



# Ensemble Kalman Filter

**Dusanka Zupanski**  
**CIRA/Colorado State University**  
**Fort Collins, Colorado**

*Advanced Numerics Seminar*  
*8 March 2006, Langen, Germany*

***Acknowledgements:***

***M. Zupanski, G. Carrio, S. Denning, M. Uliasz, R. Lokupitya, CSU***  
***A. Hou and S. Zhang, NASA/GMAO***

Dusanka Zupanski, CIRA/CSU  
Zupanski@CIRA.colostate.edu



# OUTLINE

- **Why Ensemble Data Assimilation?**
- **Kalman filter and Ensemble Kalman filter**
- **Maximum likelihood ensemble filter (MLEF)**
- **Examples of MLEF applications**
- **Future research directions**

# Why Ensemble Data Assimilation?

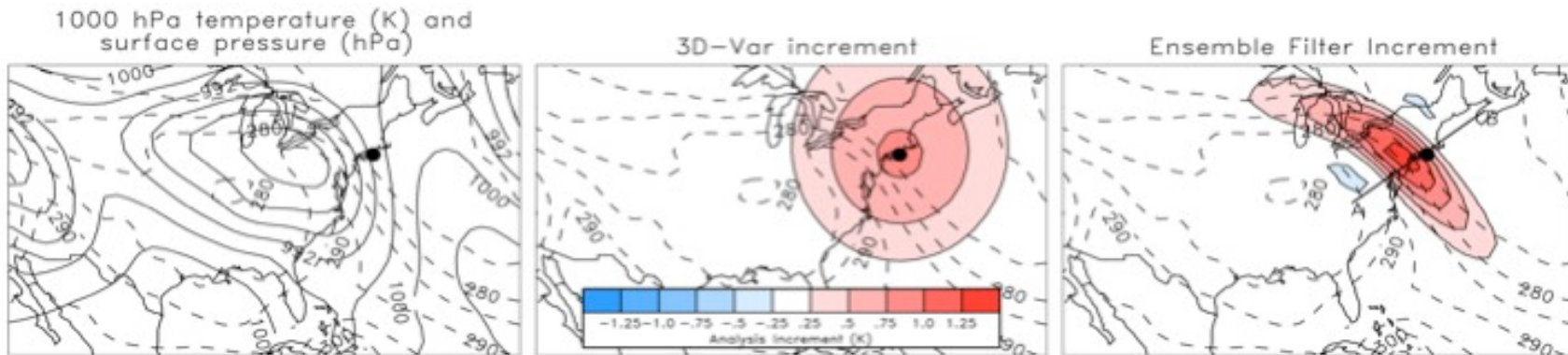
Three main reasons :

- ❑ Need for optimal estimate of the atmospheric state + **verifiable uncertainty** of this estimate;
- ❑ Need for **flow-dependent** forecast error covariance matrix; and
- ❑ The above requirements should be applicable to **most complex atmospheric models** (e.g., non-hydrostatic, cloud-resolving, LES).

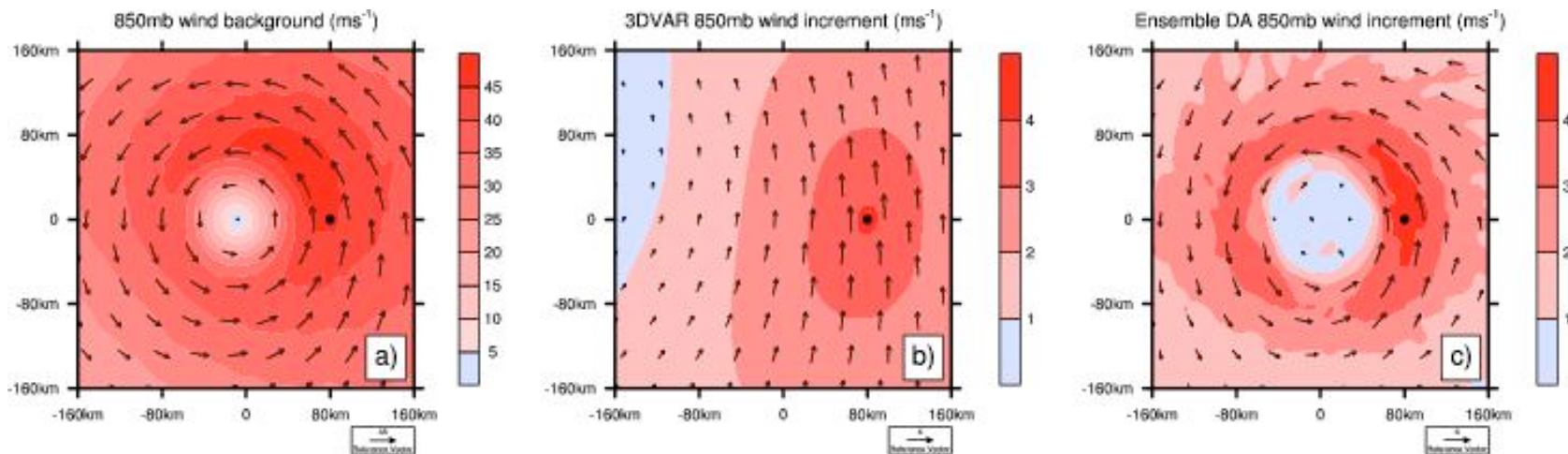
# Benefits of Flow-Dependent Background Errors

(From Whitaker et al., THORPEX web-page)

## Example 1: Fronts



## Example 2: Hurricanes



## Are there alternatives?

Two good candidates:

- ❑ **4d-var** method: It employs flow-dependent forecast error covariance, but it does not propagate it in time.
- ❑ **Kalman Filter (KF)**: It does propagate flow-dependent forecast error covariance in time, but it is too expensive for applications to complex atmospheric models.

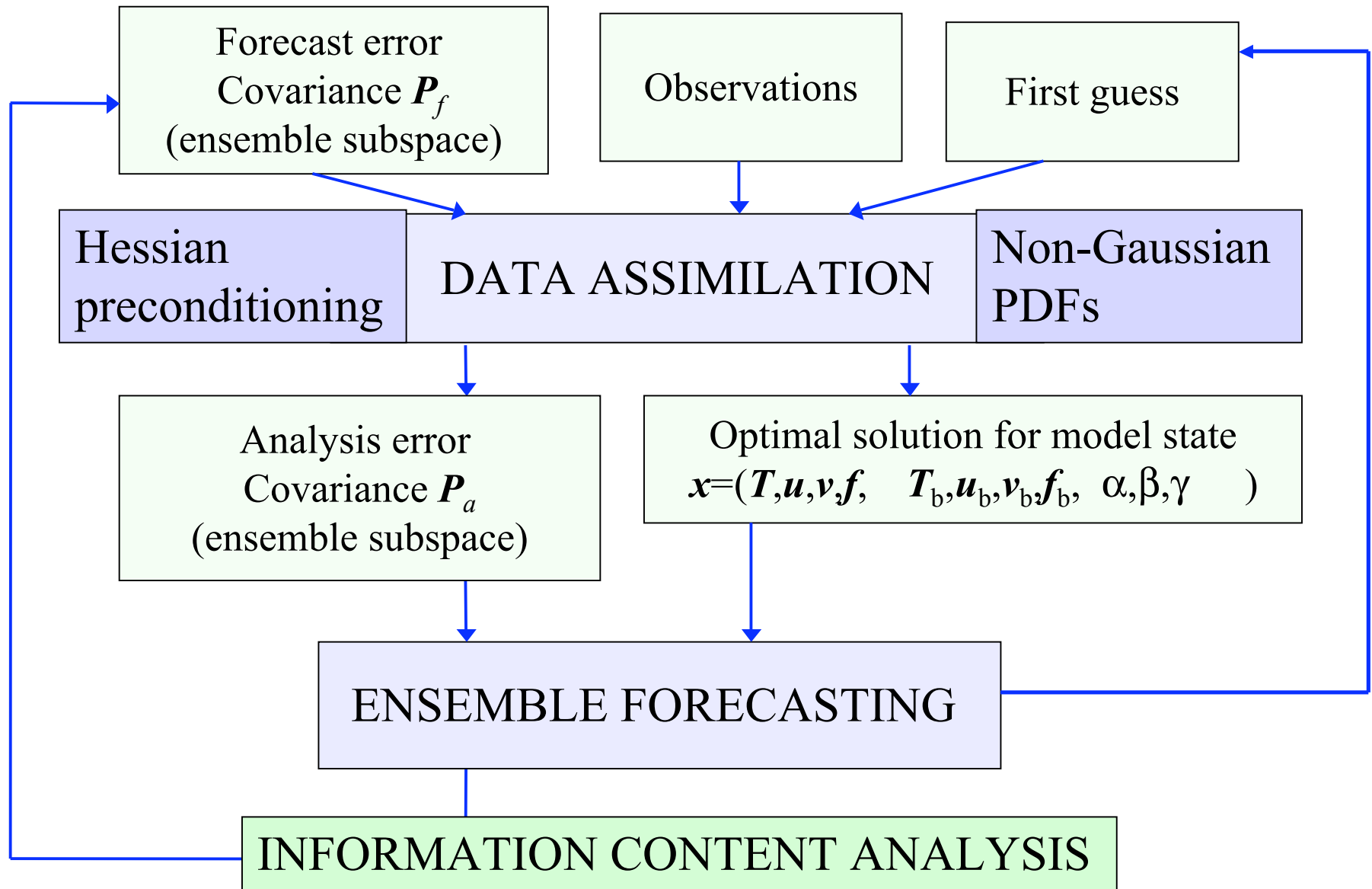


EnKF is a practical alternative to KF, applicable to most complex atmospheric models.



A bonus benefit: EnKF does not use adjoint models!

# Maximum Likelihood Ensemble Filter



# Data Assimilation Equations

**GOAL: Combine Model and Data to obtain optimal estimate of dynamical state  $x$**

**Equations in model space:**

$$\mathbf{x}_n = M_{n,n-1}(\mathbf{x}_{n-1}) + G(\mathbf{x}_{n-1})\mathbf{w}_{n-1}$$

$n$  - Time step index

$M$  - Dynamical model for model state evolution (e.g., NWP model)

$G$  - Dynamical model for state dependent model error

$\mathbf{x}$  - Model state vector of dim  $Nstate$  ;  $\mathbf{w}$  - Model error vector of dim  $Nstate$

Prior (forecast) error covariance of  $x$  (assumed known):

$$\mathbf{P}_f = E \left\{ [\mathbf{x} - E(\mathbf{x})][\mathbf{x} - E(\mathbf{x})]^T \right\} ; E - \text{Mathematical expectation}$$

Model error covariance (assumed known):

$$\mathbf{Q} = E(\mathbf{w}\mathbf{w}^T)$$

# Data Assimilation Equations

## Equations in data space:

$$y_n = H_n(x_n) + \epsilon_n$$

$n$  - Time step index (denoting observation times)

$y$  - Observations vector of dim  $Nobs$  ;  $H$  - Observation operator

$\epsilon$  - Observation error

Observation error covariance, includes also representatives error (assumed known):

$$R = E(\epsilon\epsilon^T)$$

Data assimilation should combine model and data in an optimal way. Optimal solution  $z$  can be defined in terms of optimal initial conditions  $x_a$  (analysis), model error  $w$ , and empirical parameters  $\alpha, \beta, \gamma$ .

How can we obtain optimal solution?  
Two approaches are used most often:

**Approach 1:** Optimal solution (e.g., analysis  $x_a$ ) = Minimum variance estimate, or conditional mean of Bayesian posterior probability density function (PDF) (e.g., **Kalman filter**; **Extended Kalman filter**; **EnKF**)

$$x_a = E(x|y) = \int x p(x|y) dx = \int x \frac{p(y|x)p(x)}{p(y)} dx \quad p - \text{PDF}$$

$x_a$  is defined as mathematical expectation (i.e., mean) of the conditional posterior  $p(x|y)$ , given observations  $y$  and prior  $p(x)$ .

Assuming linear  $M$  and  $H$  and independent Gaussian PDFs  
 $\Rightarrow$  **Kalman filter** solution (e.g., Jazwinski 1970)

For non-linear  $M$  or  $H$  the solution can be obtained employing  
**Extended Kalman filter**, or **Ensemble Kalman filter**.

**Approach 2:** Optimal solution (e.g., analysis  $x_a$ ) = Maximum likelihood estimate, or conditional mode of Bayesian posterior  $p(x|y)$  (e.g., **variational** methods; **MLEF**)

$$\mathbf{x}_a = \max [p(\mathbf{x}|\mathbf{y})] = \max \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})} = \min \left\{ -\log [p(\mathbf{x}|\mathbf{y})] \right\}$$

$\mathbf{x}_a$  = Maximum of posterior  $p(x|y)$ , given observations and prior  $p(x)$ .

For independent Gaussian PDFs, this is equivalent to minimizing cost function J:

$$J = \frac{1}{2} [\mathbf{x} - \mathbf{x}_b]^T \mathbf{P}_f^{-1} [\mathbf{x} - \mathbf{x}_b] + \frac{1}{2} [H[M(\mathbf{x})] - \mathbf{y}]^T \mathbf{R}^{-1} [H[M(\mathbf{x})] - \mathbf{y}]$$

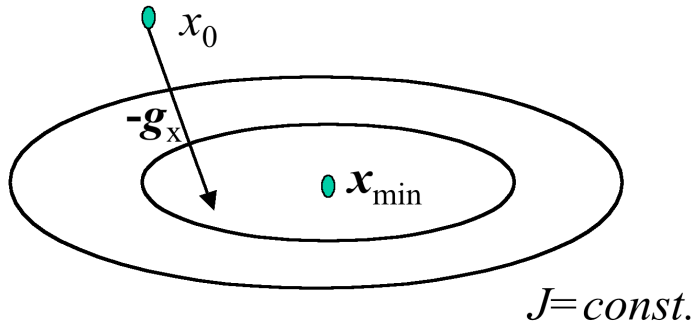
Solution can be obtained (with ideal preconditioning) in one iteration for **linear H** and **M**. Iterative solution for **non-linear H** and **M**:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \mathcal{H}^{-1} \nabla J_k$$

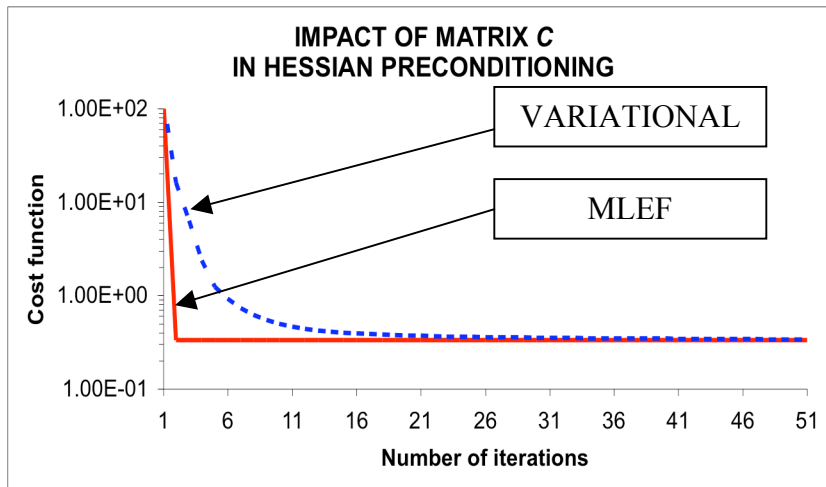
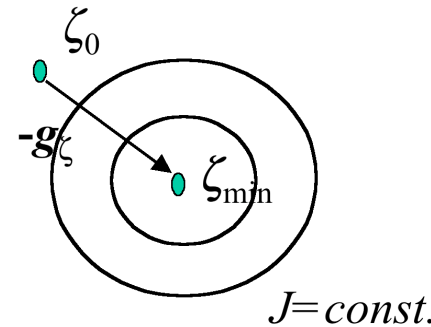
$\mathcal{H}^{-1}$  - Preconditioning matrix = inverse Hessian of J

# Ideal Hessian Preconditioning

**Physical space ( $x$ )**



**Preconditioning space ( $\zeta$ )**



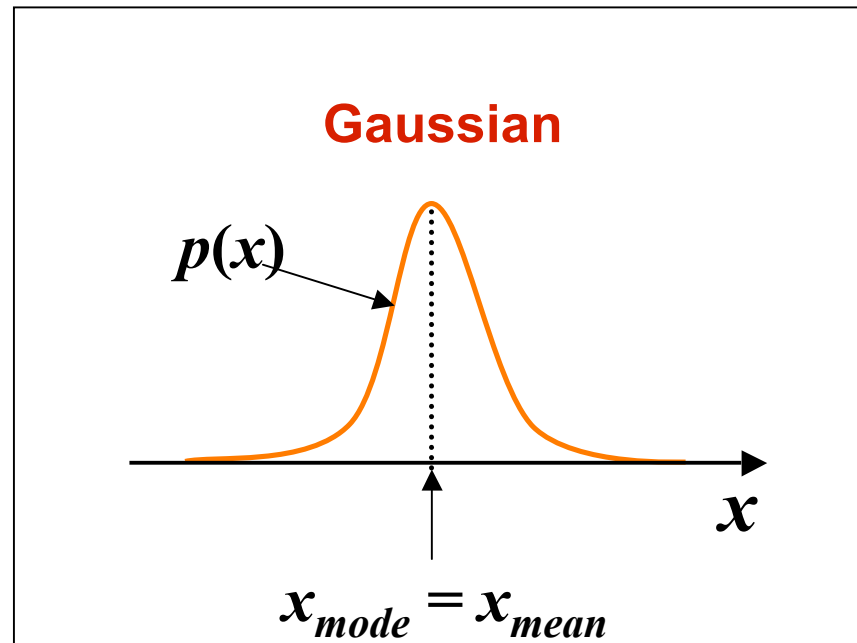
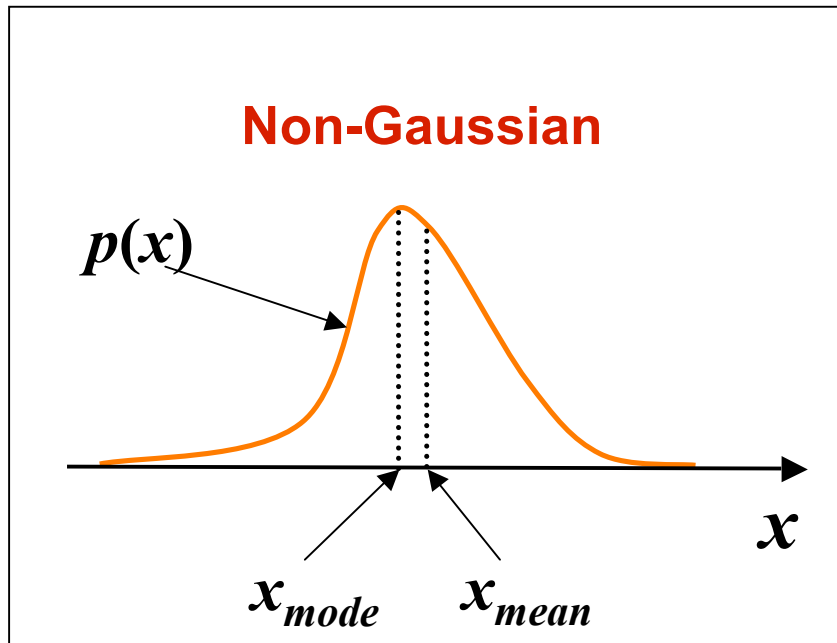
$$\mathbf{H}^{-1} = \left( \frac{\partial^2 J}{\partial \mathbf{x}^2} \right)^{-1} = \mathbf{P}_f^{1/2} (\mathbf{I} + \mathbf{C})^{-1} \mathbf{P}_f^{T/2}$$

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{P}^{-1} \mathbf{g}_k$$

$$\mathbf{P}_{MLEF}^{-1} = \mathbf{P}_f^{1/2} (\mathbf{I} + \mathbf{C})^{-1} \mathbf{P}_f^{T/2}$$

$$\mathbf{P}_{VAR}^{-1} = \mathbf{P}_f$$

## MEAN vs. MODE



Minimum variance estimate = Maximum likelihood estimate!

For Gaussian PDFs and **linear H** and **M** results of all methods [KF, EnKF (with enough ensemble members), and variational] should be identical, assuming the same  $P_f$  and  $y$  are used in all methods.

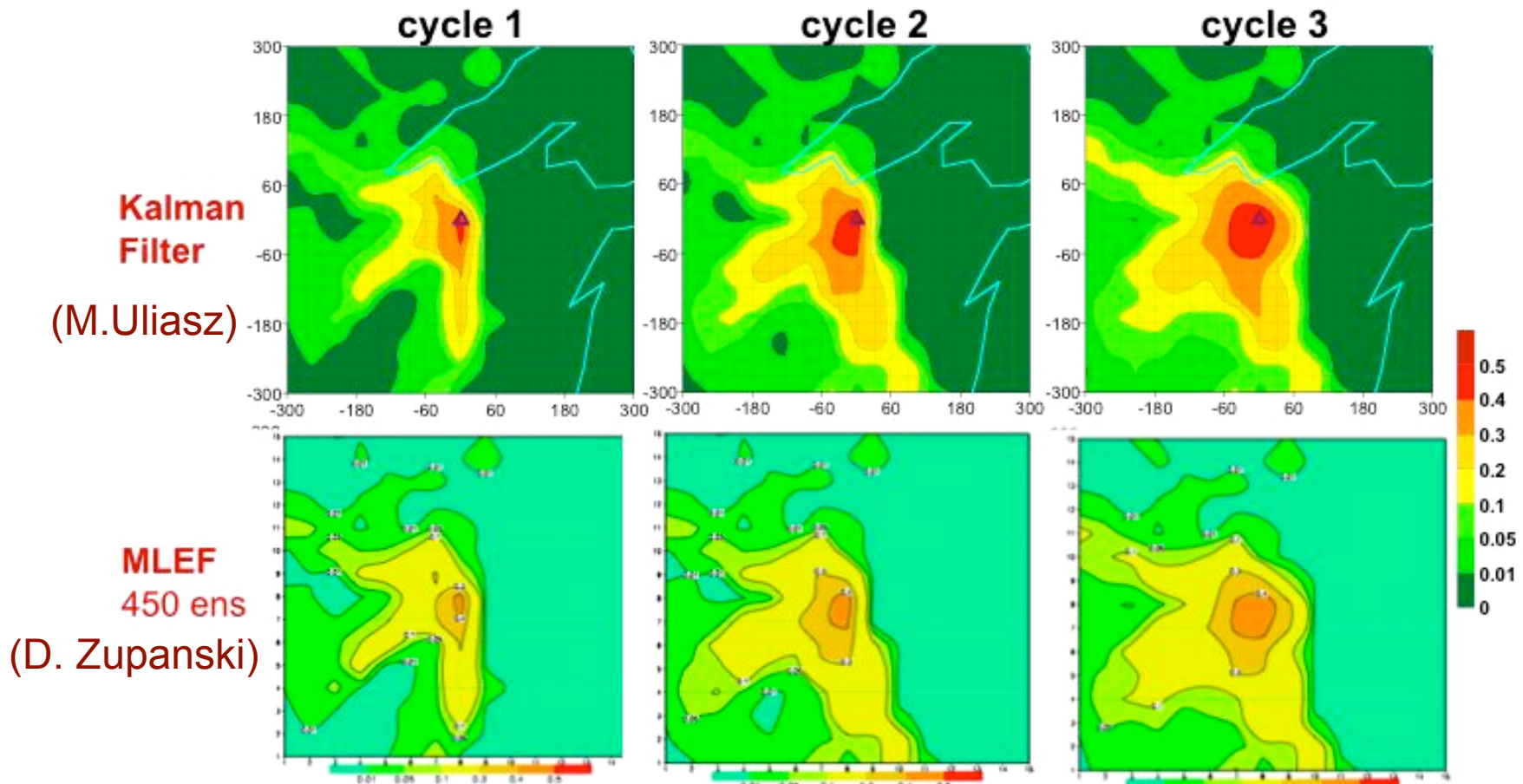


**Does this really happen?!?**

**KF,  
EnKF,  
4d-var,  
all  
created  
equal?**

# TEST RESULTS EMPLOYING A LINEAR MODEL AND GAUSSIAN PDFs

LPDM model: Estimation of respiration bias  
Reduction of uncertainty ( $\sigma_0 - \sigma$ ), Nstate=450, Nobs=600,  
three 5-day data assimilation cycles



This is a sanity check of the full-rank MLEF solution: it is equal to the Kalman filter solution for linear models (e.g., LPDM model).

## Kalman filter solution

### Analysis step:

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{P}_f \mathbf{H}^T (\mathbf{H} \mathbf{P}_f \mathbf{H}^T + \mathbf{R})^{-1} [\mathbf{y} - \mathbf{H}(\mathbf{x}_b)]$$

$\mathbf{x}_a$  - Optimal estimate of  $x$  (analysis)

$\mathbf{x}_b$  - Background (prior) estimate of  $x$

$$\mathbf{P}_a = [\mathbf{I} - \mathbf{P}_f \mathbf{H}^T (\mathbf{H} \mathbf{P}_f \mathbf{H}^T + \mathbf{R})^{-1} \mathbf{H}] \mathbf{P}_f = (\mathbf{I} - \mathbf{K} \mathbf{H}) \mathbf{P}_f$$

$\mathbf{P}_a$  - Analysis (posterior) error covariance matrix ( $N_{state} \times N_{state}$ )

$\mathbf{K}$  - Kalman gain matrix ( $N_{state} \times N_{obs}$ )

### Forecast step:

$$\mathbf{x}_0 = \mathbf{x}_a \quad ; \quad \mathbf{x}_n = \mathbf{M}_{n,n-1}(\mathbf{x}_{n-1}) + \mathbf{G}(\mathbf{x}_{n-1}) \mathbf{w}_{n-1}$$

Often neglected

$$\mathbf{P}_f = \mathbf{M} \mathbf{P}_a \mathbf{M}^T + \mathbf{G} \mathbf{Q} \mathbf{G}^T$$
 - Update of forecast error covariance

# Ensemble Kalman Filter (EnKF) solution

EnKF as first introduced by Evensen (1994) as a Monte Carlo filter.

## Analysis step:

Analysis solution defined for each ensemble member  $i$ :

$$\mathbf{x}_a^i = \mathbf{x}_b^i + \mathbf{P}_f^e \mathbf{H}^T (\mathbf{H} \mathbf{P}_f^e \mathbf{H}^T + \mathbf{R}^e)^{-1} (\mathbf{y}^i - H(\mathbf{x}_b^i))$$

Analysis ensemble perturbations:

$$\mathbf{b}_i^a = \mathbf{x}_a^i - \bar{\mathbf{x}}_a$$

Mean analysis solution:

$$\bar{\mathbf{x}}_a = \bar{\mathbf{x}}_b + \mathbf{P}_f^e \mathbf{H}^T (\mathbf{H} \mathbf{P}_f^e \mathbf{H}^T + \mathbf{R}^e)^{-1} (\bar{\mathbf{y}} - H(\bar{\mathbf{x}}_b))$$

Analysis error covariance in ensemble subspace:

$$\left( \mathbf{P}_a^e \right)^{1/2} = \begin{bmatrix} p_{1,1}^a & p_{1,2}^a & \cdot & p_{1,Nens}^a \\ p_{2,1}^a & p_{2,2}^a & \cdot & p_{2,Nens}^a \\ p_{3,1}^a & p_{3,2}^a & \cdot & p_{3,Nens}^a \\ \cdot & \cdot & \cdot & \cdot \\ p_{Nstate,1}^f & p_{Nstate,2}^f & \cdot & p_{Nstate,Nens}^f \end{bmatrix} = \begin{bmatrix} b_1^a & b_2^a & \cdot & b_{Nens}^a \end{bmatrix}$$

Sample analysis covariance

$$\mathbf{P}_a^e = \frac{1}{N_{ens} - 1} \left( \mathbf{P}_a^e \right)^{1/2} \left[ \left( \mathbf{P}_a^e \right)^{1/2} \right]^T$$

# Ensemble Kalman Filter (EnKF)

## Forecast step:

Ensemble forecasts employing a **non-linear** model  $M$

$$\mathbf{x}_n^j = M_{n,n-1}(\mathbf{x}_{n-1}^j) + G(\mathbf{x}_{n-1}^j)\mathbf{w}_{n-1}^j$$

Forecast error covariance calculated using ensemble perturbations:

$$\left(\mathbf{P}_f^e\right)^{1/2} = \begin{bmatrix} p_{1,1}^f & p_{1,2}^f & \cdot & p_{1,Nens}^f \\ p_{2,1}^f & p_{2,2}^f & \cdot & p_{2,Nens}^f \\ p_{3,1}^f & p_{3,2}^f & \cdot & p_{3,Nens}^f \\ \cdot & \cdot & \cdot & \cdot \\ p_{Nstate,1}^f & p_{Nstate,2}^f & \cdot & p_{Nstate,Nens}^f \end{bmatrix} = \left[ b_1^f \quad b_2^f \quad \cdot \quad b_{Nens}^f \right]$$

$$b_i^f = M(\mathbf{x}_a^i) - M(\bar{\mathbf{x}}_a)$$

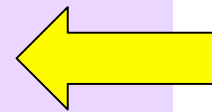
$$\mathbf{P}_f^e = \frac{1}{N_{ens} - 1} \left(\mathbf{P}_f^e\right)^{1/2} \left[ \left(\mathbf{P}_f^e\right)^{1/2} \right]^T$$

Non-linear forecast perturbations

Sample forecast covariance

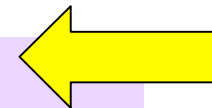
# There are many different versions of EnKF

- Monte Carlo EnKF (Evensen 1994; 2003)
- EnKF (Houtekamer et al. 1995; 2005; First operational version)
- Hybrid EnKF (Hamill and Snyder 2000)
- EAKF (Anderson 2001)
- ETKF (Bishop et al. 2001)
- EnSRF (Whitaker and Hamill 2002)
- LEKF (Ott et al. 2004)



Minimum variance solution

- MLEF (Zupanski 2005; Zupanski and Zupanski 2006)



Maximum likelihood solution

Why maximum likelihood solution? It is more adequate for employing non-Gaussian PDFs.

## Current status of EnKF applications

- EnKF is operational in Canada, since January 2005 (Houtekamer et al.). Results comparable to 4d-var.
- EnKF is better than 3d-var (experiments with NCEP T62 GFS) - Whitaker et al., THORPEX presentation ).
- Very encouraging results of EnKF in application to non-hydrostatic, cloud resolving models (Zhang et al., Xue et al.).
- Very encouraging results of EnKF for ocean (Evensen et al.), climate (Anderson et al.), and soil hydrology models (Reichle et al.).

Theoretical advantages of ensemble-based DA methods are getting confirmed in an increasing number of practical applications.

# Examples of MLEF applications



# Maximum Likelihood Ensemble Filter

(Zupanski 2005; Zupanski and Zupanski 2006)

$\mathbf{x}_n = M_{n,n-1}(\mathbf{x}_{n-1}, \mathbf{b}_{n-1}, \boldsymbol{\gamma}_{n-1})$  - Dynamical model for standard model state  $x$

$\mathbf{b}_n = G_{n,n-1}(\mathbf{b}_{n-1})$  - Dynamical model for model error (bias)  $b$

$\boldsymbol{\gamma}_n = S_{n,n-1}(\boldsymbol{\gamma}_{n-1})$  - Dynamical model for empirical parameters  $\gamma$

Define augmented state vector  $\mathbf{z}$

$$\mathbf{z}_n = (\mathbf{x}_{n-1}, \mathbf{b}_{n-1}, \boldsymbol{\gamma}_{n-1})$$

And augmented dynamical model  $F$

$$\mathbf{z}_n = F_{n,n-1}(\mathbf{z}_{n-1})$$



Find optimal solution (augmented analysis)  $\mathbf{z}_a$  by minimizing  $J$  (MLEF method):

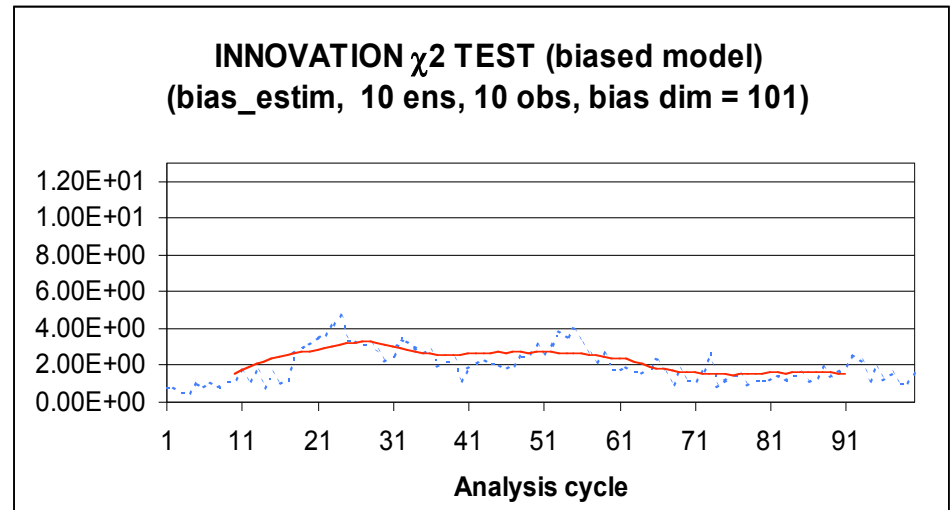
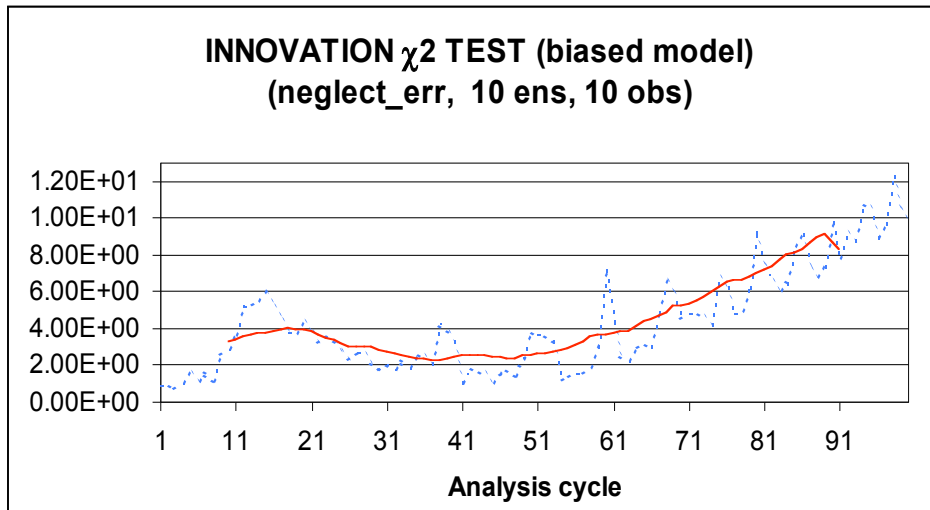
$$J = \frac{1}{2}[\mathbf{z} - \mathbf{z}_b]^T \mathbf{P}_f^{-1}[\mathbf{z} - \mathbf{z}_b] + \frac{1}{2}[H[F(\mathbf{z})] - \mathbf{y}_{obs}]^T \mathbf{R}^{-1}[H[F(\mathbf{z})] - \mathbf{y}_{obs}] = \min$$

# BIAS ESTIMATION, KdVB model

(Zupanski and Zupanski 2006)

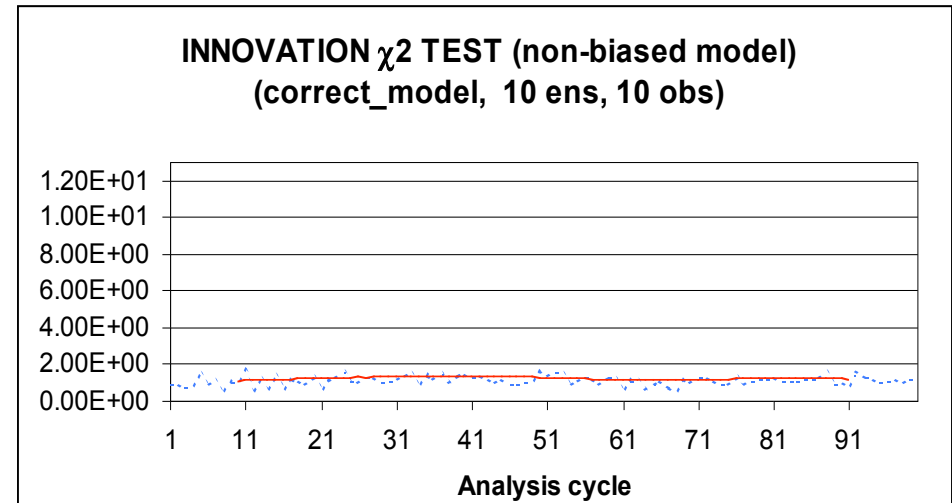
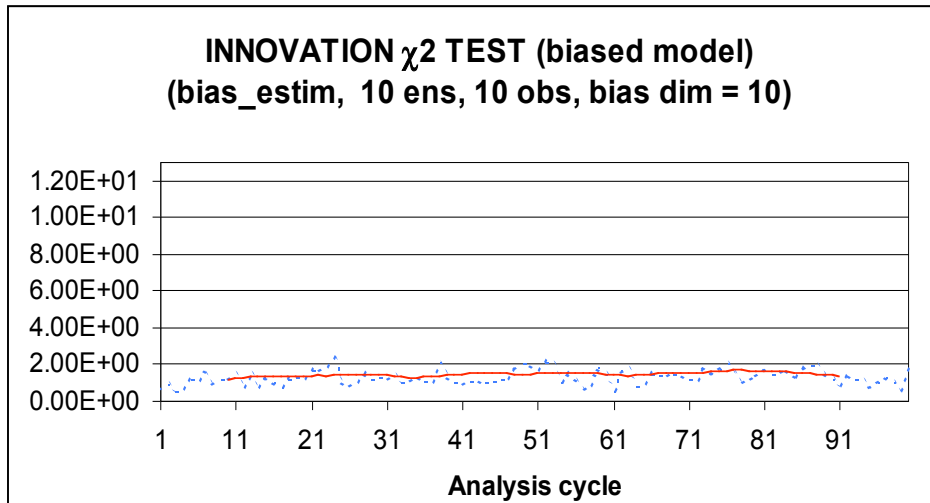
**NEGLECT BIAS**

**BIAS ESTIMATION (vector size=101)**



**BIAS ESTIMATION (vector size=10)**

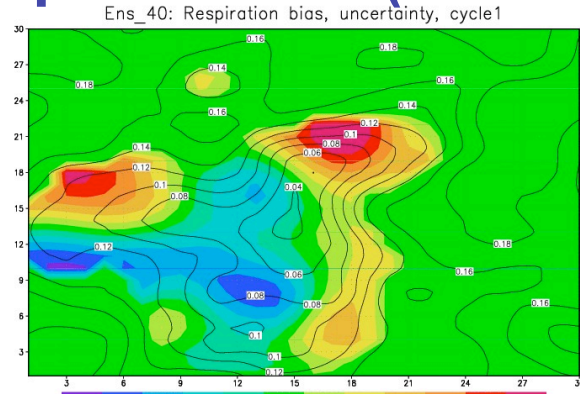
**NON-BIASED MODEL**



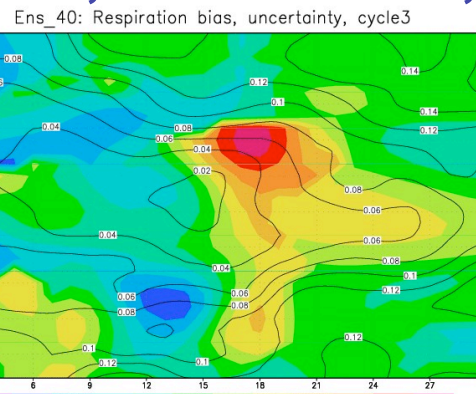
It is beneficial to reduce degrees of freedom of the model error.

# Bias estimation: Respiration bias $\beta_R$ , using LPDM carbon transport model (Nstate=1800, Nobs=1200, DA interv=10 days)

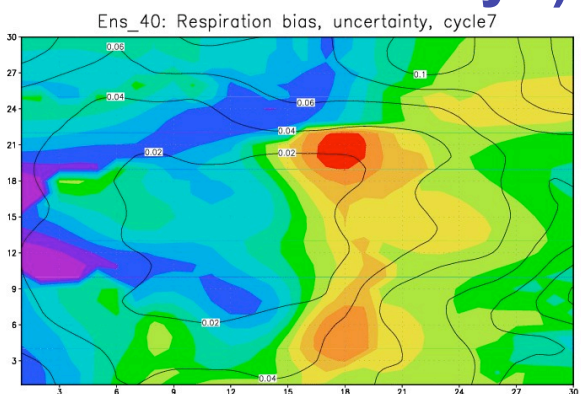
40  
Ens



Cycle 1

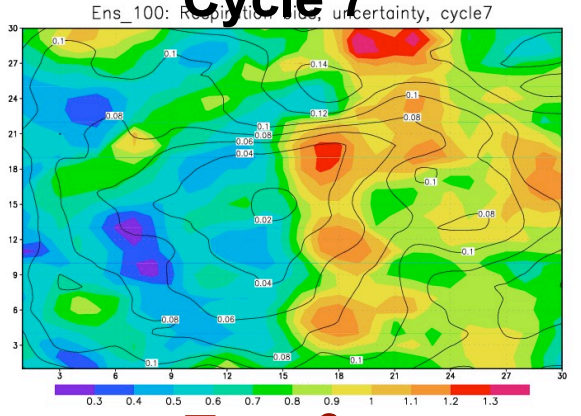
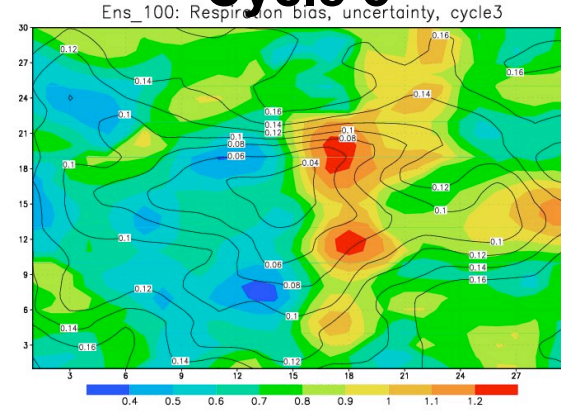
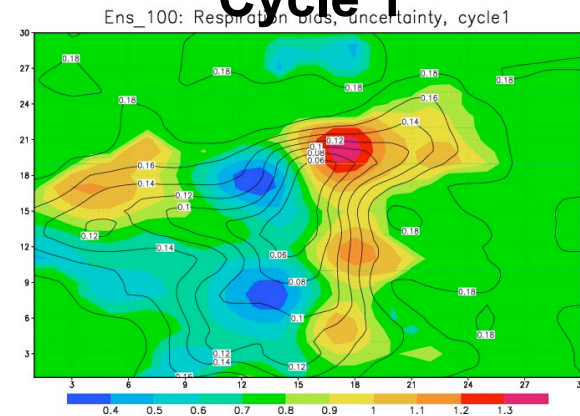


Cycle 3



Cycle 7

100  
Ens

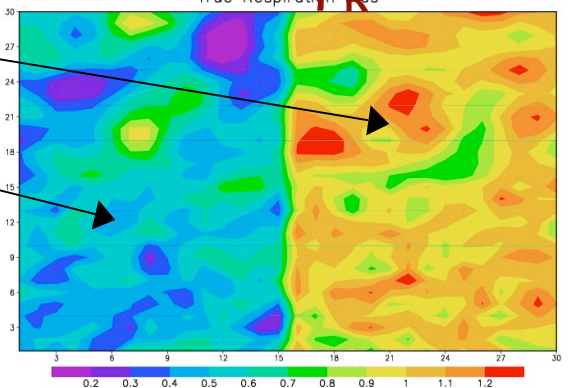


Domain with larger bias (typically land)

Domain with smaller bias (typically ocean)

Both the magnitude and the spatial patterns of the true bias are successfully captured by the MLEF.

True  $\beta_R$



## Information measures in ensemble subspace

(Bishop et al. 2001; Wei et al. 2005; Zupanski et al. 2005, 2006)

$\mathbf{C} = \mathbf{Z}^T \mathbf{Z}$      $\mathbf{C}$  - information matrix in ensemble subspace of dim  $N_{ens} \times N_{ens}$

$\mathbf{z}^i = R^{-1/2} H[M(x + p_f^i)] - R^{-1/2} H[M(x)]$      $\mathbf{z}^i$  - are columns of  $\mathbf{Z}$

$\mathbf{x} - \mathbf{x}_b = \mathbf{P}_f^{1/2} (\mathbf{I} + \mathbf{C})^{-1/2} \boldsymbol{\zeta}$      $\boldsymbol{\zeta}$  - control vector in ensemble space of dim  $N_{ens}$   
 $\mathbf{x}$  - model state vector of dim  $N_{state} \gg N_{ens}$

**Degrees of freedom (DOF) for signal (Rodgers 2000):**

$d_s = \text{tr} [(\mathbf{I} + \mathbf{C})^{-1} \mathbf{C}] = \sum_i \frac{\lambda_i^2}{(1 + \lambda_i^2)}$      $\lambda_i^2$  - eigenvalues of  $\mathbf{C}$

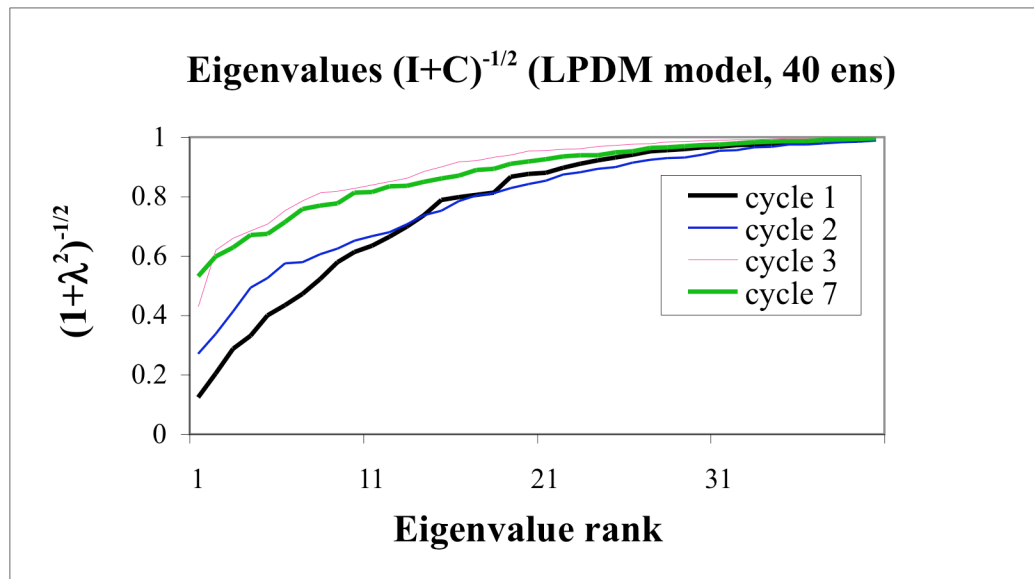
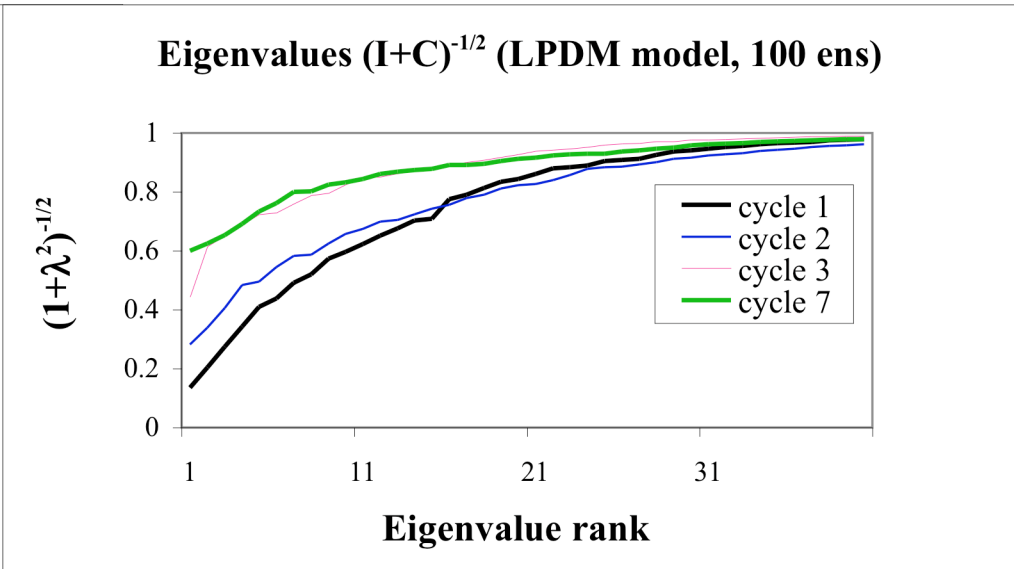
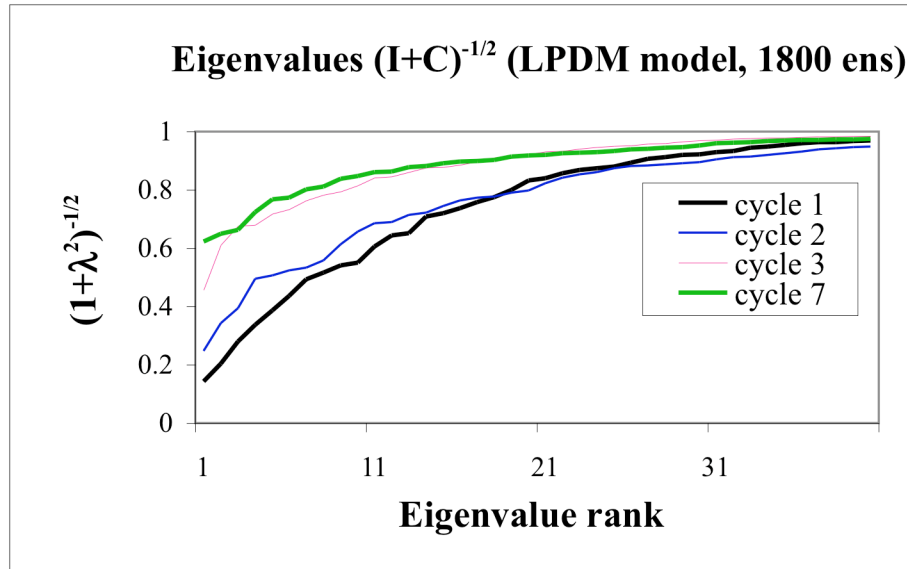
**Shannon information content,  
or entropy reduction**

$$h = \frac{1}{2} \sum_i \ln(1 + \lambda_i^2)$$

Errors are assumed Gaussian in these measures.



# LPDM Model CO<sub>2</sub>-flux BIAS estimation: Eigenvalue spectrum of $(I+C)^{-1/2}$ (First 40 eigenvalues, N<sub>ens</sub> = 1800, 100, and 40)

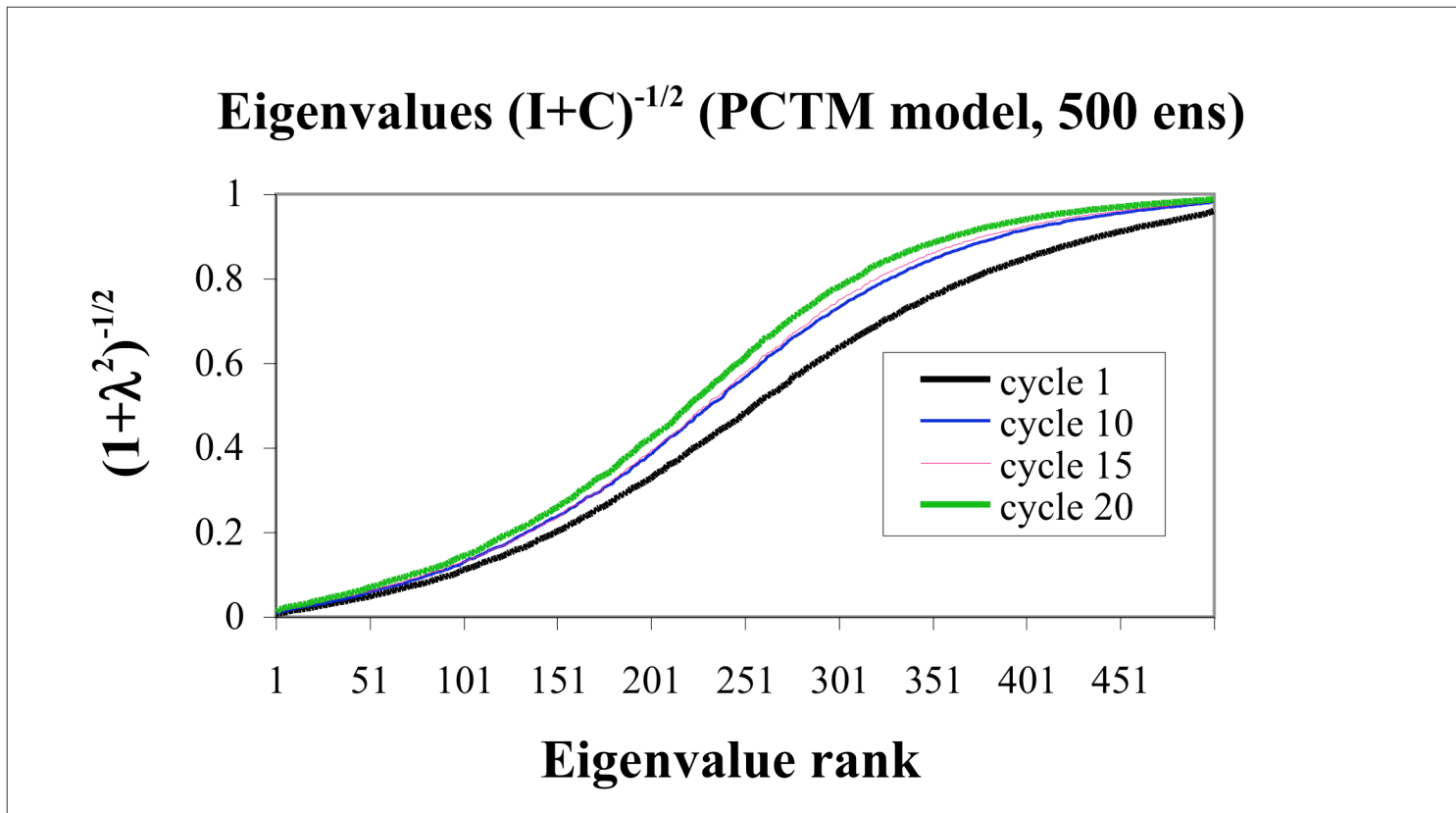


**Eigenvalue spectrum is very similar for all 3 ensemble sizes!**



## PCTM Global Model CO<sub>2</sub>-flux estimation: Eigenvalue spectrum of C

(Nstate=13104, Nobs=13104, fully observed system, Nens= 500)

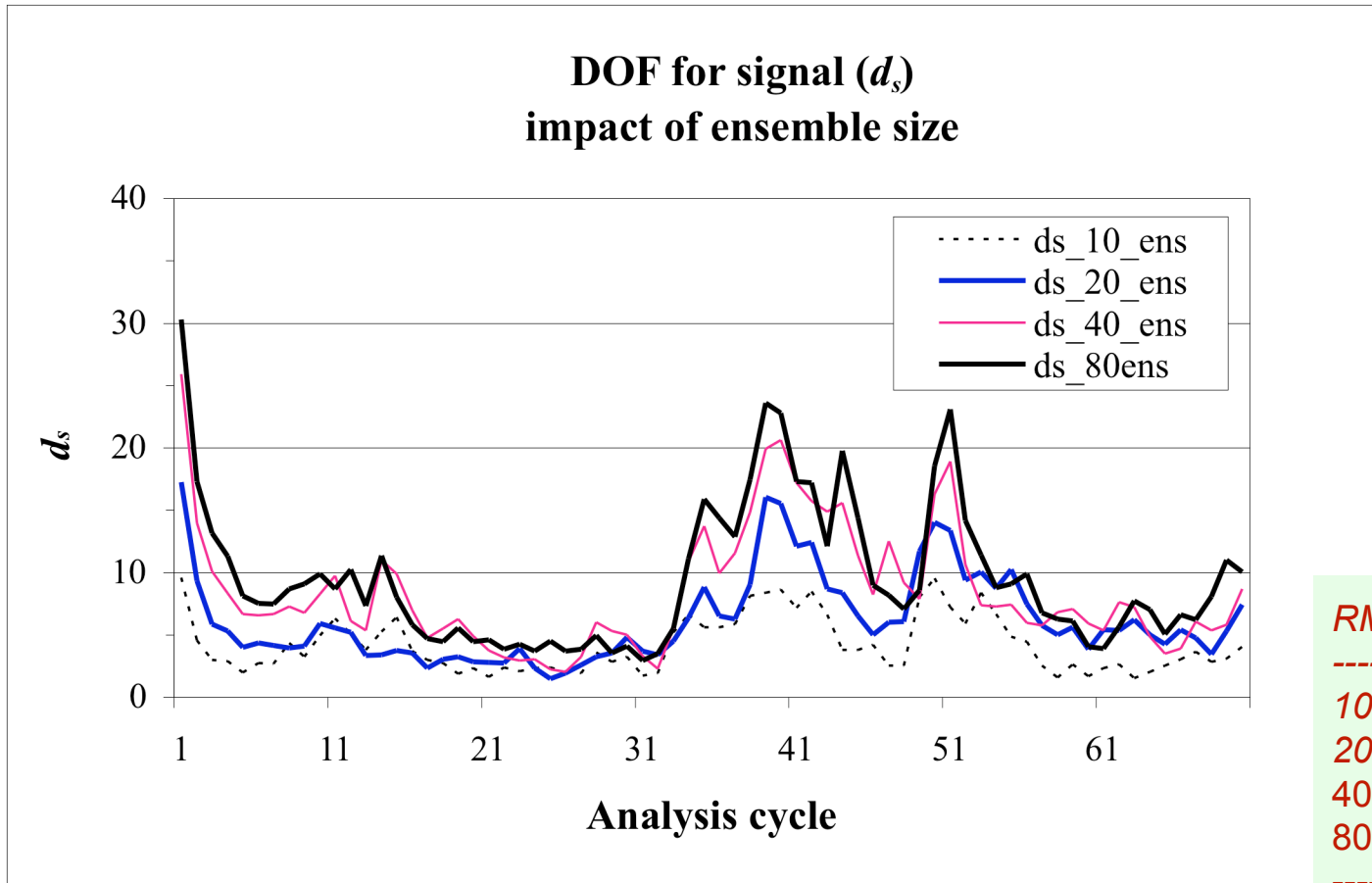


Ensemble size of 500 is adequate for describing all DOFs of this fully observed system.

In later cycles more eigenvalues are approaching value 1 (no information).



## GEOS-5 Single Column Model: DOF for signal (Nstate=80; Nobs=80, seventy 6-h DA cycles, assimilation of simulated T,q observations)



DOF for signal varies from one analysis cycle to another due to changes in atmospheric conditions.

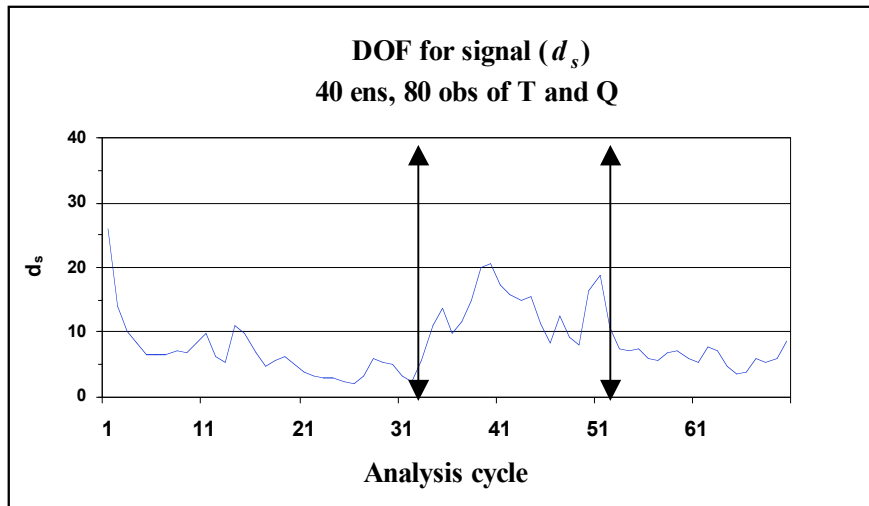
*RMS Analysis errors for T, q:*

-----  
 10ens ~ 0.50K; 0.566g/kg  
 20ens ~ 0.32K; 0.462g/kg  
 40ens ~ 0.27K; 0.417g/kg  
 80ens ~ 0.20K; 0.362g/kg  
 -----

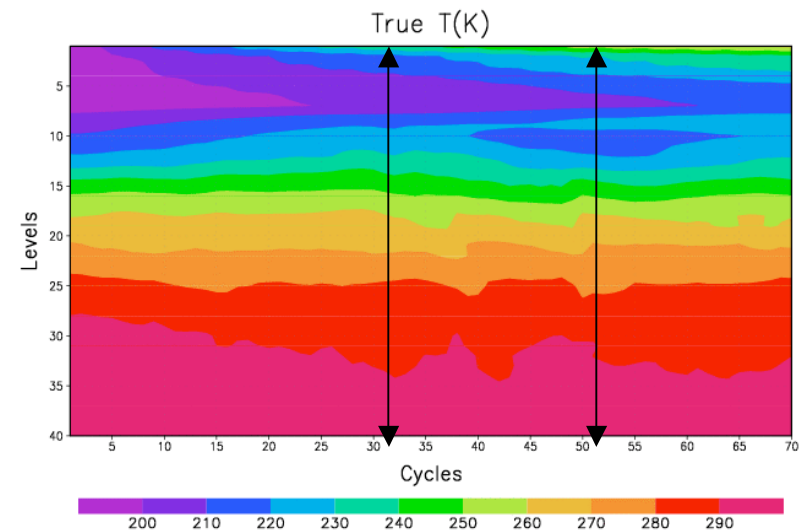
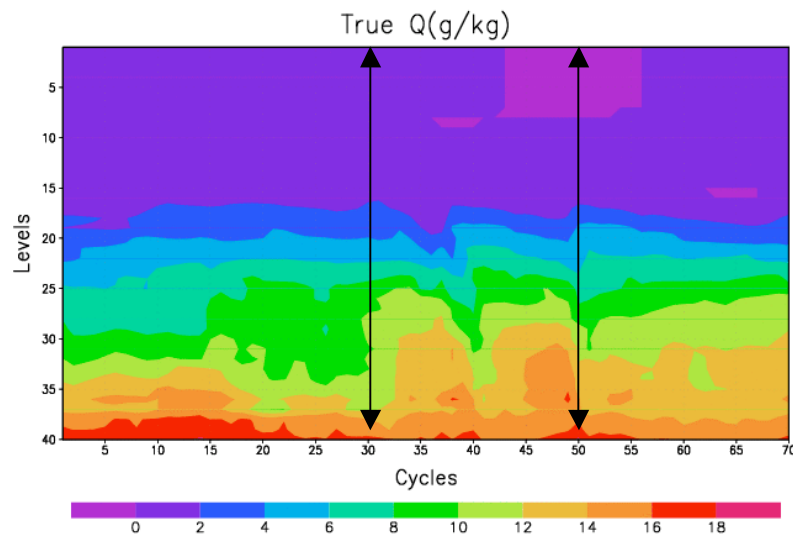
No\_obs ~ 0.82K; 0.656g/kg

**Small ensemble size (10 ens), even though not perfect, captures main data signals.**

# MLEF application to GEOS-5 single column model: DOF for signal (ds)



$$d_s = \text{tr}[(\mathbf{I} + \mathbf{C})^{-1} \mathbf{C}] = \sum_i \frac{\lambda_i^2}{(1 + \lambda_i^2)}$$



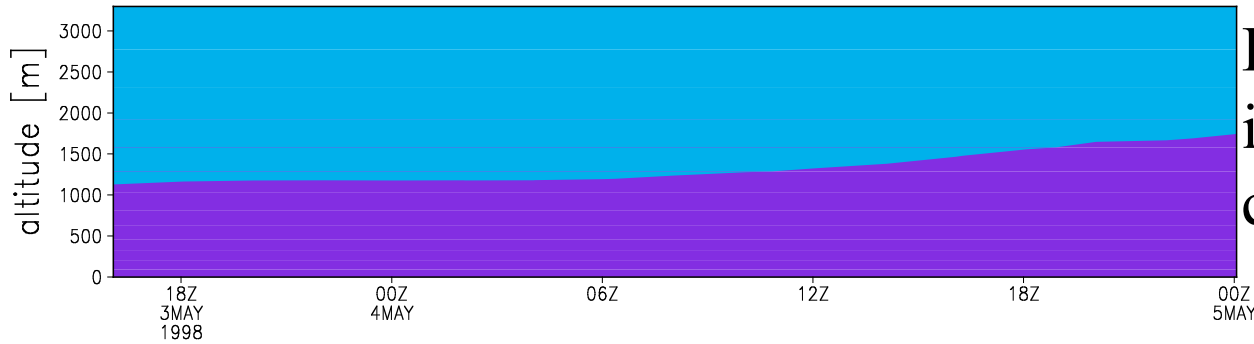
DOF for signal reflects dynamics of the true state!

Dusanka Zupanski, CIRA/CSU  
Zupanski@CIRA.colostate.edu

# MLEF assimilation of radar/lidar IWP and LWP using RAMS/LES

IFN concentration, Control run

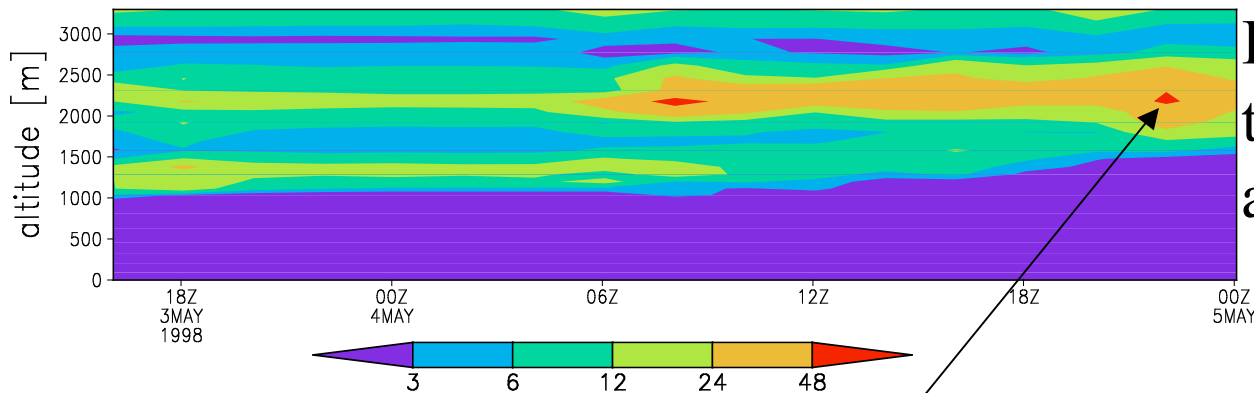
CONTROL



IFN ↓ below inversion as cloud forms

IFN concentration, 48 members

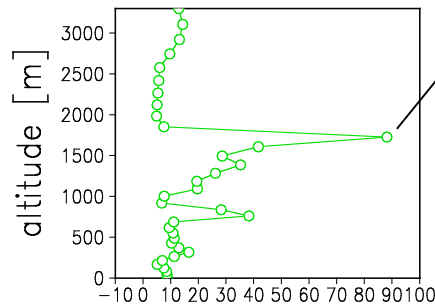
EXP



IFN ↑ above the inversion, as observed

Observed IFN concentration (at ~ 22Z 4May)

VERIF



Independent observation

Courtesy of G. Carrió

Dusanka Zupanski, CIRA/CSU  
Zupanski@CIRA.colostate.edu

# Non-Gaussian (lognormal) MLEF framework: CSU SWM (Randall et al.)

*Cost function derived from posterior PDF*

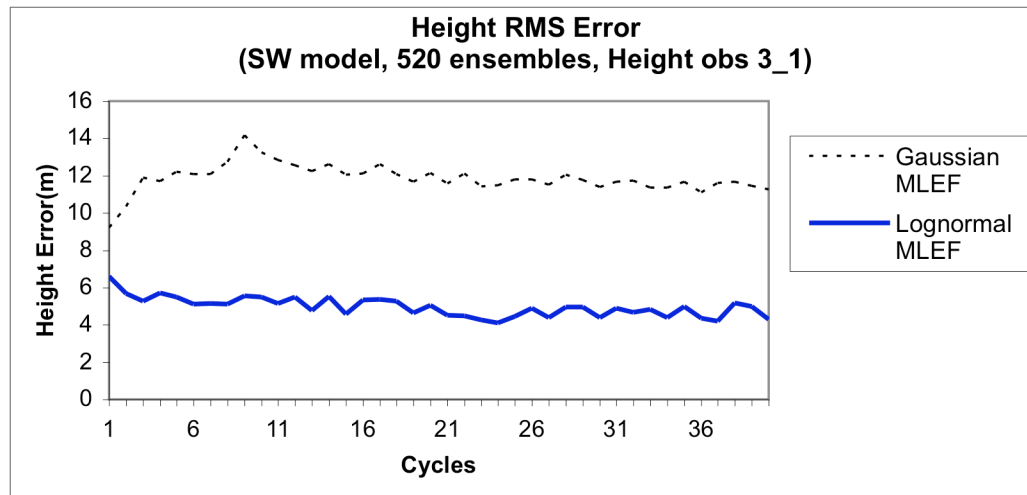
*( $\mathbf{x}$ -Gaussian,  $\mathbf{y}$ -lognormal):*

$$J(\mathbf{x}) = \frac{1}{2} (\mathbf{x} - \mathbf{x}^f)^T \mathbf{P}_f^{-1} (\mathbf{x} - \mathbf{x}^f) + \frac{1}{2} \left( \ln \left[ \frac{\mathbf{y}}{\mathcal{H}(\mathbf{x})} \right] - m \right)^T \mathbf{R}_S^{-1} \left( \ln \left[ \frac{\mathbf{y}}{\mathcal{H}(\mathbf{x})} \right] - m \right) + \sum_{i=1}^{N_{obs}} \ln \left[ \frac{\mathbf{y}}{\mathcal{H}(\mathbf{x})} \right]_i$$

*Normal  
(Gaussian)*

$$\mathcal{H}(\mathbf{x}) = \exp \left( a \frac{\mathbf{x} - \mathbf{x}_{ref}}{b} \right)$$

*Lognormal  
additional nonlinear term*



**Beneficial impact of correct PDF assumption – practical advantages**

*Courtesy of M. Zupanski*

Dusanka Zupanski, CIRA/CSU  
Zupanski@CIRA.colostate.edu



## Future Research Directions

- Covariance inflation and localization need further investigations: Are these techniques necessary?
- Model error and parameter estimation need further attention: Do we have sufficient information in the observations to estimate complex model errors?
- Information content analysis might shed some light on DOF of model error and also on the necessary ensemble size.
- Non-Gaussian PDFs have to be included into DA (especially for cloud variables).
- Characterize error covariances for cloud variables.
- Account for representativeness error.

## References for further reading

- Anderson, J. L., 2001: An ensemble adjustment filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903.
- Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**, (C5), 10143–10162.
- Evensen, G., 2003: The ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean Dynamics*. **53**, 343–367.
- Hamill, T. M., and C. Snyder, 2000: A hybrid ensemble Kalman filter/3D-variational analysis scheme. *Mon. Wea. Rev.*, **128**, 2905–2919.
- Houtekamer, Peter L., Herschel L. Mitchell, 1998: Data Assimilation Using an Ensemble Kalman Filter Technique. *Monthly Weather Review*: Vol. 126, No. 3, pp. 796–811.
- Houtekamer, Peter L., Herschel L. Mitchell, Gerard Pellerin, Mark Buehner, Martin Charron, Lubos Spacek, and Bjarne Hansen, 2005: Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations. *Monthly Weather Review*: Vol. 133, No. 3, pp. 604–620.
- Ott, E., and Coauthors, 2004: A local ensemble Kalman filter for atmospheric data assimilation. *Tellus.*, **56A**, 415–428.
- Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, 2003: Ensemble square root filters. *Mon. Wea. Rev.*, **131**, 1485–1490.
- Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913–1924.
- Zupanski D. and M. Zupanski, 2006: Model error estimation employing an ensemble data assimilation approach. *Mon. Wea. Rev.* (in press).
- Zupanski, M., 2005: Maximum likelihood ensemble filter: Theoretical aspects. *Mon. Wea. Rev.*, **133**, 1710–1726

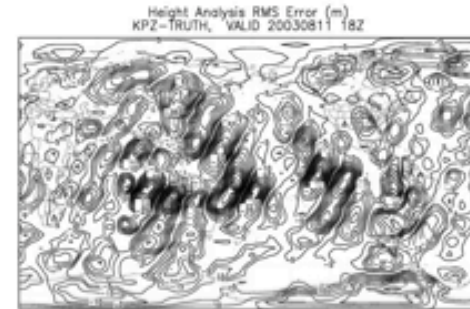
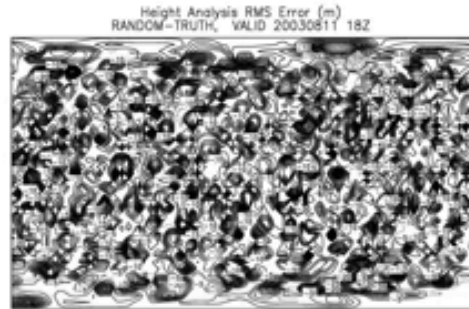
Thank you.

Dusanka Zupanski, CIRA/CSU  
Zupanski@CIRA.colostate.edu

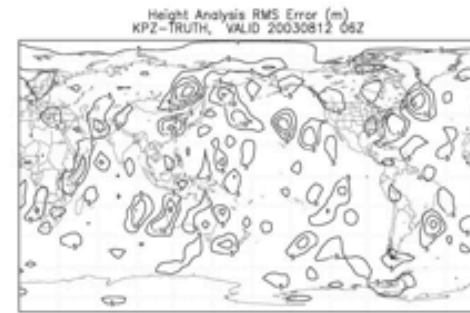
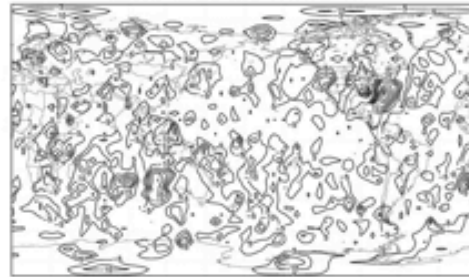


## Example: CSU SWM model of Randall et al. Solution-Truth shown

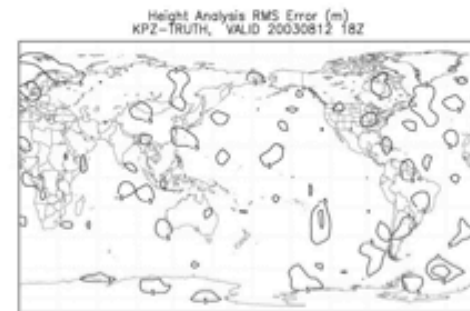
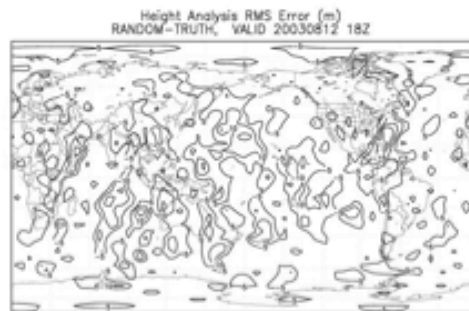
Cycle 1



Cycle 3



Cycle 5



**Exp1:  
Random  
noise in  
cycle 1**

**Exp2:  
Correlated  
random  
noise in  
cycle 1**

*(From M.  
Zupanski et al.,  
Tellus)*

**Time evolving model dynamics significantly reduces the noise in the initially prescribed perturbations!**

## MLEF experiments with CSU/RAMS Large Eddy Simulation (LES) model

- 23 2-h DA cycles: 18UTC 2 May 1998 – 00 UTC 5 May 1998 (Mixed phase Arctic boundary layer cloud at Sheba site)
- Experiments initialized with typical *clean aerosol* concentrations
- May 4 was abnormal: *high* IFN and CCN above the inversion
- $\Delta x = 50\text{m}$ ,  $\Delta z_{\text{max}} = 30\text{m}$  (2d domain: 50col, 40lev),  $\Delta t = 2\text{s}$ , Nens=48
- Sophisticated microphysics in RAMS/LES
- Control variables:  $\Theta_{\text{il}}$ ,  $u$ ,  $v$ ,  $w$ ,  $N_x$ ,  $R_x$  (8 species), IFN, CCN (dim= 22 variables x 50 columns x 40 levels = 44000)
- Radar/lidar real observations of IWP, LWP are assimilated
- IWP and LWP are vertically integrated quantities (no information about the profiles of IFN, CCN is observed)

### Acronyms:

IFN - Ice Forming Nuclei

CCN - Cloud Concentration Nuclei

IWP (LWP) - Ice (Liquid) Water Path



# Basic characteristics of Maximum Likelihood Ensemble Filter (MLEF)

*(Zupanski 2005; Zupanski and Zupanski 2006)*

- MLEF is similar to 4dvar because it seeks a **maximum likelihood** solution (i.e., minimum of  $J$ ).
- It is also similar to EnKF methods because it uses ensembles to calculate forecast error covariance.
- MLEF uses the same definition of transformation matrix as in the ETKF (Bishop et al. 2001).
- It has a capability to estimate and reduce several major sources of forecast uncertainties simultaneously: **Initial conditions, model error, boundary conditions, and empirical parameters.**
- MLEF has also a capability to take into account **non-Gaussian (log-normal) PDFs** (Fletcher and M. Zupanski 2006)