Chemical Data Assimilation – The Need for Closer Integration of Measurements and Models Greg Carmichael, Dept. Chemical & Biochemical Engineering & Center for Global & Regional Environmental Research



This picture, adapted from Katsushika Hokusai's masterpiece "The Great Wave off Kanagawa," artistically displays the spirit of supercomputing. Complex phenomena, such as waves on the surface of a fluid, are modeled by covering space with a grid and then solving the laws of physics at discrete points on that grid. The finer the grid, the closer the numerical simulation is to the actual solutions of the mathematical laws of nature that govern the physical world. Chemical Data Assimilation – The Need for Closer Integration of Measurements and Models Greg Carmichael, Dept. Chemical & Biochemical Engineering & Center for Global & Regional Environmental Research

University of Iowa



TODAY:

- Continue the introduction into chemical data assimilation (focus on variational methods)
- Illustrate power and limitations through examples

This picture, adapted from Katsushika Hokusai's masterpiece "The Great Wave off Kanagawa," artistically displays the spirit of supercomputing. Complex phenomena, such as waves on the surface of a fluid, are modeled by covering space with a grid and then solving the laws of physics at discrete points on that grid. The finer the grid, the closer the numerical simulation is to the actual solutions of the mathematical laws of nature that govern the physical world.

Forecast Skill (Persistence vs Model)

Persistence





Ozone - summer

Challenge: Improving Predictions through 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!!

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)



Why do data assimilation?

- 1. To obtain an initial state for launching forecasts
- 2. To make consistent estimates of the atmospheric state for diagnostic studies.
 - reanalyses
- 3. To challenge models with data and vice versa

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

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 wellquantified, 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)

$$\tau_m' = \tau_m + K \big(\tau_o - H \tau_m \big)$$

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

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

•K – Kalman gain matrix

•H – Interpolator from model to observation space

•O/B Observed/Background error covariances

•f and $\epsilon\,$ - fractional error and RMS uncertainty

 $\bullet L_{xy}$ horizontal correlation length scale for errors in model fields

 $\cdot d_x$ 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) (AOD is not unique; aerosol, size, composition, RH etc.)
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); PM₁₀ and SO₄ from EANET



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.

Elements of NWP at convective scale : DA algorithms

At convective scale, most of operational NWP centers (MF, UKMO, NCEP...) use 3DVar schemes with short assimilation/forecast cycles to limit the gap in time between observations and the forecast to be corrected

$$J(\mathbf{x}) = \frac{1}{2} (\mathbf{x} - \mathbf{x}^{\mathbf{b}})^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}^{\mathbf{b}}) + \frac{1}{2} (\mathbf{y}^{\mathbf{o}} - H(\mathbf{x}))^T \mathbf{R}^{-1} (\mathbf{y}^{\mathbf{o}} - H(\mathbf{x}))$$



+ Cheap, fast, no TL/AD of M

- no integration in time: only observations valid around the analysis time are considered

- as in 4DVar, **B** is not flow dependent

Four-dimensional Variational (4D-Var) Chemical Data Assimilation (Using STEM)

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

Control variables can be initial conditions, parameters, emissions, boundary conditions

• Use operator splitting to solve CTM

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

Data assimilation



Model state

Basic idea of 4D-Var

•Define a cost functional

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

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. $\lambda_i = \frac{\partial J}{\partial r}$

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

4D-Var application with CTMs



Backward adjoint model integration

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)

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



NMC method results



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) $y^{0}(u), \quad u = \text{input}; \quad \psi(\mathbf{y}^{N}) = \text{output}$ $\Delta u \to \Delta y^{0} \to \dots \to \Delta y^{N} \to \Delta \psi(\mathbf{y}^{N})$ $\frac{\partial \psi}{\partial u} = \frac{\partial \psi}{\partial \mathbf{y}^{N}} \left(\frac{\partial \mathbf{y}^{N}}{\partial \mathbf{y}^{N-1}} \cdots \frac{\partial \mathbf{y}^{1}}{\partial \mathbf{y}^{0}}\right) \frac{\partial \mathbf{y}^{0}}{\partial u}$

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

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



The Adjoints Are Themselves Very Valuable

CO as a tracer of fossil fuel CO_2 ...

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



1e-06 2.15e-06 4.64e-06 1e-05 2.15e-05 4.64e-05 1e-04 0.000215 0.000464 0.001

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

Analysis of the impact of assimilation of observations on the calculation of States contribution to ozone violations ...





Building Assimilation Systems for WRF-Chem



We are developing the adjoint of the full WRF-Chem Model

Ref: Saide et al., 2013, Saide et al., PNAS 2012, GRL 2014, 2015a,b; Guerette and Henze GMD, 2015 for use in 3d (GSI) and 4dVar applications

Building Assimilation Systems for WRF-Chem

WRFPLUS-Chem Current Capabilities

AD and TL developed for:

Guerette and Henze GMD, 2015

- Emissions (Anthropogenic and Biomass Burning for GOCART)
- ACM2 PBL (Pleim et al. 2007) w/ integrated tracer transport (2013)
- Pleim-Xiu and SFCLAY surface layer schemes
- Pleim-Xiu and SLAB LSMs
- Wesely Dry Deposition
- Chem advection (similar to tracer advection by Xin Zhang in 2012)
- GOCART Aerosols
 - BC aging
 - PM summation
 - Sulfate chemistry (pending testing)
- 2nd-order checkpointing for long duration (>6 hr) AD/TL simulations
- New adjoint forcing scheme for in-situ chemical observations
- Standalone AD and TL:
 - Grell-Freitas Convection (2013) [chemical tracers in development]

Ref: Saide et al., 2013, Saide et al., PNAS 2012, GRL 2014, 2015a,b; Guerette and Henze GMD, 2015 for use in 3d (GSI) and 4dVar applications

ARW-WRF/Chem and the Gridpoint Statistical Interpolation (GSI) Analysis System (3dVar)



R

Bias

Grell et al., 2009

Models Constrained with Observations Play increasing Important Roles in Research and Applications



 ✓ Need for *More* aerosol and atmospheric composition data for use in assimilation Good News: The global observing systems for atmospheric composition are growing



Balloons

Models Constrained with Observations Play increasing Important Roles in Research and Applications



- Need for *More* aerosol and atmospheric composition data for use in assimilation
- New observations streams are in the pipe-line ...

 Are our modeling & assimilation systems ready to use these data?

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

- Near future more assets, e.g., GEMS,
- Objective: 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+GÓCI. (Only oversea AOD used)

Saide et al., GRL, 2014



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

Near future more assets , e.g., GEMS!!

- **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+GÓCI (Only oversea AOD used)



50

45

40

35



35

Impact of GOCI on PM10



Fractional Bias reduction

Assimilation Technique along with GOCI Assimilation Changes 4-d Fields --





KORUS-AQ forecasting system



- WRF-Chem with MOSAIC aerosols and a Reduced Hydrocarbon chemistry (Pfister et al. JGR 2014), including simplified SOA formation (Hodzic and Jimenez, GMD 2011)
- GFS and MACC meteorological and chemical boundary conditions
- KORUS-AQ anthro (Jung-Hun Woo) and QFED fire emissions
- AOD data assimilation using GSI (Saide et al., ACP 2013). MODIS and GOCI data were assimilated simultaneously every three hours. First NRT system assimilating GEO AOD
- Four days of forecasts were available for the outer domain, 2 days for the inner domain

Example of AOD assimilation impact during KORUS-AQ

Day-2

Day-3

Day-1





WRF-Chem Forecast Evaluation vs MODIS and **AERONET AOD**



40

Impact of Assimilation on AOD Forecasts (2nd







Technique already Extended to include Assimilation of Surface PM2.5 Observations

Haze JAN 2013



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

Application #3a. Adding More Information (AOD is not unique!)- multiple wavelength AOD

Fractional error reductions (FER)

- Compute fractional error for guess and assimilation
- Take the difference
- Positive values mean improvements





Shift in size distribution

Ref: Saide et al ACP 2013

Application #3b. Adding More Information – problem with clouds Challenge in the Use of AOD in AQ Forecasting/Reanalysis



Current Base Layer: MODIS/Terra Corrected Reflectance (True Color



Challenge

In cloudy regions there is little AOD information – also it provides limited info on size/composition

Idea:

Can we use the cloud retrievals to improve sub-cloud aerosol distributions?



20° S VOCALS-REx stats

Idea use cloud drop information

- SEP marine Sc N_d satellite retrievals show evidence of aerosol load and agree with observations
- Aerosol indirect effects simulated with some skill in WRF-Chem
- Hypothesis: variational assimilation with N_d retrievals can improve below-cloud aerosols in models (see aerosols through clouds)



Aerosol data assimilation using cloud satellite retrievals

50



.

Ref: Saide et al., PNAS 2012

Assimilation results: + & - biases 51 reduced

- Assimilate MODIS Terra
 N_d
- Aerosol mass and number are changed





Daytime N_d after assimilation vs GOES and in-situ aerosol

- Large improvements during the first 2 days for all domain
- GOES Assimilation improves agreement with VOCALS-Rex C130 aerosol number and mass observations





Assimilation of ICARTT Ozone Observations -- Assessing Information Content (4dVar)



Assimilation Produces An Optimal State Space

the importance of measurements above the surface!



background (B) error estimates

Chai et al., JGR 2007

Information below 4 km most important

Region-mean profile

100

50

Ozone (ppbv)

в

D

Framework To improve modeled ozone contributions from non-local and local sources

¹Using multi-model experiment and multi-scale chemical data assimilation of Aura observations, we aim at improving the estimated impacts on ozone in California from trans-boundary pollutants and US ozone precursors (NOx & isoprene) emissions



Huang et al., JGR, 2015

Framework To improve modeled ozone contributions from non-local and local sources

[•]Using multi-model experiment and multi-scale chemical data assimilation of Aura observations, we aim at improving the estimated impacts on ozone in California from trans-boundary pollutants and US ozone precursors (NOx & isoprene) emissions



Impact of the multi-scale assimilation on STEM ozone:

We repartitioned ozone source contributions from local and non-local sources

Period-mean near-surface daytime (<2 km a.g.l., 8am-7pm) ozone in STEM, after-before assimilation

> 0.5 -1.5

2.5

3.5

4.5



Contributions from "transported background" increased by 2-6 ppbv.

net changes



w/ updated emissions



Contributions from US NOx & isoprene emissions dropped by ~0-6 ppbv.

- Background ozone enhanced by \sim 4 ppbv (\sim 8%)
- Monitoring sites unable to capture some strong changes made by the assimilation
- High terrain regions: more sensitive to extraregional sources
- Central Valley: more sensitive to local NOx and isoprene emissions

Constraining biomass burning emissions for improved prediction skill and assessing smoke-weather interactions

- WRF with aerosol-aware microphysics (AAM) (Thompson and Eidhammer, 2014) and WRF-Chem emissions.
- Inversion based on Saide et al. (GRL 2015b) using WRF tracers (no adjoint, no ensembles)
- Plans for using it operationally for the NASA ORACLES and NOAA FIREX field experiments
- Run 2 forecasts needed by the inversion method (with and without feedbacks) so that aerosol feedbacks impacts are also forecasted





Opportunity: Assimilation of Key Meteorological Parameters Improve AQ Prediction Skill





The Growing Interest in Improving Air Quality Predictions/Services and the Role of Atmospheric Composition in Weather and Climate Applications Offer Great Opportunities for Our Community

