

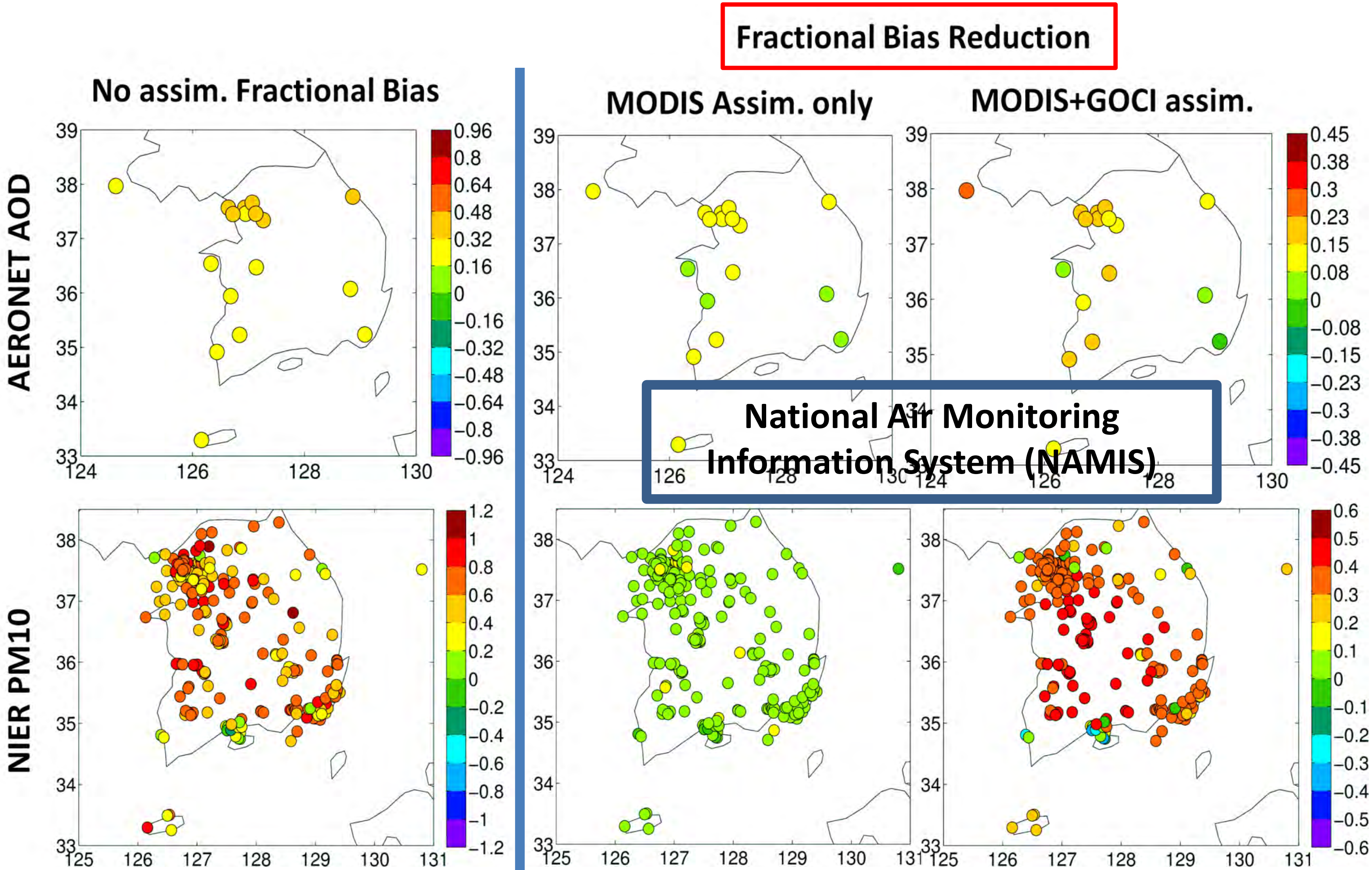
Saide et al., GRL, on-line, 2014

Impact of GOCI on AOD Prediction



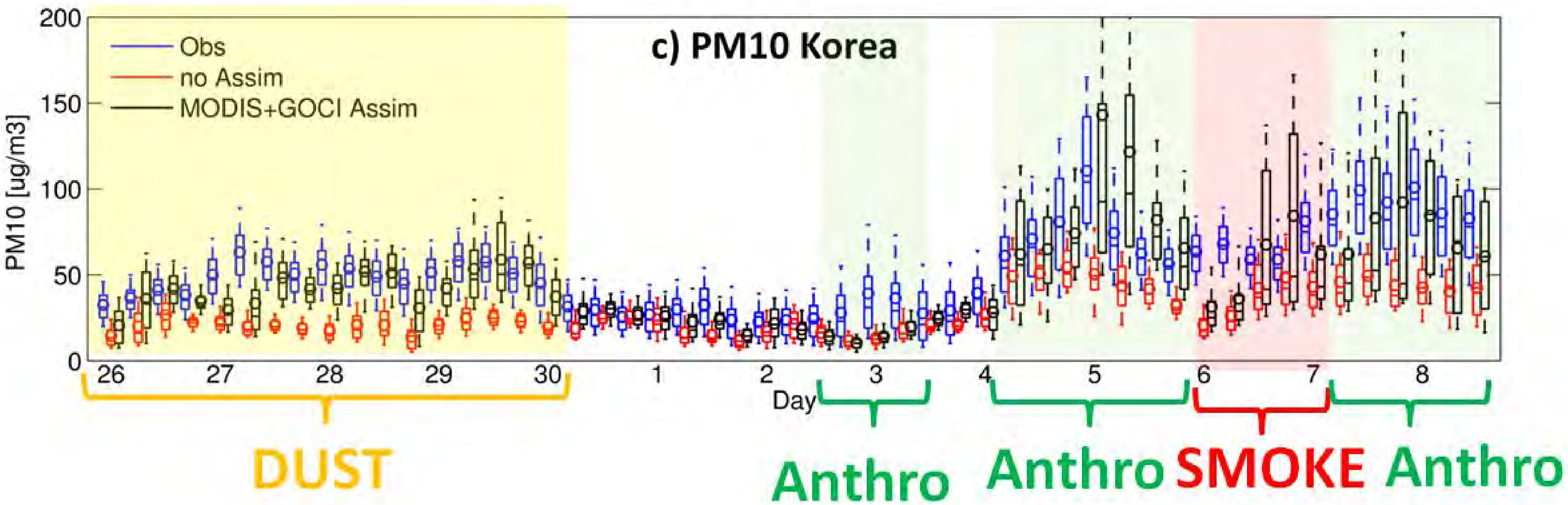
- AERONET DRAGON network

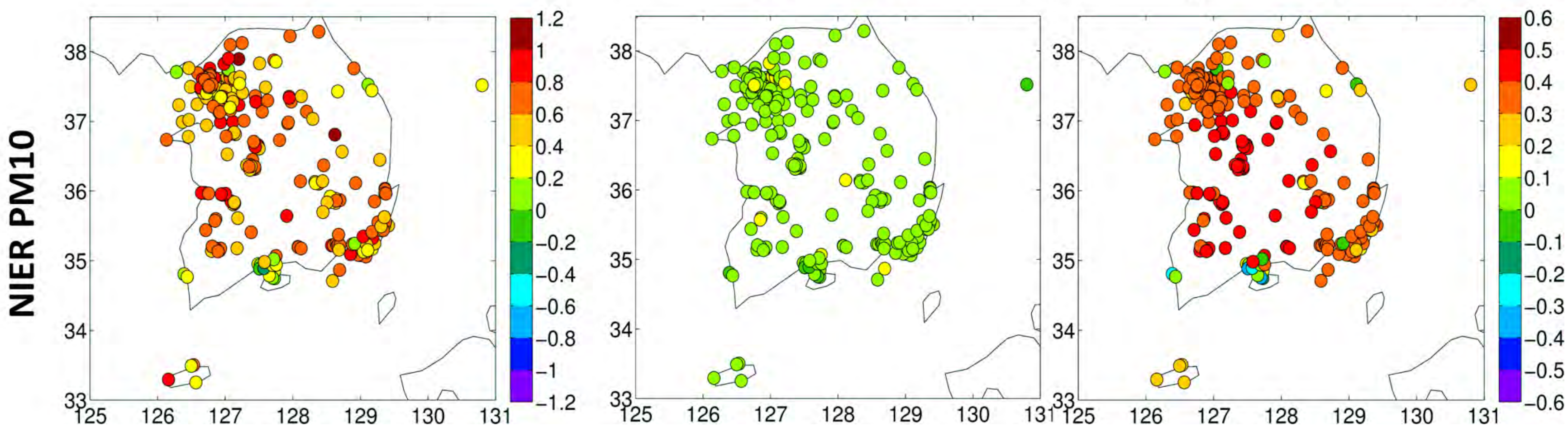
Impact of GOCI on PM10 Prediction



Data	Assimilation	FBR	FER	AER	%FBR	%FER	%AER
PM10	MODIS only	0.07	0.03	1.47	0.98	0.80	0.80
	MODIS+GOCI	0.33	0.16	2.90	0.97	0.93	0.74

Impact of GOCI on PM10 Prediction





Fractional Bias reduction

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

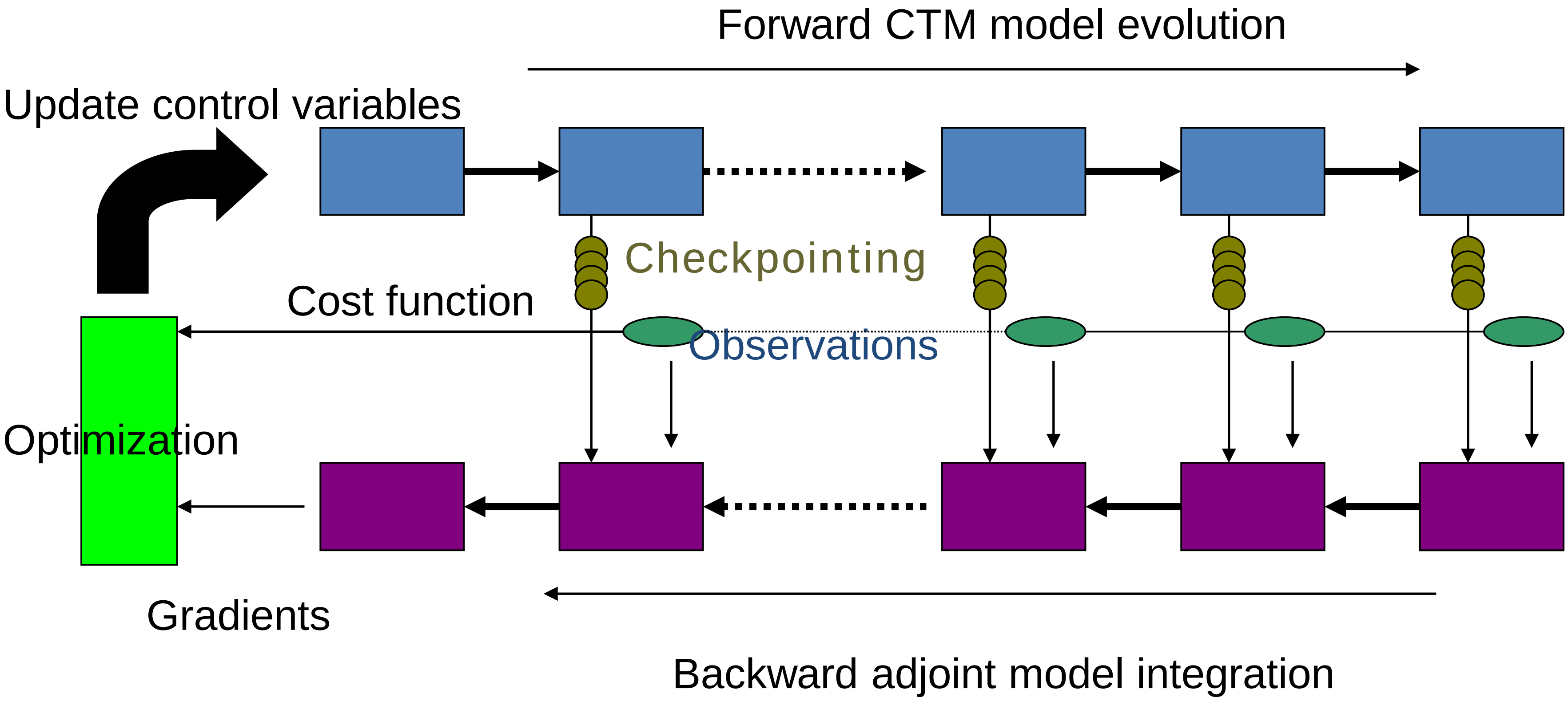
- Derive adjoint of tangent linear model

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

Where φis 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

4D-Var application with CTMs



Adjoints Provide a Powerful Tool for Sensitivity Analysis

The sensitivity can be obtained either via the direct chain rule (TLM/DDM) or via its transpose (ADJ)

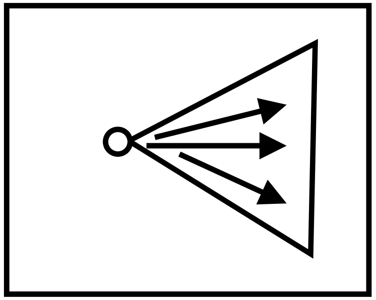
$$y^0(u), \quad u = \text{input}; \quad \psi\left(\mathbf{y}^N\right) = \text{output}$$

$$\Delta u \rightarrow \Delta y^0 \rightarrow \cdots \rightarrow \Delta y^N \rightarrow \Delta\psi\left(\mathbf{y}^N\right)$$

$$\frac{\partial \psi}{\partial u} = \frac{\partial \psi}{\partial \mathbf{y}^N}\left(\frac{\partial \mathbf{y}^N}{\partial \mathbf{y}^{N-1}} \cdots \frac{\partial \mathbf{y}^1}{\partial \mathbf{y}^0}\right)\frac{\partial \mathbf{y}^0}{\partial u}$$

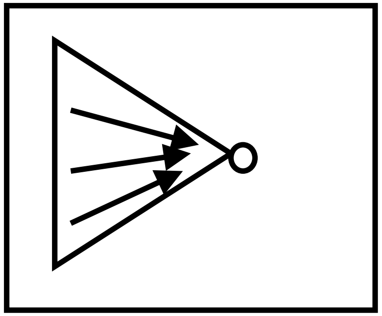
TLM (DDM) = source-oriented approach

$$\frac{\partial \psi}{\partial u} = \frac{\partial \psi}{\partial \mathbf{y}^N} \cdot \frac{\partial \mathbf{y}^N}{\partial u}; \quad \frac{\partial \mathbf{y}^k}{\partial u} = \frac{\partial \mathbf{y}^k}{\partial \mathbf{y}^{k-1}} \cdot \frac{\partial \mathbf{y}^{k-1}}{\partial u}, \quad k = 1, \cdots, N$$



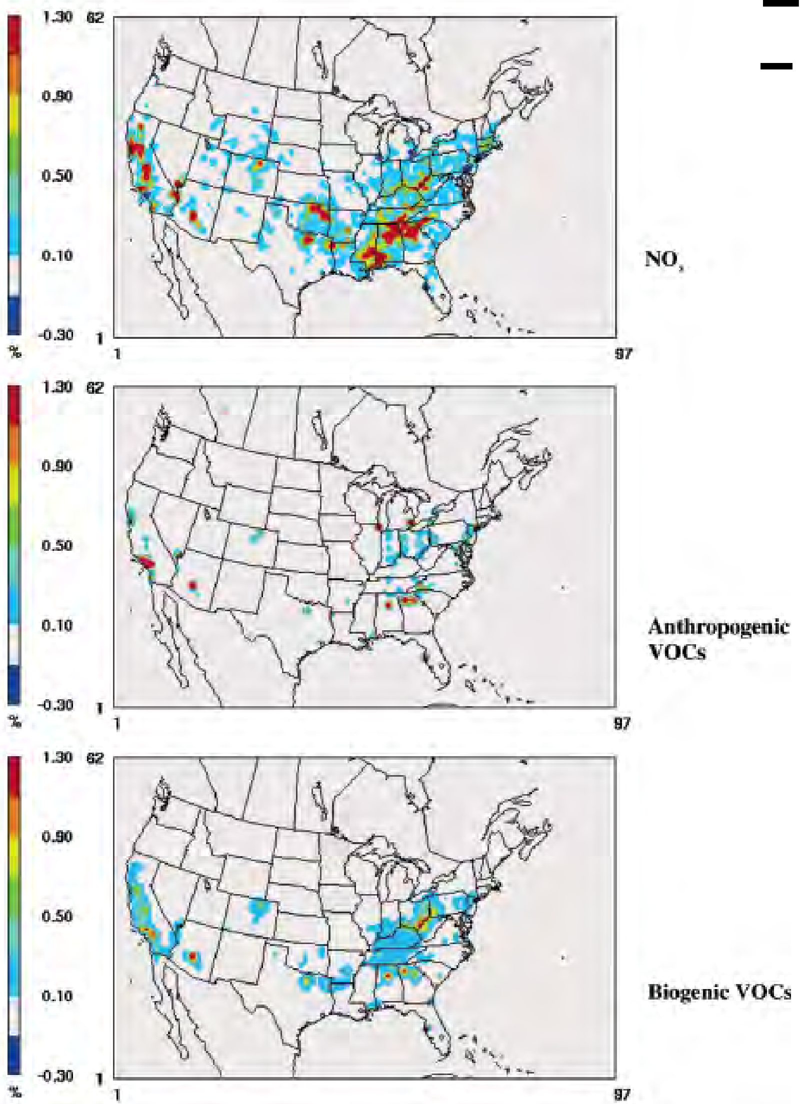
ADJ = receptor-oriented approach

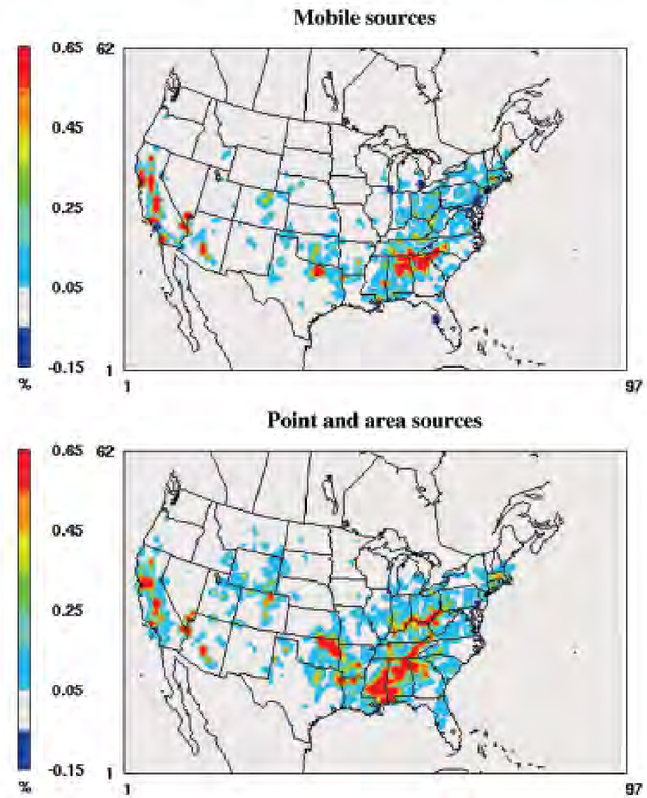
$$\frac{\partial \psi}{\partial u} = \frac{\partial \psi}{\partial \mathbf{y}^0} \cdot \frac{\partial \mathbf{y}^0}{\partial u}; \quad \left(\frac{\partial \psi}{\partial \mathbf{y}^{k-1}}\right)^T = \left(\frac{\partial \mathbf{y}^k}{\partial \mathbf{y}^{k-1}}\right)^T \cdot \left(\frac{\partial \psi}{\partial \mathbf{y}^k}\right)^T, \quad k = N, \cdots, 1$$



Sensitivity of ozone violations wrt emissions

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



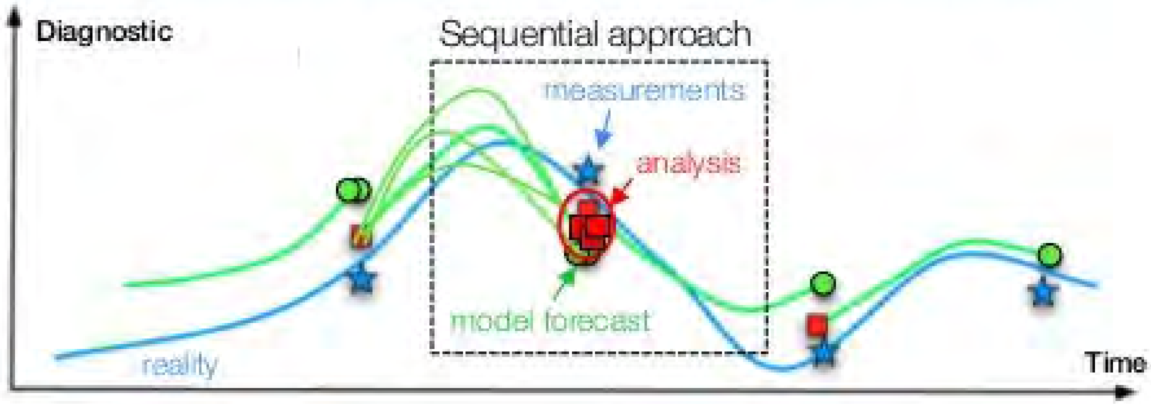


Hakami et al., ES&T 2006

INTRODUCTION

Data assimilation: why? how?

→ Key idea: "optimal combination of observations and forward model"



Ensemble Kalman filter (EnKF)

- Forecast step → uncertainty propagation
  - Explicit propagation of the error statistics
  - Nonlinear extension of the Kalman filter
- Analysis step → Kalman filter update equation

Stochastic characterization
Estimation of error covariance matrices

Kalman gain matrix

$$\blacksquare = \bullet + \boxed{\mathbf{K}}\left[\underbrace{\star - \mathcal{G}(\bullet)}_{\text{Distance to observations}}\right]$$

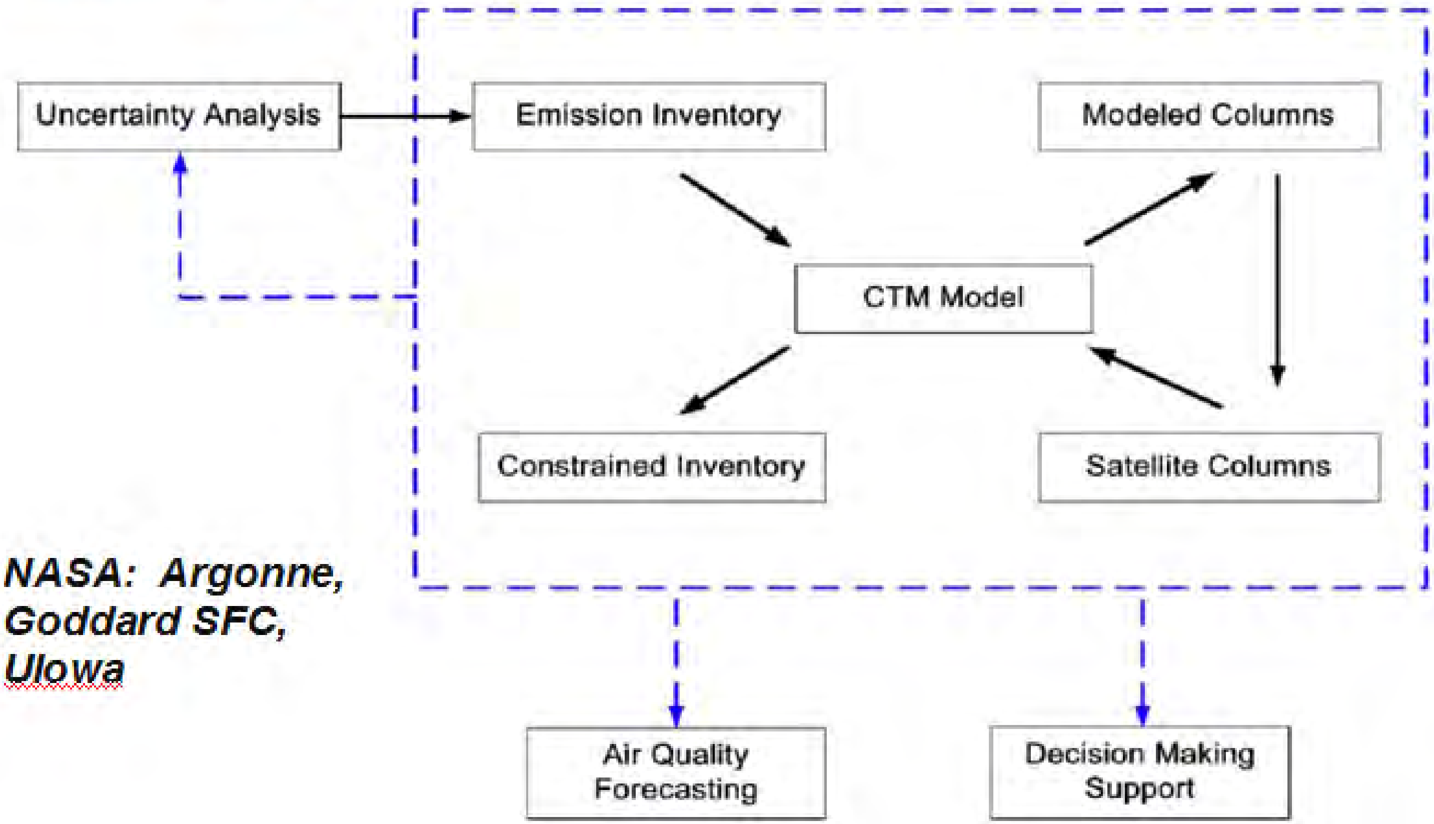
Control variables

Newer methods

- ANALOG Kalman Filter

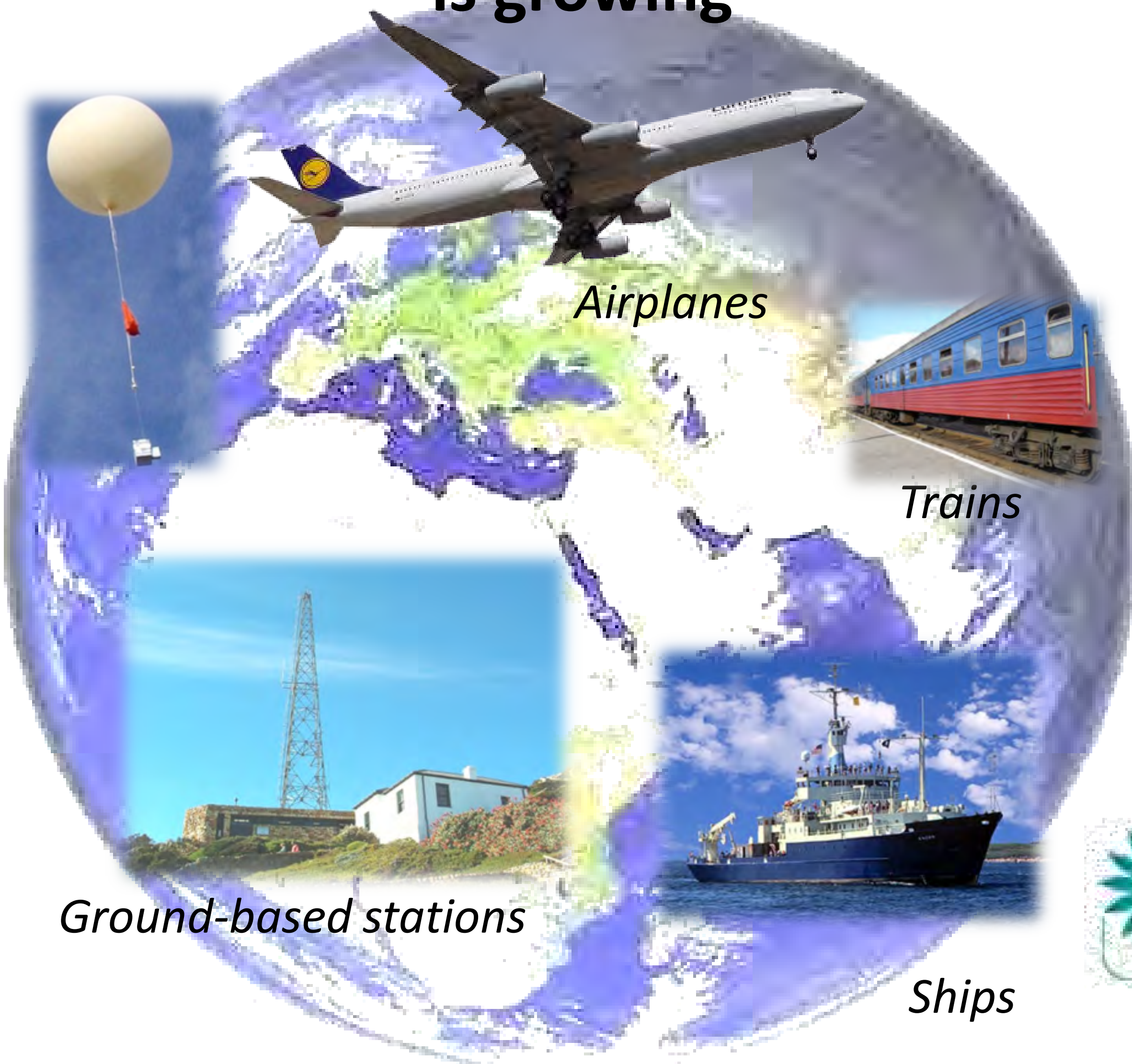
Rapid Updates of Emissions Are Needed

We are developing new approaches to integrate satellite data with chemical transport models and emission inventories for improved AQM

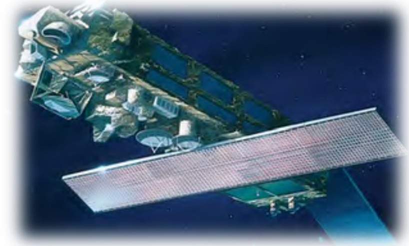


NASA: Argonne, Goddard SFC, UIowa

Good News: The global observing syst atmospheric composition is growing



Balloons



Satellites

GAW

GLOBAL ATMOSPHERE WATCH

ICOS integrated carbon observation system

IAGOS

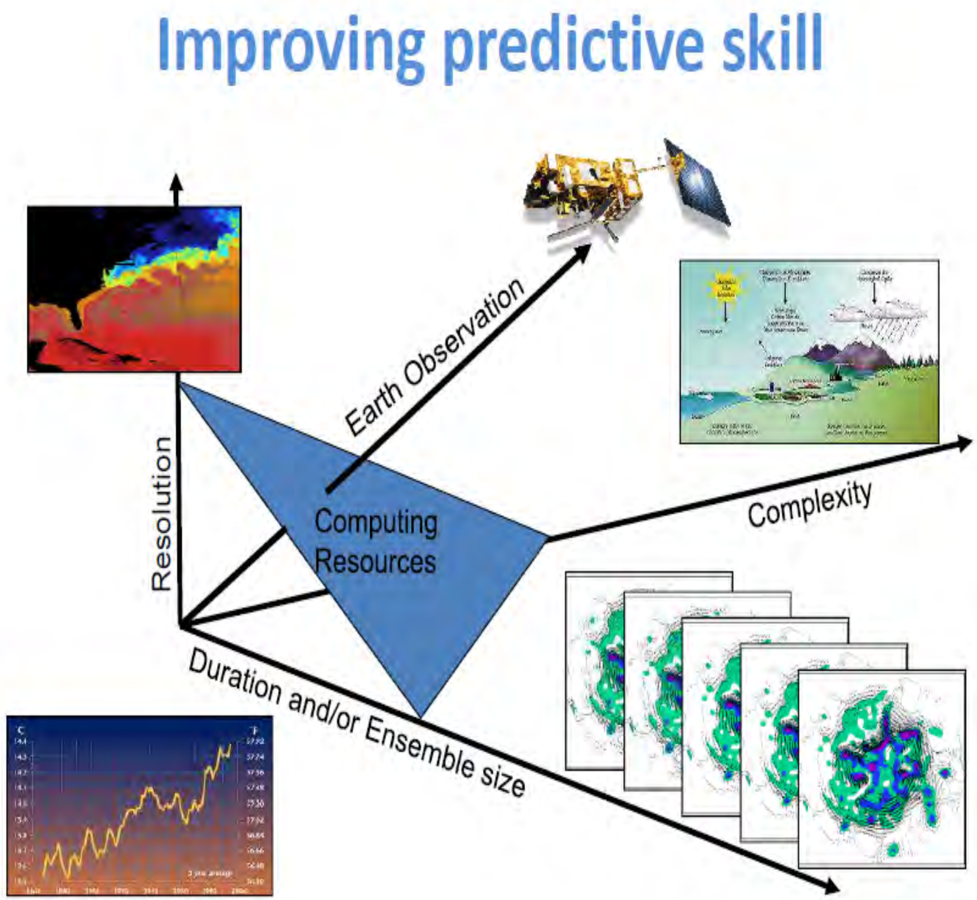
ACTRIS

EEA

AERONET Aerosol Robotic Network

...

Improving Predictive Capabilities
Exciting Times Ahead!!

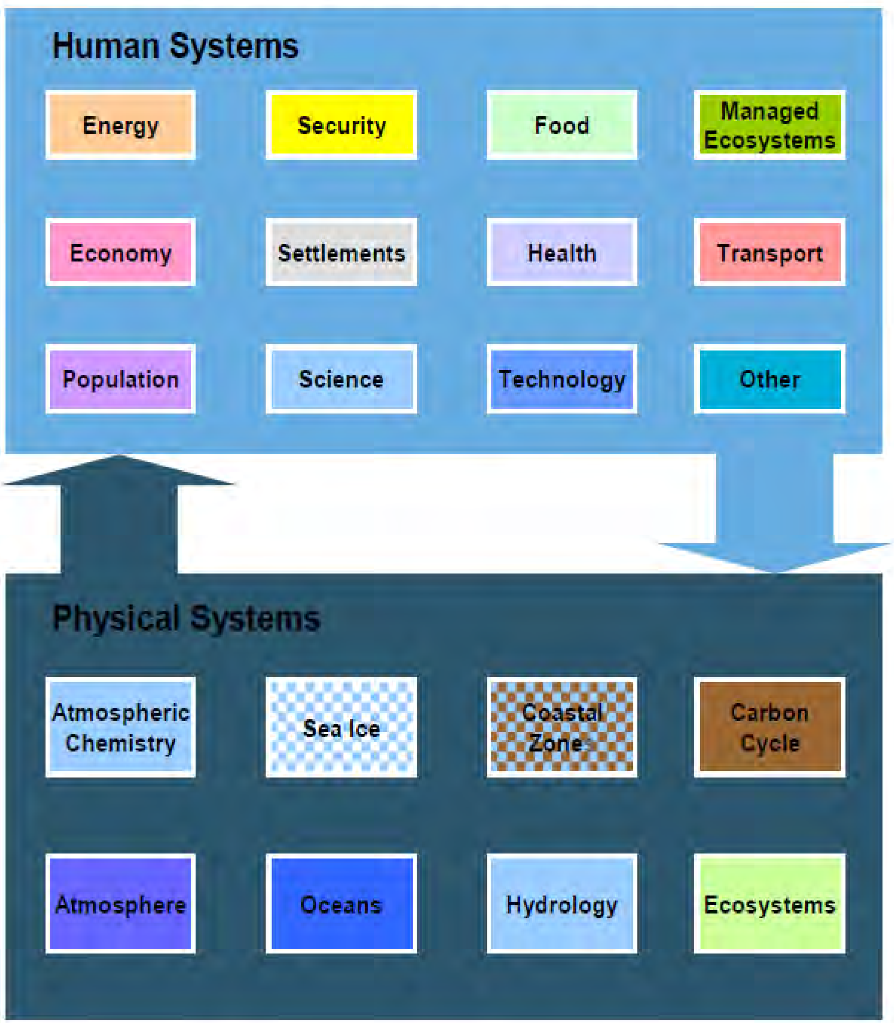


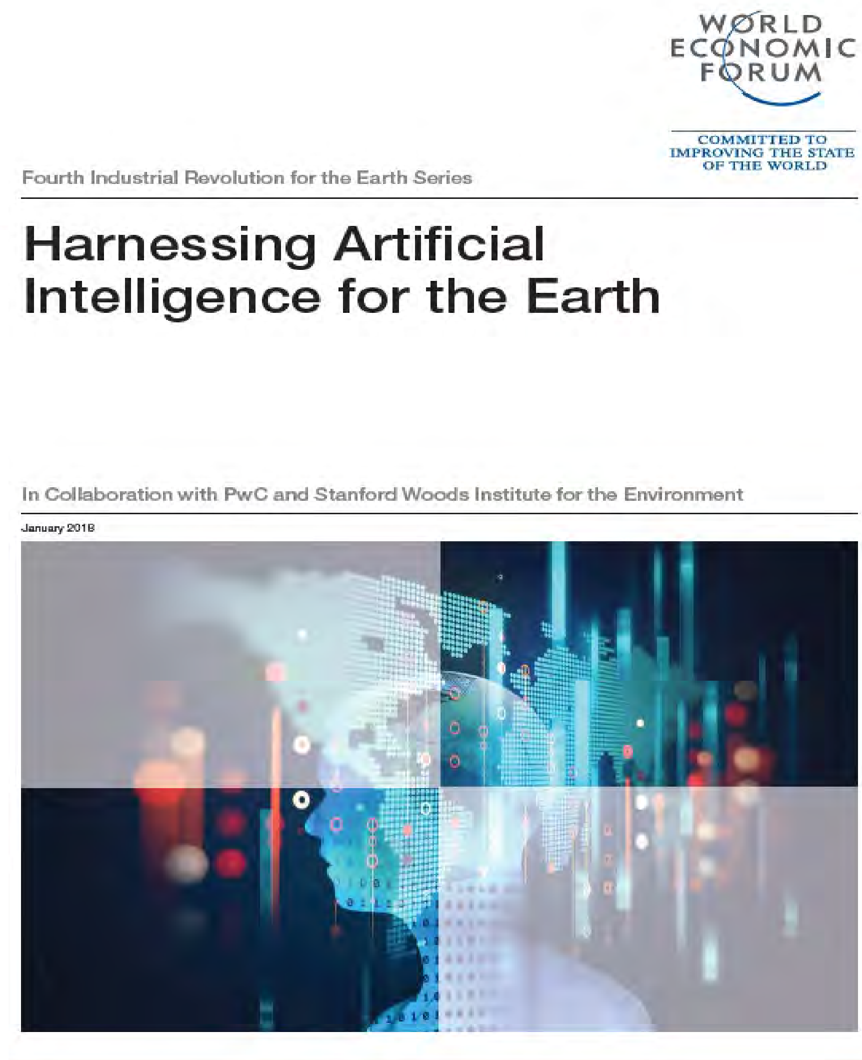
- ✓ Trend toward closer linkages of weather, atmospheric composition, and climate related services
- ✓ Information needed at higher resolution (and longer lead times) to address societal needs
- ✓ Further improvements require advances in observing systems, models and assimilation systems

Exciting Times Ahead!

Beyond physical models

Integrated Assessment Model





WORLD ECONOMIC FORUM

COMMITTED TO IMPROVING THE STATE OF THE WORLD

Fourth Industrial Revolution for the Earth Series

Harnessing Artificial Intelligence for the Earth

In Collaboration with PwC and Stanford Woods Institute for the Environment

January 2018

Motivation: A new concept for Seamless Meteorology and Chemistry Modelling

- Physical and Chemical Weather: dependence of meteorological processes (incl. precipitation, thunderstorms, radiation, clouds, fog, visibility and PBL structure) on atmospheric concentrations of chemical components (especially aerosols).
- Meteorological data assimilation (in particular assimilation of radiative properties) also depends on chemical composition.
- Air quality forecasts loose accuracy when CTMs are run offline.
- Climate modeling: large uncertainty of SLCFs, water vapor feedbacks, etc.

=> Need for a new generation of seamless integrated meteorology and chemistry modelling systems for predicting atmospheric composition, meteorology and climate evolution !