International Symposium on Data Assimilation

July 18-22, 2016, University of Reading

Coupled data assimilation with Maximum Likelihood Ensemble Filter (MLEF)

Milija Zupanski
Cooperative Institute for Research in the Atmosphere
Colorado State University
Fort Collins, Colorado

Thanks to:

Kazuyoshi Suzuki (JAMSTEC, Japan)

Seon-Ki Park, Ebony Lee (Ewha Univ., South Korea)

NASA MAP #NNX13AO10G

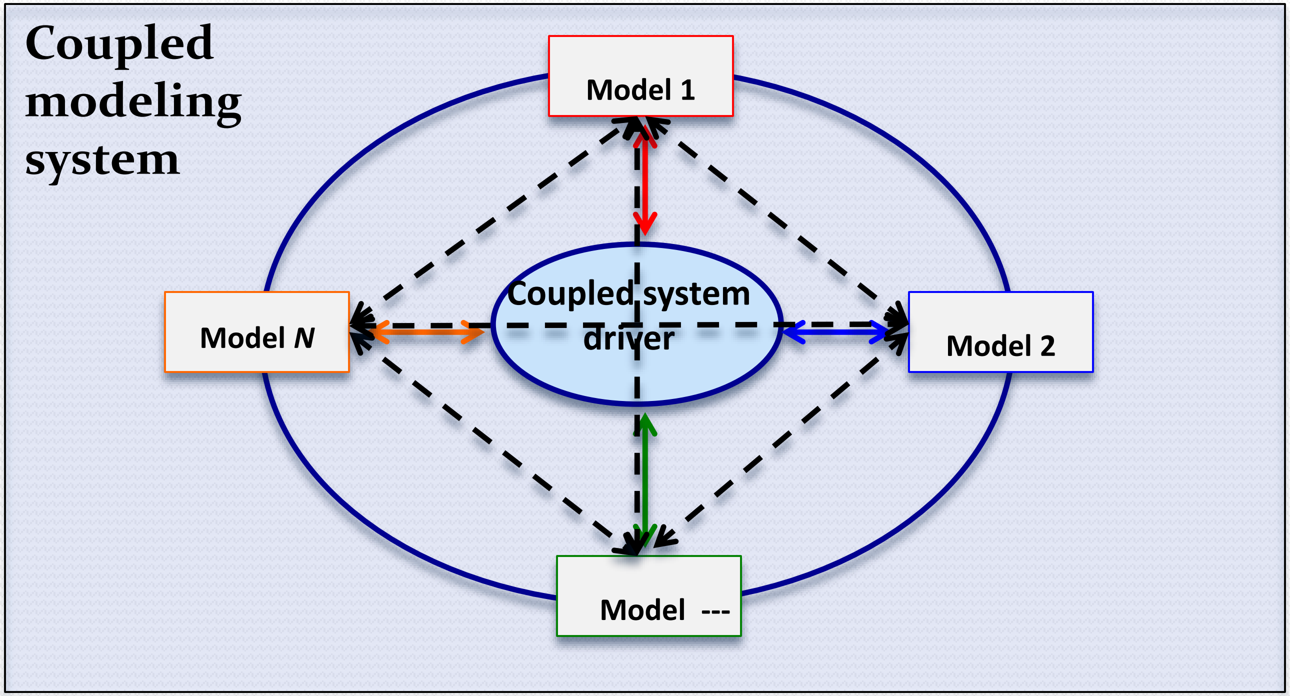
ONR #N000149169192040

Overview

- Coupled systems and coupled data assimilation (DA)
- Forecast error covariance in coupled DA
- Information flow in coupled DA
- Maximum Likelihood Ensemble Filter (MLEF)
- Coupled data assimilation results with MLEF
- Some challenges of coupled DA
- Future

Coupled models

- **Complex interactions between coupled system components**
- Each component model interacts with the system driver
- System components may interact with each other
- Single or multiple executables



Coupled data assimilation

- What is coupled data assimilation?

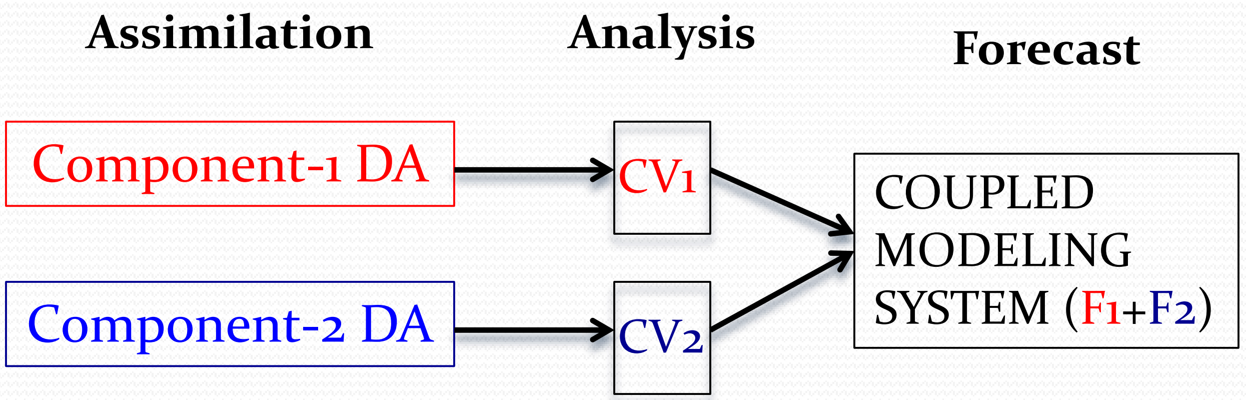
Coupled DA is data assimilation with coupled modeling systems

- How does coupled data assimilation differ from a standalone data assimilation for each component?

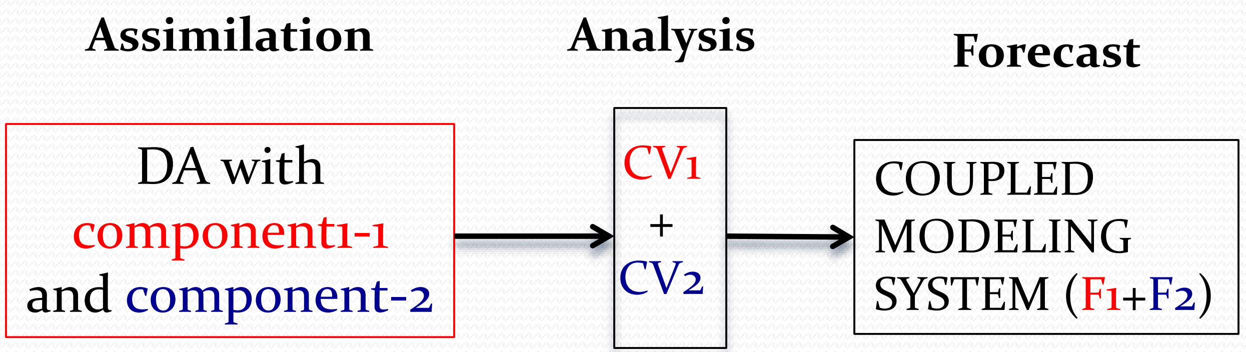
  - Increased number of control variables (increased dimension)
  - Increased complexity of control variables
  - Generally unknown cross-component correlations
  - Possibly different spatiotemporal scales

Data assimilation options

(1) Standalone data assimilation for each component



(2) Coupled data assimilation with all components



Mathematical details of coupled DA

- Coupled DA can be interpreted as an augmented DA system

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_S \end{pmatrix} \qquad P_f = \begin{pmatrix} P_{11} & & P_{1S} \\ & \ddots & \vdots \\ P_{1S}^T & \cdots & P_{SS} \end{pmatrix}$$

- Forecast

$$x^{n+1} = m(x^n)$$

- Analysis

$$J(x) = \frac{1}{2}(x - x^f)^T P_f^{-1}(x - x^f) + \frac{1}{2}[y_1 - h_1(x_1)]^T R_1^{-1}[y_1 - h_1(x_1)] + \cdots + \frac{1}{2}[y_S - h_S(x_S)]^T R_S^{-1}[y_S - h_S(x_S)]$$

No major difference between coupled and standalone equations.

Review: Analysis correction

Singular Value Decomposition: $$P_f^{1/2} = USV^T = \sum_i \sigma_i u_i v_i^T$$

Generic change of variable to avoid matrix inversion: $$x = x^f + P_f^{1/2} w \qquad (x - x^f)^T P_f^{-1} (x - x^f) \Rightarrow w^T w$$

Analysis update: $$x^a = x^f + P_f^{1/2} z_{obs}$$

KF example: $$z_{obs} = P_f^{T/2} H^T (H P_f H^T + R)^{-1} (y - h(x^f))$$

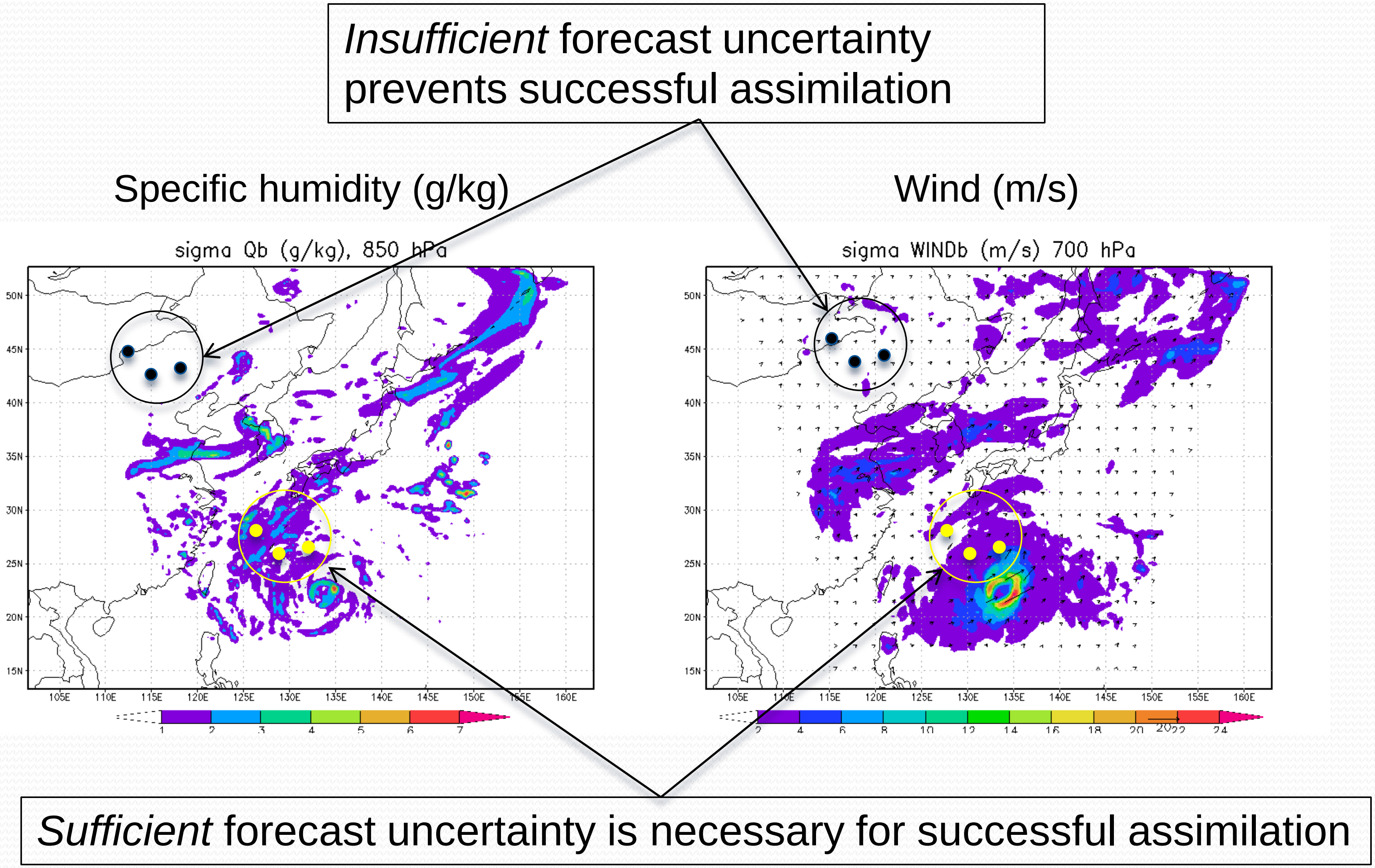
The DA analysis update is

$$x^a - x^f = \left(\sum_i \sigma_i u_i v_i^T \right) z_{obs} = \sum_i \mu_i u_i \qquad \mu_i = \sigma_i v_i^T z_{obs}$$

- Analysis increments are defined in the subspace spanned by the left SVs of the (SQRT) forecast error covariance

- Projection of (transformed) observations onto the right SVs of the (SQRT) forecast error covariance is critical for allowing the impact of observations

Forecast uncertainty – 32 ensembles (Typhoon Nabi, valid 03 Sep 2005 0300 UTC)



Comments on forecast error covariance in coupled DA

- Analysis correction subspace

As for the standalone multivariate system, forecast error covariance singular vectors define the space of analysis corrections, as well as whether or not the observations can have an impact on the analysis.

- Flow-dependent error covariance

May be even more relevant than in standalone DA, as cross-component correlations may be completely unknown.

- Spatiotemporal scales

By using model forecasts to define flow dependent error covariance, dependencies that develop during the couple system forecast will be present in the coupled forecast error covariance. Otherwise, the covariance modeling has to explicitly include interactions between scales.

- Error covariance localization may be challenging

Coupled DA and information

- (Coupled) data assimilation can be interpreted as *processing and exchange of information between system components with the aim of maximizing the efficiency of information flow*

- Shannon entropy

$$H\{X\} = -\int p(x)\log(p(x))\,dx$$

- Impact of data assimilation: change of entropy due to observations

$$\Delta H = H\{X\} - H\{X|Y\}$$

- Joint entropy can be used to represent the entropy of a coupled system

$$H(X_1, X_2) = H(X_1) + H(X_2) - MI(X_1, X_2) \qquad MI(X_1, X_2) \geq 0$$

$$H(X_1, X_2) \leq H(X_1) + H(X_2)$$

Total uncertainty of the coupled system is reduced compared to the uncertainty of individual components

Maximum Likelihood Ensemble Filter (MLEF)

***Forecast*: Uncertainty evolution using a nonlinear prediction model**

Transport uncertainty in time by a *nonlinear* model m (one span vector at a time)

$$x^f = m(x^a) \qquad x_i^f = m(x^a + p_i^a)$$

$$p_i^f = x_i^f - x^f = m(x^a + p_i^a) - m(x^a)$$

Each uncertainty column-vector is a member of an "ensemble" (i.e. span)

***Analysis*: Iterative minimization of an arbitrary nonlinear cost function**

- **Maximum a-posteriori (MAP) method**
- **Use best applicable minimization method**
- **Optimal Hessian preconditioning**

$$J(x) = \frac{1}{2}\left(x - x^f\right)^T P_f^{-1}\left(x - x^f\right) + \frac{1}{2}\left(y - h(x)\right)^T R^{-1}\left(y - h(x)\right)$$

Role of Hessian (second derivative)

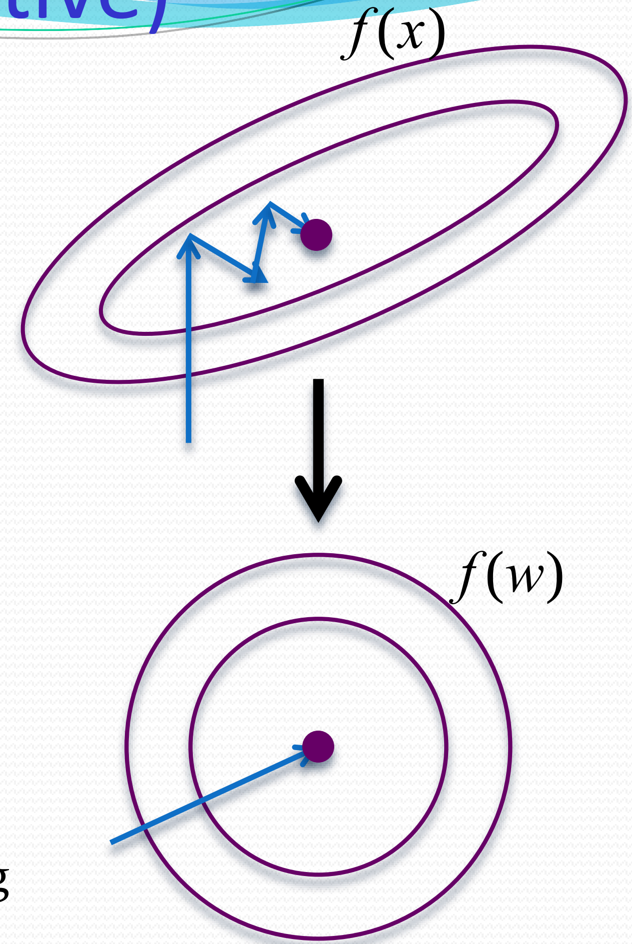
- Optimal preconditioning is the transpose of the inverse square root of the Hessian

$$G = P_f^{-1} + H^T R^{-1} H$$

$$G = EE^T$$

$$x = x^f + E^{-T} w$$

$$G_w = E^{-1}(EE^T)E^{-T} = I$$



- For Gaussian cost function, MLEF applies a two-step change of variable to achieve Hessian preconditioning

$$x = x^f + P_f^{1/2} w$$

$$w = (I + Z^T Z)^{-1/2} z \qquad Z_i = R^{-1/2}[h(x + p_i^f) - h(x)]$$

Addressing nonlinearity and non-differentiability of cost function in MLEF

$$J(w) = \frac{1}{2} w^T w + \frac{1}{2}\left[y - h(x^f + P_f^{1/2} w)\right]^T R^{-1}\left[y - h(x^f + P_f^{1/2} w)\right]$$

Standard Taylor expansion of cost function: $\mathrm{d}w = \boldsymbol{1}_i$

$$J(w + \mathrm{d}w) = J(w) + \left(\frac{\partial J}{\partial w}\right)^T \mathrm{d}w + \frac{1}{2!}(\mathrm{d}w)^T \frac{\partial^2 J}{\partial w^2}\mathrm{d}w + O(\|\mathrm{d}w\|^3)$$

Calculate the difference without Taylor expansion:

$$J(w + \mathrm{d}w) = J(w) + (\mathrm{d}w)^T w - \left[h(w + \mathrm{d}w) - h(w)\right]^T R^{-1}\left[y - h(x)\right] +$$
$$+ \frac{1}{2}(\mathrm{d}w)^T \mathrm{d}w + \frac{1}{2}\left[h(w + \mathrm{d}w) - h(w)\right]^T R^{-1}\left[h(w + \mathrm{d}w) - h(w)\right]$$

Use finite-difference representation of derivatives in iterative minimization:

$$g(w) = \mathrm{d}w^T w - \left[h(w + \mathrm{d}w) - h(w)\right]^T R^{-1}\left[y - h(x)\right]$$

$$G(w) = \frac{1}{2}\mathrm{d}w^T \mathrm{d}w + \frac{1}{2}\left[h(w + \mathrm{d}w) - h(w)\right]^T R^{-1}\left[h(w + \mathrm{d}w) - h(w)\right]$$

Nonlinear and non-differentiable problem: Burgers model

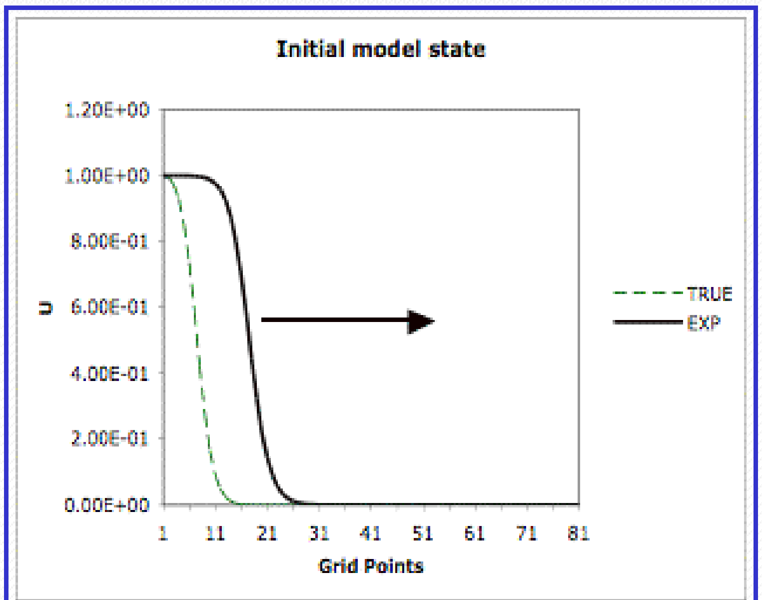
Test case: Discontinuous cubic observation operator

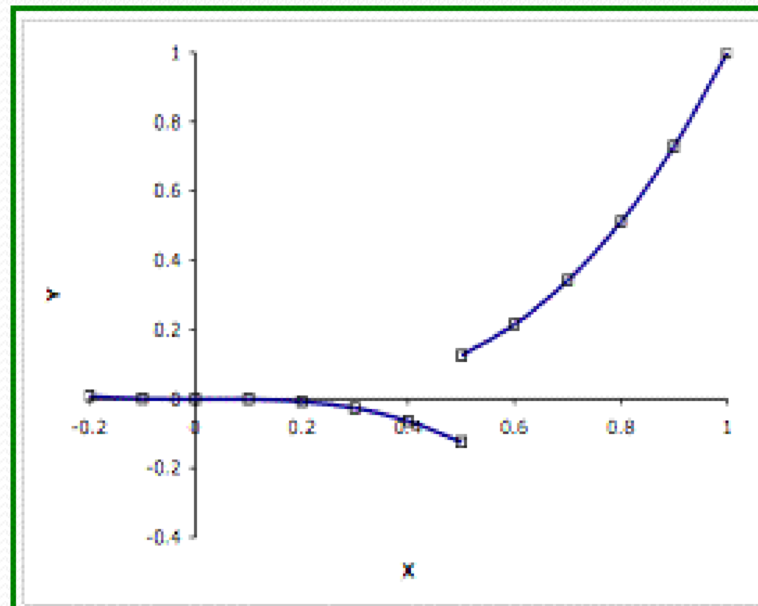
1-dimensional Burgers model simulating a shock-wave

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = \mu\frac{\partial^2 u}{\partial x^2}$$

Observation operator is a **non-differentiable cubic** function

$$h(u) = \begin{cases} u^3 & for \quad u \geq 0.5 \\ -u^3 & for \quad u < 0.5 \end{cases}$$





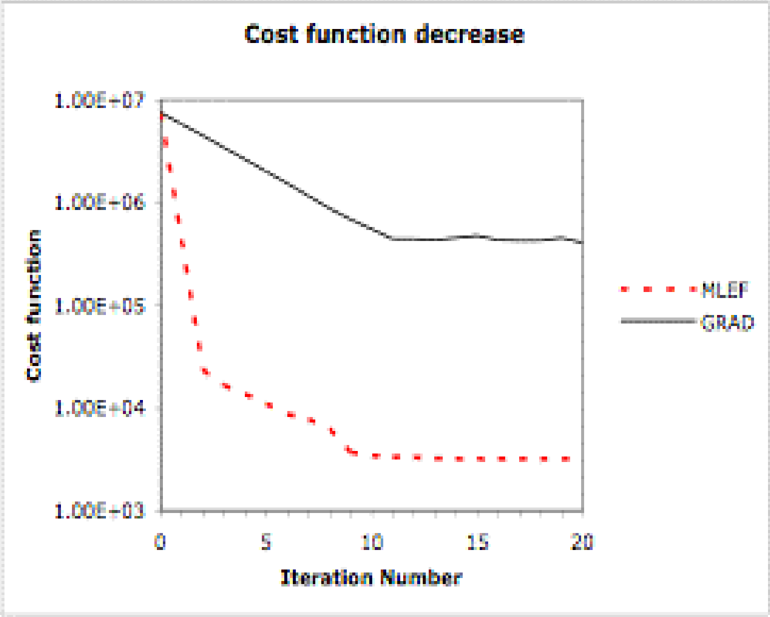
MLEF nonlinear minimization

(Zupanski et al. 2008)

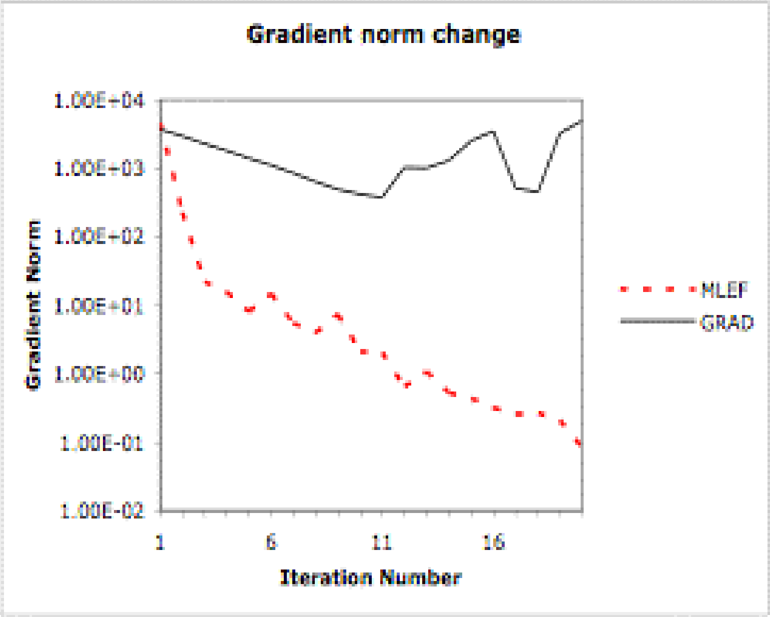
Exp 1 (MLEF): Use nonlinear differences $h(u + p_i^f) - h(u)$

Exp 2 (GRAD): Use the linearity/differentiability assumption $H\,\boldsymbol{p}_i^f$

Cost function

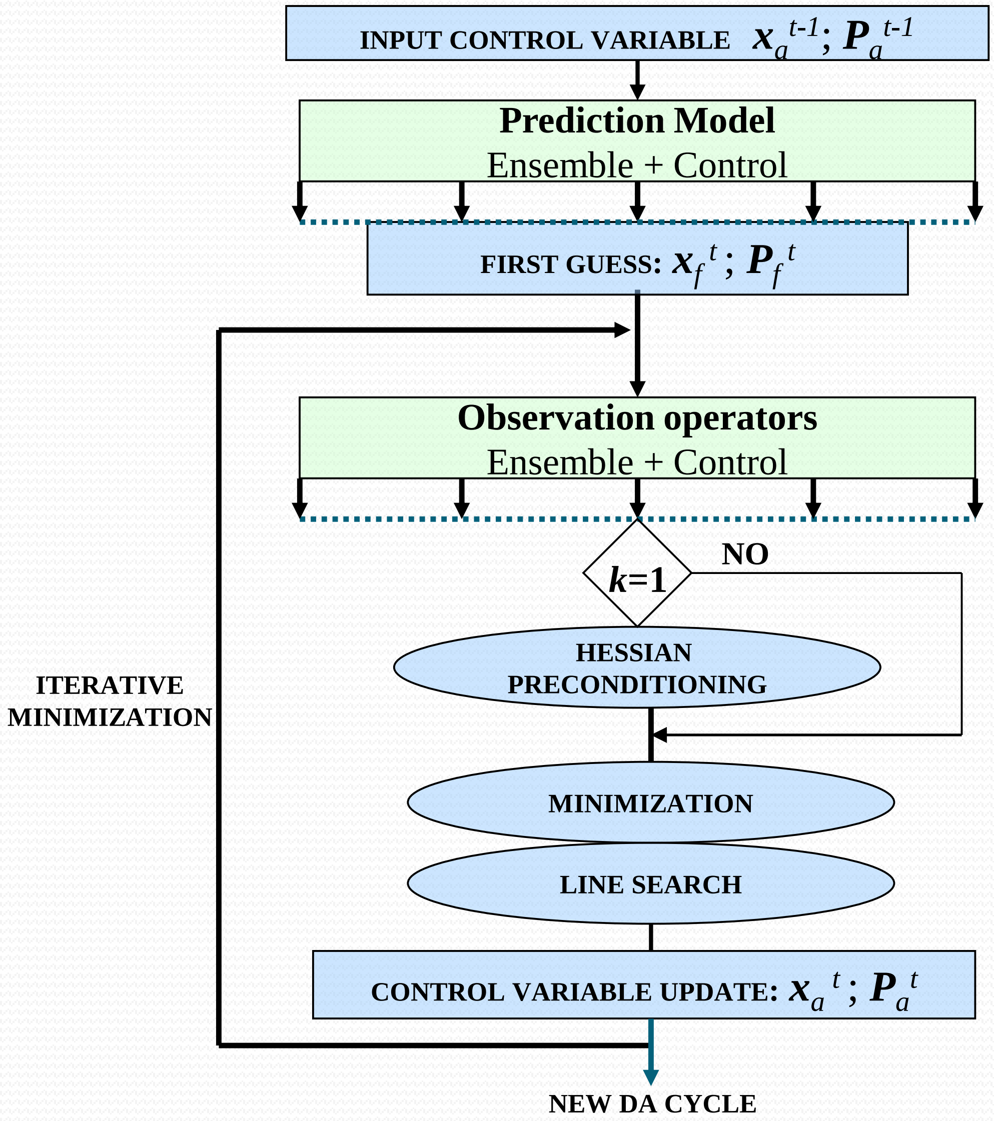


Gradient norm



Clear benefit of using finite difference representation of derivatives

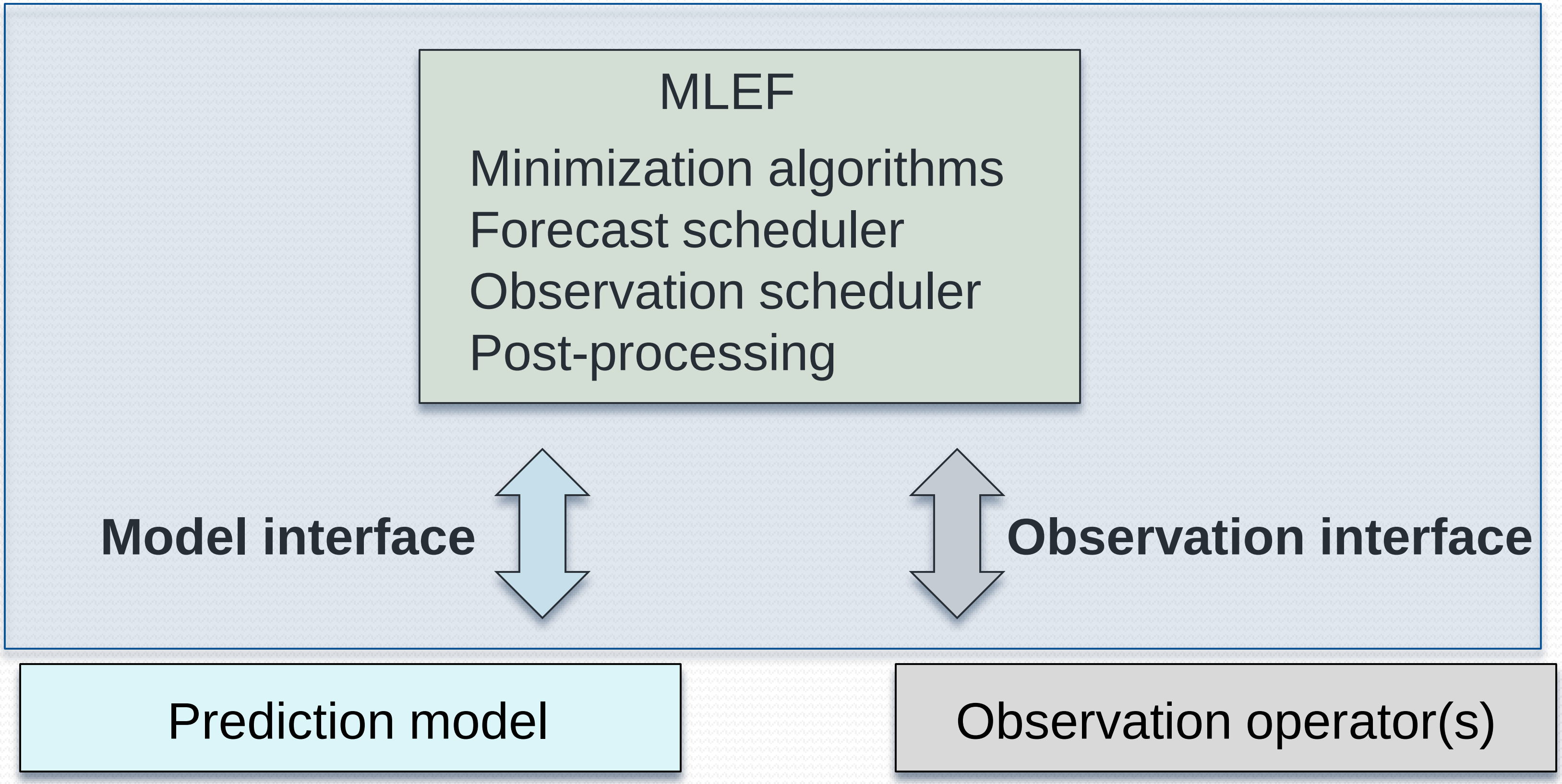
MLEF flow chart



- A hybrid between ensemble and variational DA
- Full-rank or reduced-rank
- Deterministic first guess forecast
- MAP: Analysis is the maximum of a posterior pdf
- Nonlinear analysis solution by an iterative minimization
- Improved minimization efficiency by an implicit Hessian preconditioning

Modular algorithm

- User-friendly compilation (self-contained libraries: LAPACK)
- Easy experiment specifications (control variable, minimization, localization)
- Parallel (MPI) – optional
- Fortran 90/95 - based



Understanding information flow in coupled DA:

single-point, 2-variable coupled system

- Consider a single-point coupled atmosphere-aerosol system with two variables:

(1) Atmosphere (x_{atms})

(2) Aerosol (x_{aero})

- Assume single observation of atmospheric variable y_{atms}

$$x = \begin{pmatrix} x_{atms} \\ x_{aero} \end{pmatrix} \qquad P_f = \begin{pmatrix} \sigma_{atms}^2 & \sigma_{atms}\sigma_{aero}\rho_{atms,aero} \\ \sigma_{atms}\sigma_{aero}\rho_{atms,aero} & \sigma_{aero}^2 \end{pmatrix}$$

$\rho_{atms,aero}$

correlation between atmosphere and aerosol

$$y = \begin{pmatrix} y_{atms} \\ 0 \end{pmatrix} \qquad R = \begin{pmatrix} (\sigma_R^2)_{atms} & 0 \\ 0 & (\sigma_R^2)_{aero} \end{pmatrix}$$

Analysis solution

Atmospheric analysis is identical to a standalone atmospheric DA

$$x^a_{atms} = \frac{1}{1+\varepsilon^2} x^f_{atms} + \frac{\varepsilon^2}{1+\varepsilon^2} y_{atms} \qquad \varepsilon^2 = \frac{(\sigma^2_f)_{atm}}{(\sigma^2_R)_{atm}}$$

Aerosol analysis depends on the cross-component correlation $\rho_{atms,aero}$

$$\rho_{atms,aero} = 0 \quad \Rightarrow \quad x^a_{aero} = x^f_{aero}$$

$$\rho_{atms,aero} \neq 0 \quad \Rightarrow \quad x^a_{aero} = x^f_{aero} + \rho_{atms,aero} \left(\frac{\sigma_{aero}}{\sigma_{atms}} \right) \frac{\varepsilon^2}{1+\varepsilon^2} (y_{atm} - x^f_{atm})$$

1- When the cross-component correlation is zero, there is no transfer of information between the two components

2- When the cross-component correlation exists, the information from atmospheric observation is transferred to the aerosol initial conditions, and thus the information flow and efficiency of the system is improved

Some coupled DA results with MLEF

Models (regional)

- WRF-Chem
  - CBMZ and MOZART chemistry
  - GOCART aerosol
- WRF-ARW
  - Noah land surface model (LSM)
  - Noah-MP LSM

Observations

- Atmospheric observations
  - NOAA GSI forward nonlinear component as observation operator
- Chemistry observations
  - NASA OMI total column o3, no2, so2
- Aerosol observations
  - NASA OMI aerosol optical depth (AOD)
- Gravity variations
  - NASA GRACE

Assimilation of OMI AOD observations

Experimental setup

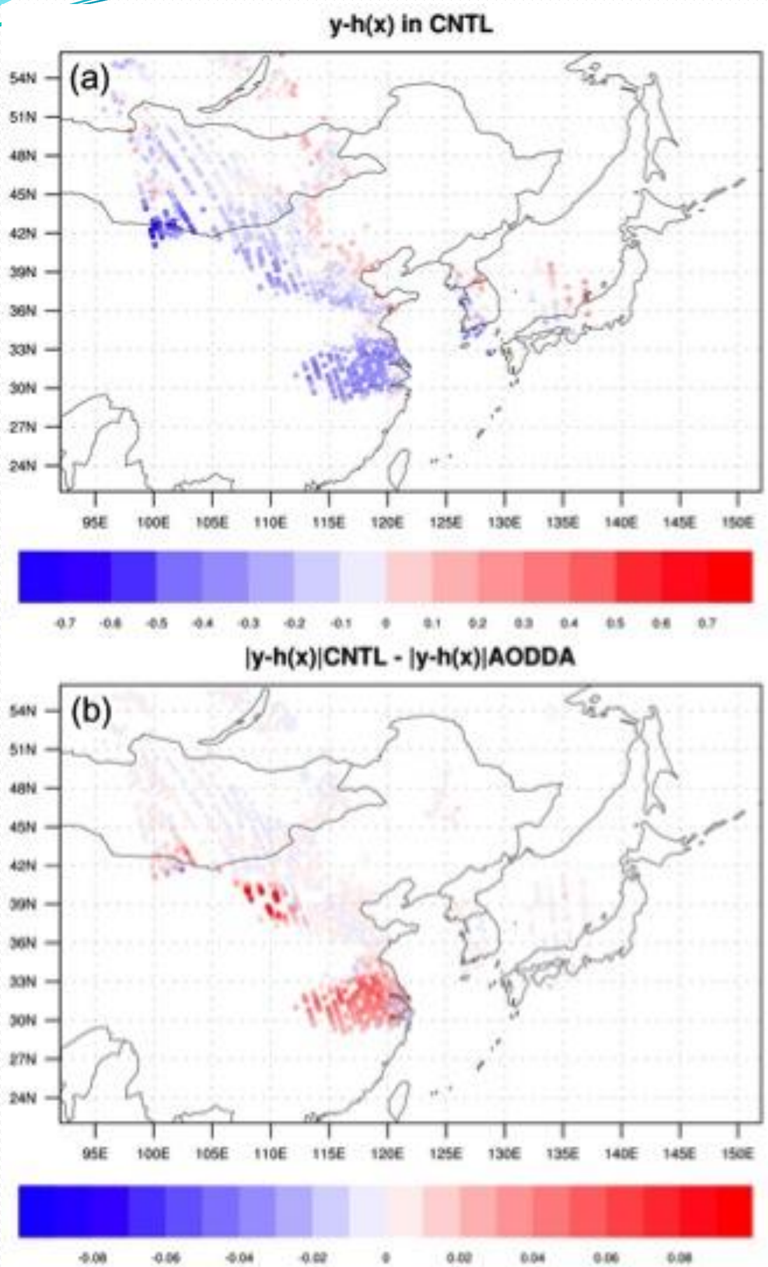
- WRF-CHEM model with with CBMZ chemistry and GOCART aerosol options
- 30 km / 31 layer
- OMI AOD observations at 500 nm (AOD observation operator)
- NOAA atmospheric observations (forward GSI operator)
- 32 ensembles, 6-hour assimilation period
- Control variables include mixing ratios of GOCART dust species (500, 1400, 2400, 4500, 8000 nm) and atmospheric variables (pressure, temperature, winds, humidity)
- Dust storm over Korea and Japan: 12-13 May 2011

Experiments for assessing the impact of AOD observations in data assimilation:

(1) Assimilation of atmospheric observations only (**ATM**),

(2) Assimilation of AOD and atmospheric observations (**AOD+ATM**)

DA verification in AOD observation space (c5)



Analysis Residuals
Valid 0600 UTC 13 May 2011

Control
(atmospheric observations only)

$$y - h(x^a)$$

AOD difference

$$\left|y - h(x^a)\right|_{CNTL} - \left|y - h(x^a)\right|_{AOD}$$

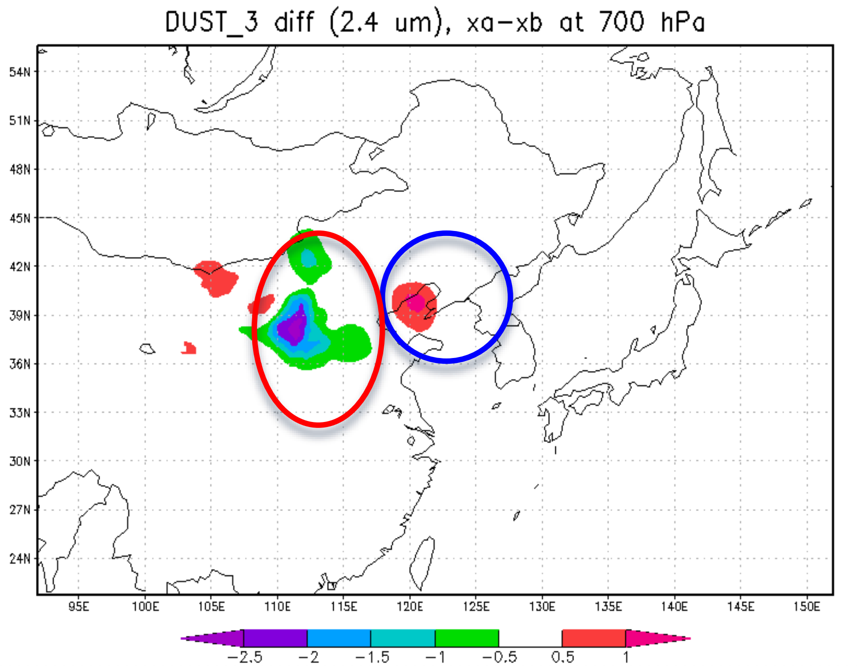
Positive difference (red color) implies an improvement of analysis due to additional AOD observations

Analysis increments (x^a-x^f)

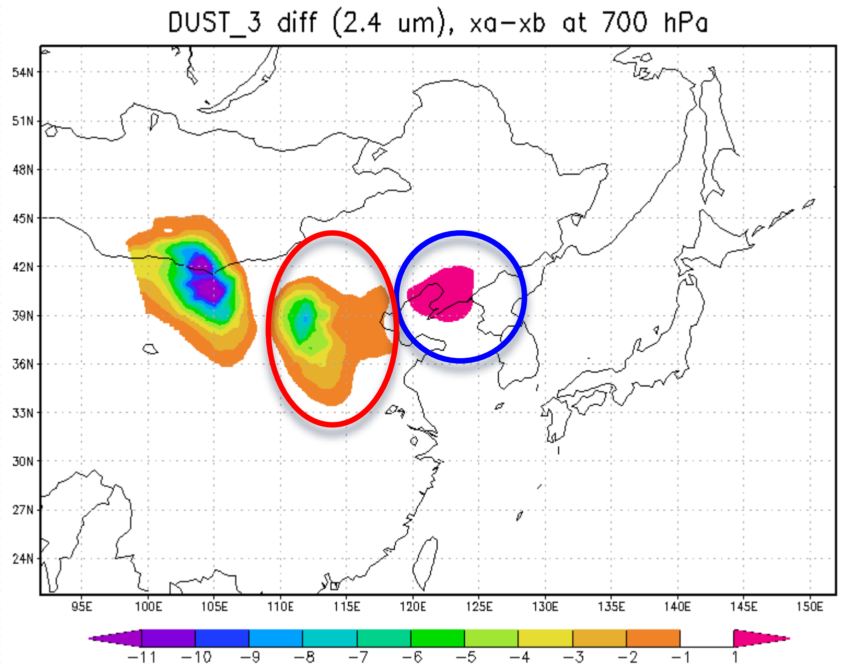
Impact of observations on dust variable

DUST_3 (2400 nm) at 700 hPa, valid 0600 UTC 12 May 2011

ATM (υg/kg-dry air)



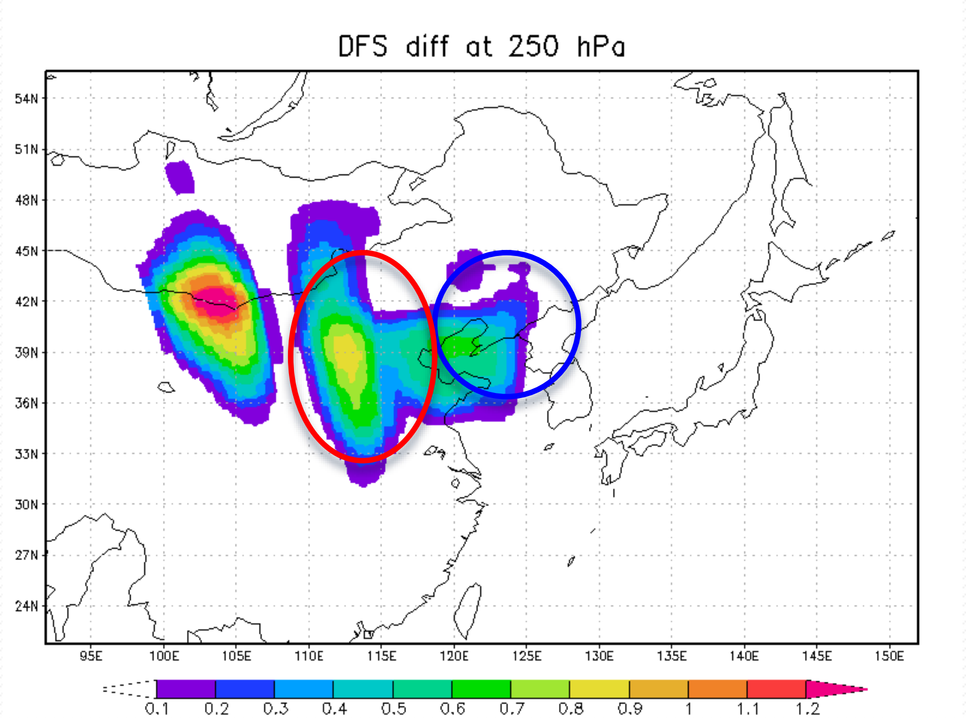
AOD+ATM (ug/kg-dry air)



- Atmospheric observations impact chemistry/aerosol analysis
- The impact is small in magnitude, but could be valuable
- AOD observations still bring important new information

Impact of AOD observations measured by DFS: 250 hPa

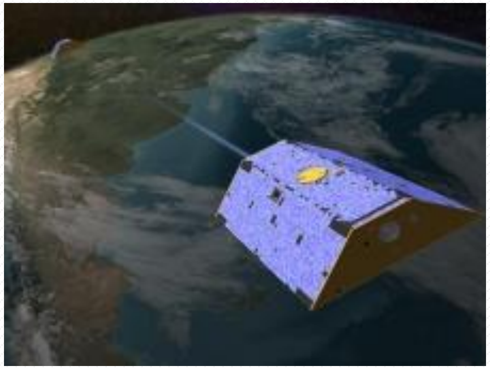
Difference: DFS (AOD+ATMS) – DFS (ATM)



- Most positive impact over the Gobi desert in Mongolia
- Some positive impact over China and Yellow sea

Assimilation of Gravity Recovery and Climate (GRACE) Equivalent Water Thickness (EWT) Observations

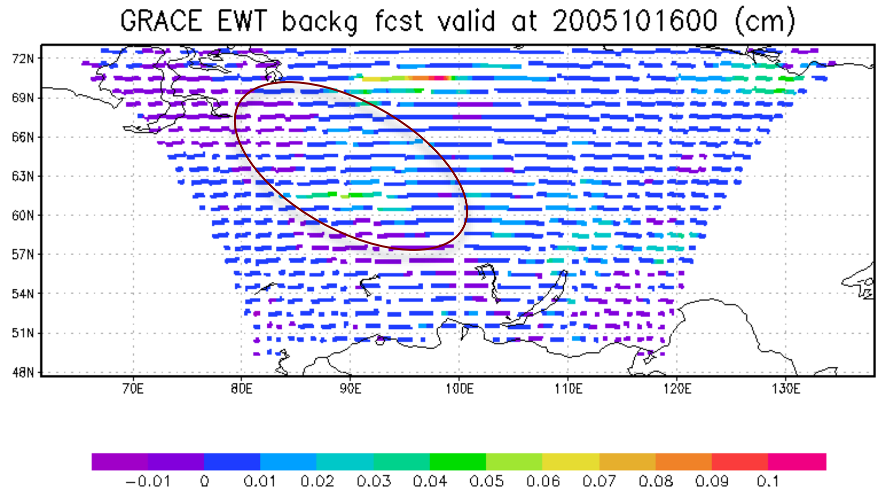
- Coupled atmosphere-soil hydrology data assimilation
- Control variables: temperature, pressure, wind, atmospheric humidity, snow, cloud variables, soil moisture, soil temperature
- EWT observations: *averaged 30 days differences of the column soil moisture, snow, and canopy water*
- Data assimilation interval: 30 days
- WRF model (30 km /31 layer) with Noah-MP LSM
- Geographical area: Siberia
- MLEF data assimilation method



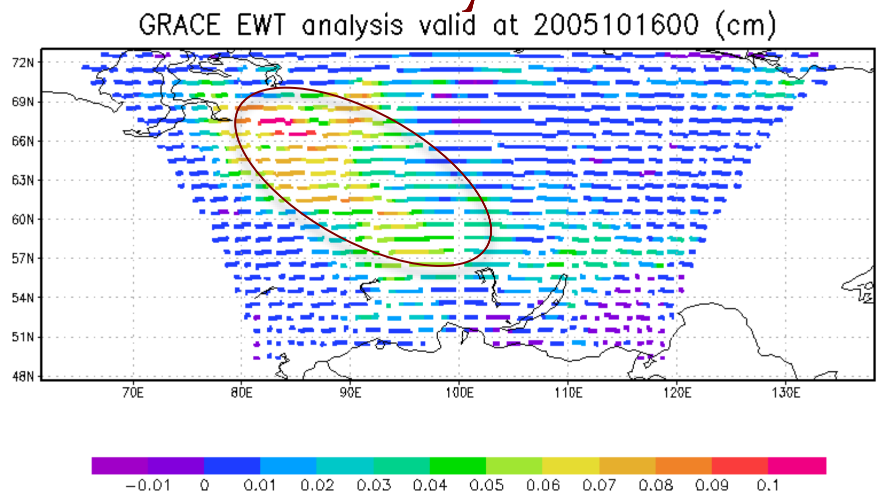
Examples of other GRACE related research

GRACE Data Assimilation Results: EWT Guess, Analysis and Observations

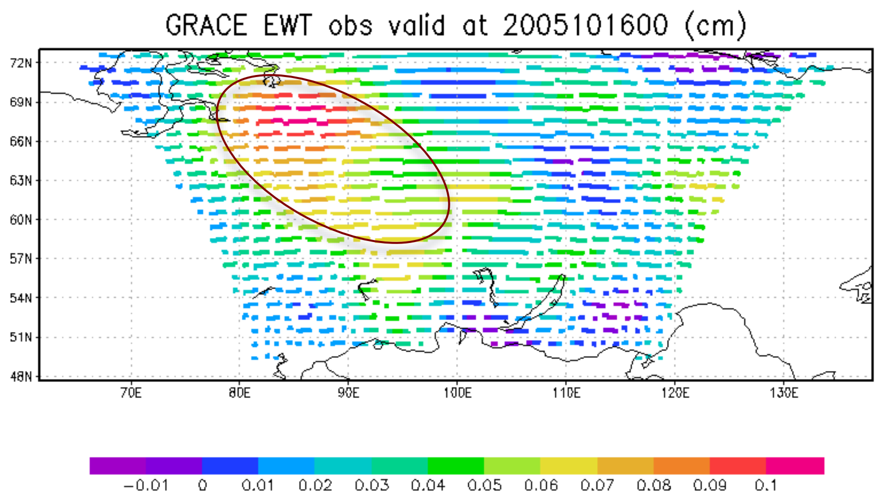
Guess



Analysis



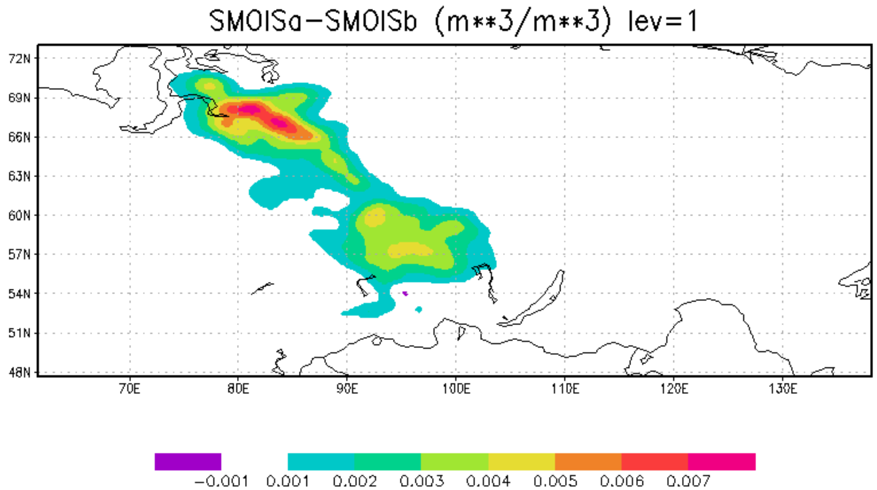
Observations



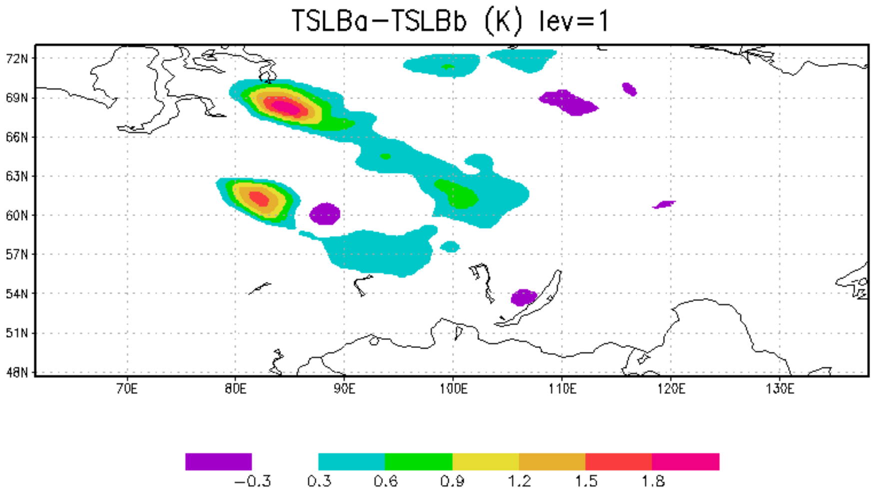
Analysis is closer to observations

GRACE Data Assimilation Results:
Impact on Land Surface analysis increments

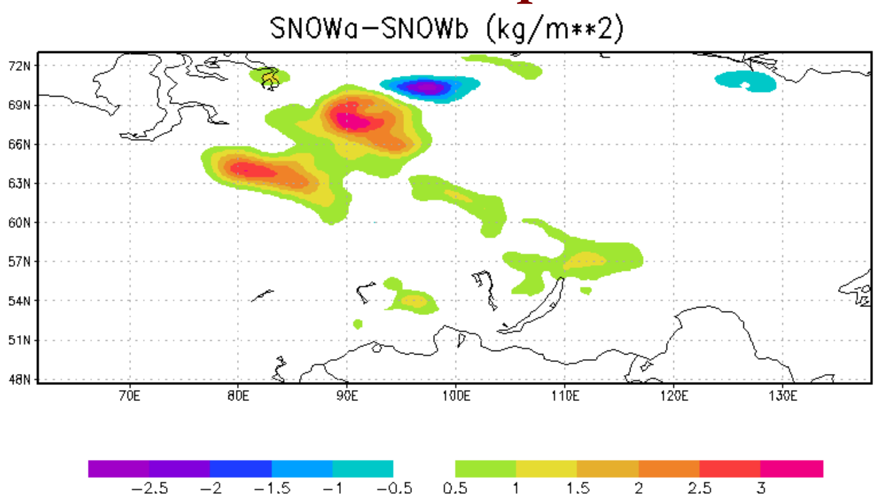
Surface Soil Moisture



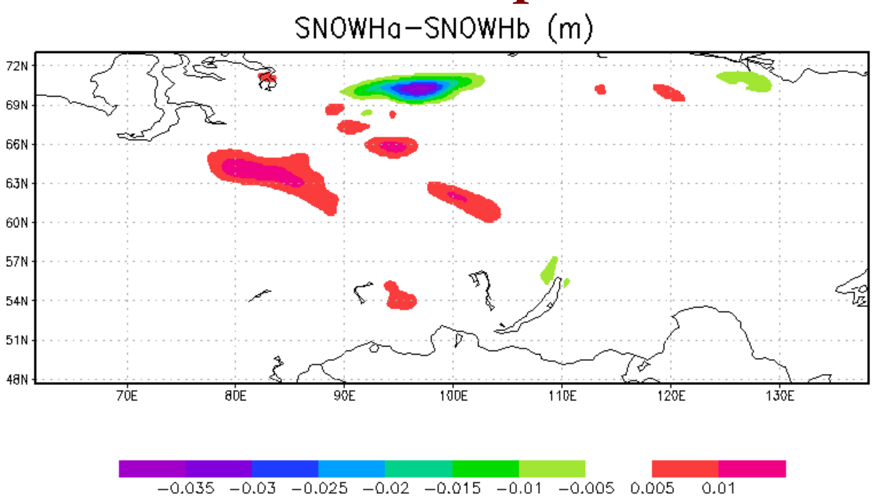
Surface Soil Temperature



Snow Water Equivalent

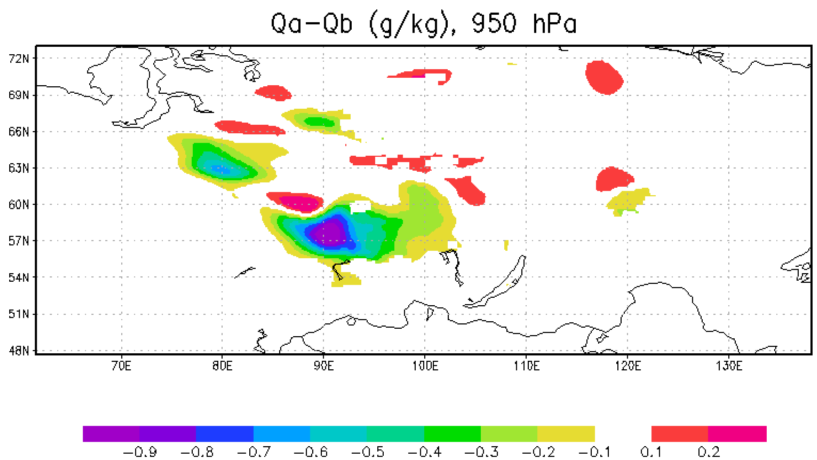


Snow Depth

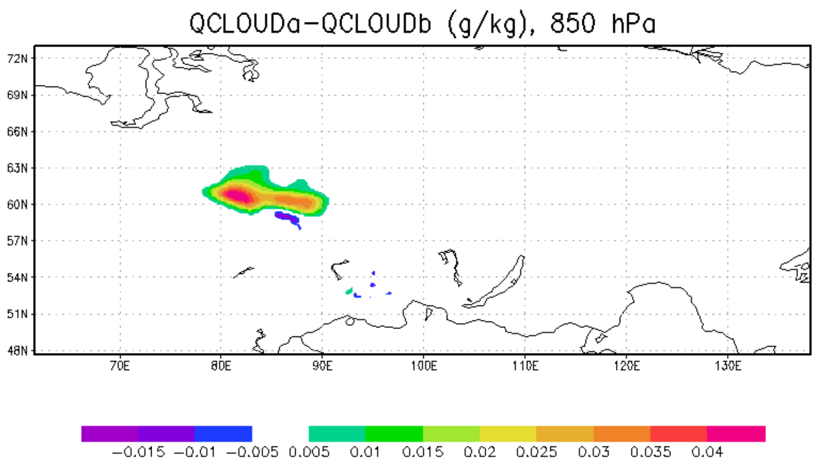


GRACE Data Assimilation Results:
Impact on *Unobserved* Atmospheric analysis increments

Specific humidity at 950 hPa



Cloud Water at 850 hPa



- Impact of EWT observations on atmospheric initial conditions
- Possible due to cross-component correlations between land surface and atmosphere

Coupled atmosphere – land surface DA: single observation experiments

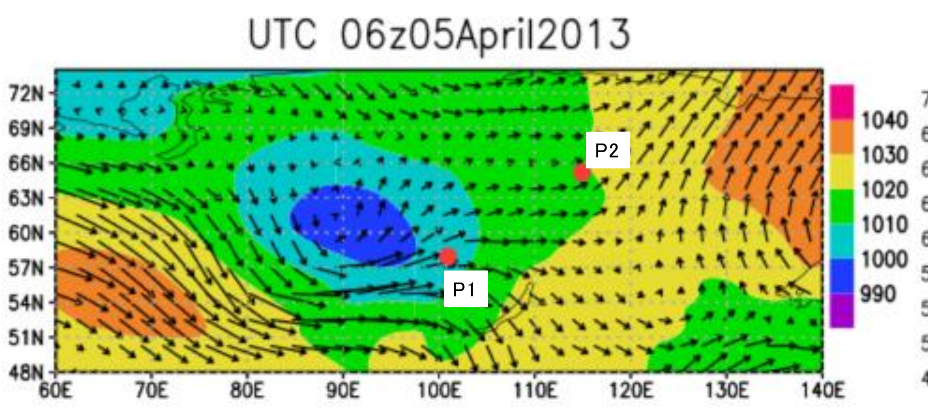
Synoptic situation
- 0600 UTC on April 5, 2013
- Passing low-pressure system over south Siberia

Modeling/DA system:
- WRF with Noah and Noah-MP LSMs
- 30 km/31 layers
- MLEF with 32 ensembles
- 6-hour assimilation window
- Control variables include atmospheric and soil variables

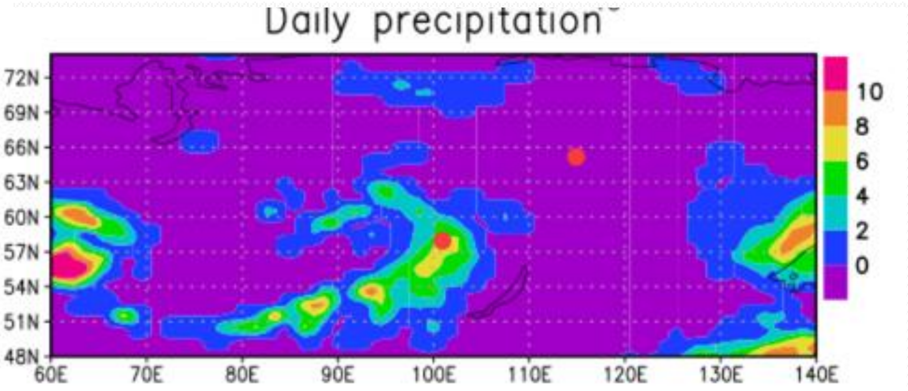
Experiments:
- Examine the impact of cross-component covariance between T-2m and soil variables by performing a single observation experiments with atmospheric T-2m observation
- Examine the flow-dependent impact of land surface models on cross-component covariance

Synoptic situation

Wind at 500 hPa
Valid 0600 UTC on April 5, 2013



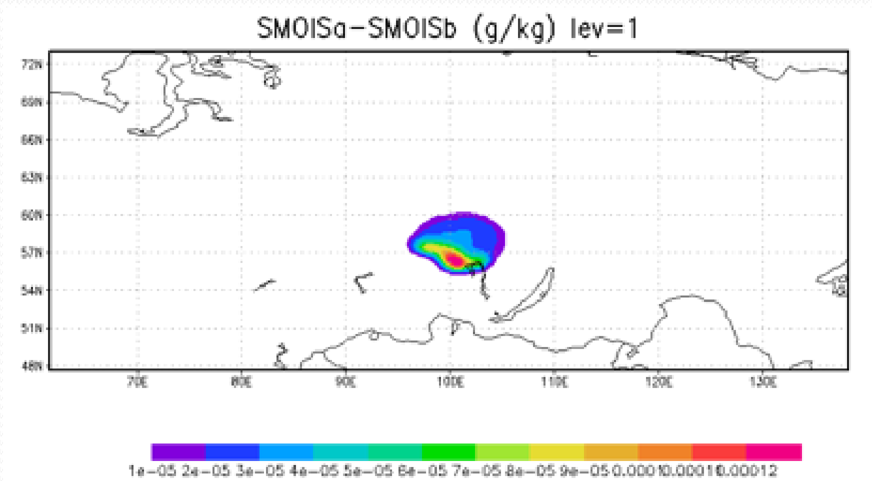
Accumulated precipitation (mm)
Valid 0600 UTC on April 5, 2013



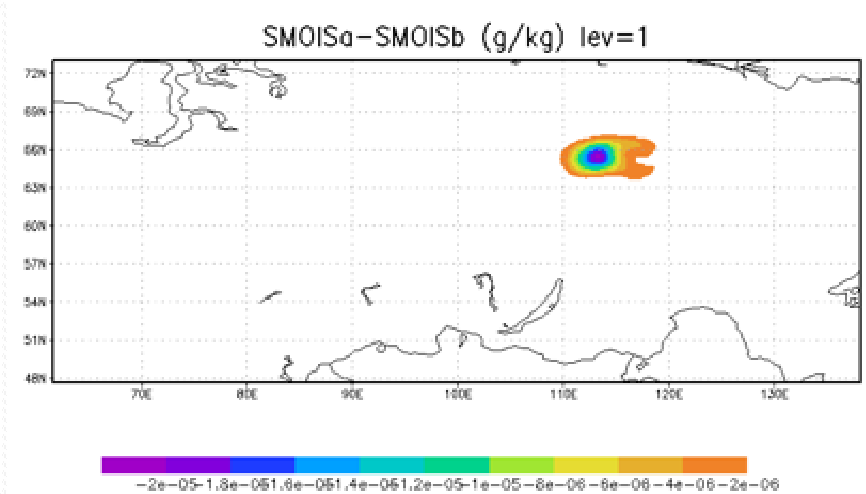
- Sea level pressure (shaded) indicates a strong low-pressure system
- Precipitation pattern suggest a frontal zone associated with the low-pressure system
- Above freezing values at site P1, but below freezing at site P2
- Deeper snow depth and complete coverage over the site P2, while at site P1 the snow depth is smaller

Soil moisture response with Noah-MP (top soil layer)

P1



P2



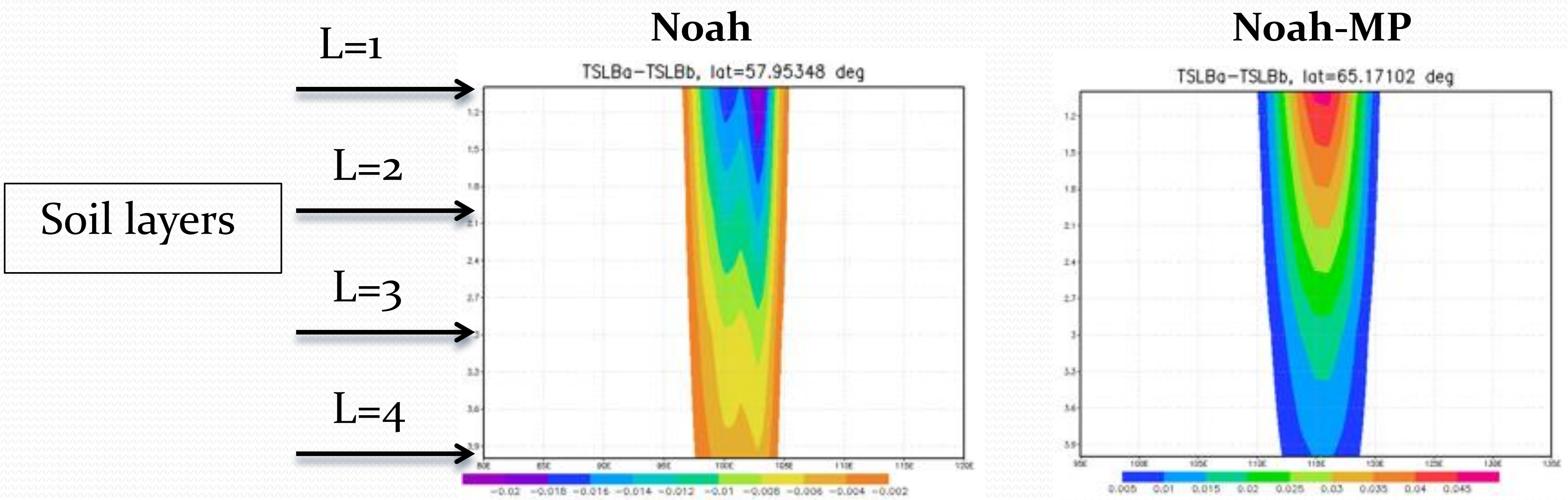
- Different sign and strength of the response:

  - P1 has positive and relatively strong response (0.1 g/kg)

  - P2 has negative and weak response (0.001 g/kg) – snow-covered surface

Flow-dependent covariance important for accurate representation of coupled DA system

Soil temperature response, P1 site



- Well-defined, localized response in vertical and horizontal
- Strongest response near the top layer
- Noah has a weak negative response (~0.02 K)
- Noah-MP has a stronger positive response (0.045 K)

Cross-component covariance is significantly impacted by the LSM model

Coupled chemistry-atmosphere: Assimilation of synthetic GEMS radiances

Experimental setup

- WRF-CHEM model with with MOZART chemistry
- 30 km / 51 layer
- Model top at 10 hPa
- 32 ensembles, 6-hour assimilation period
- Control variables include the mixing ratios of MOZART chemistry (o3,no2,so2,co,hcho) and atmospheric variables (pressure, temperature, winds, humidity)
- Storm over Korea and Japan: 10-13 May 2011

Observations:

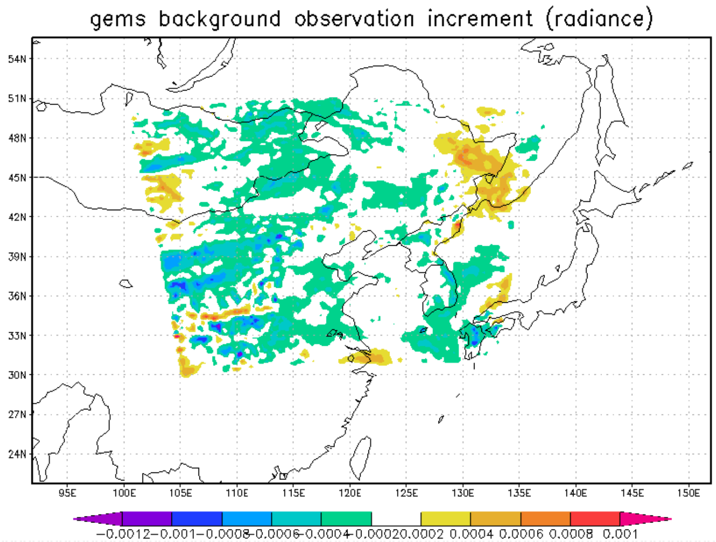
- GEMS synthetic radiance observations at 320 nm wavelength
- OMI total column chemistry observations (o3, no2, so2)
- NOAA atmospheric observations (forward GSI operator)

DA impact in GEMS radiance observation space

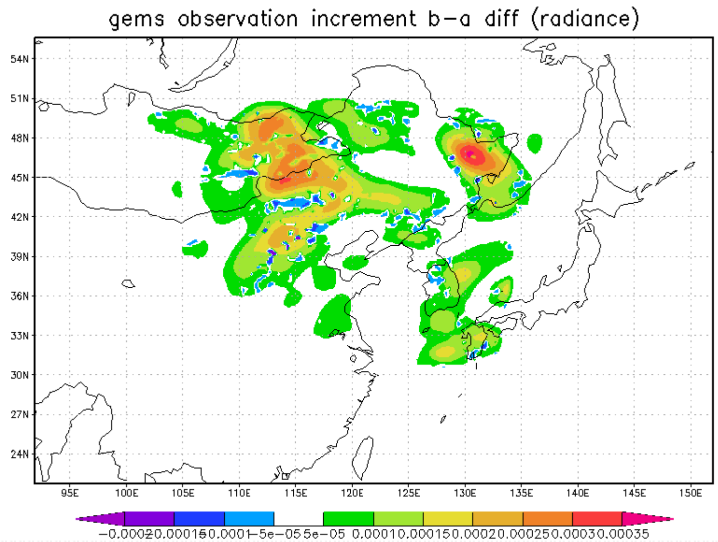
Innovation vector for radiance

Valid 0600 UTC 10 May 2011

$$y_{obs} - h(x^f)$$



$$|y_{obs} - h(x^f)| - |y_{obs} - h(x^a)|$$



- GEMS radiance innovation shows reduction due to assimilation, indicated by positive values on the right panel
- GEMS RMS error reduced by 10 %
- Total cost function decrease by 20%

Assimilation of atmospheric, chemistry and synthetic radiance observations helps improve all components of the coupled system

Important issues: Error covariance localization in coupled DA

- Model/analysis space: $r \circ P_f$
- Observation space: $r \circ R^{-1}$

Consider land surface – atmosphere coupled system

- In general, different localization length for land and for atmosphere
- Assume atmospheric observation close to surface:
  - surface temperature, surface specific humidity, low-frequency microwave radiance (e.g., AMSR2)
  - impacts both land surface and atmosphere
- With observation space localization, it is difficult to use proper localization scales when observation is "shared" between two components
- Model space localization appears to be a more natural, straightforward approach
- Important choice of vertical coordinate in the analysis

Important issues: Quantifying coupling strength

Joint entropy of two processes, X_1 and X_2, is:

$$H(X_1, X_2) = -\int_{x_1}\int_{x_2} p(x_1, x_2) \log p(x_1, x_2)\, dx_1\, dx_2$$

Mutual information (*MI*) measures the information shared by two processes:

$$MI(X_1, X_2) = -\int_{x_1}\int_{x_2} p(x_1, x_2) \log\left[\frac{p(x_1, x_2)}{p(x_1)p(x_2)}\right] dx_1\, dx_2$$

Mutual information has values between 0 and infinity:

MI=0	=> independent processes
MI="large"	=> fully dependent processes

One can use mutual information to quantify the coupling strength:
(1) **Strong coupling** implies intense sharing of information, thus **large MI**
(2) **Weak coupling** implies marginal sharing of information, thus **small MI**

Quantifying coupling strength: Gaussian pdfs

(Zupanski 2016)

With Gaussian pdfs, the mutual information is defined in terms of the covariance P_f

$$P_f = \begin{pmatrix} P_{11} & P_{12} \\ P_{12}^T & P_{22} \end{pmatrix}$$

$$MI(X_1, X_2) = -\frac{1}{2} \ln \frac{\det(P_f)}{\det(P_{11})\det(P_{22})} \quad \textbf{(A)}$$

Using $\det(P_f) = \det(P_{11})\det(P_{22})\det(I - P_{11}^{-1}P_{12}P_{22}^{-1}P_{12}^T)$

$$MI(X_1, X_2) = -\frac{1}{2} \ln\left[\det\left(I - P_{11}^{-1}P_{12}P_{22}^{-1}P_{12}^T\right)\right] \quad \textbf{(B)}$$

- Mutual information can be used in practical data assimilation to quantify coupling strength
- Formulation (**A**) may be preferable for computational stability since it avoids matrix inversion
- Formulation (**B**) is more revealing as it relates to correlations

Summary and Future

- **Information flow is enhanced by using cross-component covariances, leading to more efficient use of available observations**
- **Initial effort in developing coupled DA should focus on estimating, and eventually utilizing, cross-component covariances**
- **Need to consider additional options for the control variable, such as empirical parameters, systematic model error, moving time average**
- **Introduce entropy measures for quantifying coupling strength**
- **Error covariance localization for coupled systems requires more research and better understanding**
- **Possibly use more general formulation of analysis coordinates to describe localization length in coupled systems**

Thank you!

References:

Lim, S. J., Park, S. K., and M. Zupanski, 2015: Ensemble data assimilation of total column ozone using a coupled meteorology–chemistry model and its impact on the structure of Typhoon Nabi (2005). *Atmos. Chem. Phys.*, **15**, 10019-10031. doi:10.5194/acp-15-10019-2015

Lokupitiya, R.S., D. Zupanski, A.S. Denning, S.R. Kawa, K.R. Gurney, M. Zupanski, W. Peters, 2008: Estimation of global CO_2 fluxes at regional scale using the maximum likelihood ensemble filter. *J. Geophys. Res.*, **113**, D20110, doi:10.1029/2007JD009679.

Park SK, Lim S, Zupanski M (2015) Structure of forecast error covariance in coupled atmosphere–chemistry data assimilation. *Geosci. Model Dev*. 8: 1315–1320.

Tran, A. P., M. Vanclooster, M. Zupanski, and S. Lambot, 2014: Joint estimation of soil moisture profile and hydraulic parameters by ground-penetrating radar data assimilation with maximum likelihood ensemble filter. *Water Resour. Res*., **50**, doi:10.1002/2013WR014583.

Zupanski, M., 2005: Maximum Likelihood Ensemble Filter: Theoretical Aspects. *Mon. Wea. Rev*., **133**, 1710–1726.

Zupanski, M., I. M. Navon, and D. Zupanski, 2008: The Maximum Likelihood Ensemble Filter as a non-differentiable minimization algorithm. *Q. J. R. Meteorol. Soc*., **134**, 1039-1050.

Zupanski M, 2016: Data assimilation for coupled modeling systems. *Data Assimilation for Atmospheric, Oceanic, and Hydrologic Applications (Vol. III)*, S.-K. Park and L. Xu, Eds, Springer Heidelberg New York Dordrecht London, in print.

Degrees of Freedom for Signal (DFS) calculation in MLEF

Change of entropy / degrees of freedom for signal in Gaussian framework:

$$DFS = trace\left[I - P_a P_f^{-1} \right]$$

In reduced-rank ensemble system, such as MLEF, d_s can be computed exactly

$$DFS = trace\left[(I + Z^T Z)^{-1} Z^T Z \right]$$

$$Z_i = R^{-1/2}[h(x + p_i^f) - h(x)] \qquad Z^T Z = U \Lambda U^T$$

$$DFS = \sum_i \frac{\lambda_i}{1+\lambda_i}$$

- Eigenvalue decomposition is used in MLEF to invert the Hessian matrix
- DFS is a by-product of MLEF