International Conference on Ensemble Methods in Geophysical Sciences
November 12-18, 2012, Toulouse, France

Towards the development of hybrid variational-ensemble data assimilation: Minimization, Hessian preconditioning, and static error covariance model

Milija Zupanski

Cooperative Institute for Research in the Atmosphere
Colorado State University
Fort Collins, Colorado, U. S. A.

Acknowledgements:

- Prof. Michael Navon, Florida State Univ.
- National Science Foundation (NSF) - Collaboration in Mathematical Geosciences (CMG)
- Joint Centers for Satellite Data Assimilation
- NCAR CISL high-performance computing support (“Bluefire”)

Outline

- Hessian preconditioning
- Hybrid variational-ensemble methods
- Static error covariance in hybrid methods
- Future development

Why hybrid variational-ensemble DA?

1. Variational DA:

Full rank forecast error covariance

Nonlinear minimization

2. Ensemble DA:

Flow-dependent forecast error covariance

Uncertainty feedback between analysis and forecast

3. Unresolved issue:

Optimal Hessian preconditioning

4. Computational issues (user and application dependent)

Code development and updates, computing time

Take advantage of both DA methodologies

Minimization and Hessian preconditioning

Minimize cost function

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

Using change of variable $\quad x-x^{f}=P_{f}^{1/2}w$

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

Gradient

$$g=\nabla J(w)$$

Hessian

$$F=\nabla^{2}J(w)=I+P_{f}^{T/2}H^{T}R^{-1}HP_{f}^{1/2}=I+C$$

Descent direction in each minimization iteration is the solution of linear equation:

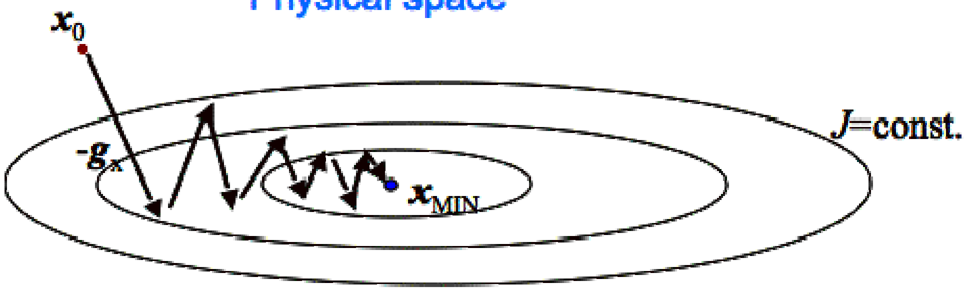
$$Fd_{k}=-g_{k} \qquad or \qquad F(w_{k}-w_{k-1})=g_{k}-g_{k-1}$$

Optimal Hessian preconditioning can:
(1) Improve minimization convergence
(2) Reduce error of the analysis solution

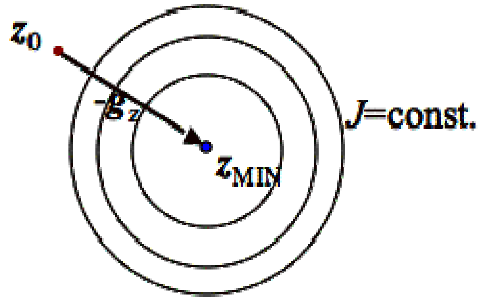
Improving convergence with Hessian preconditioning

Geometric interpretation of Hessian preconditioning:

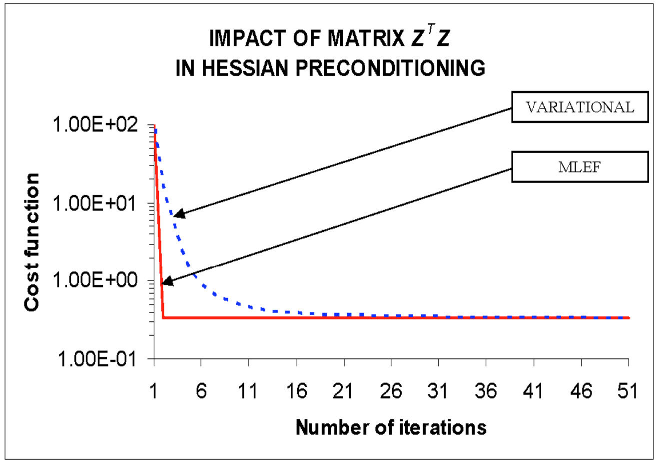
Physical space



Preconditioning space



Hessian Preconditioning



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

$$\mathrm{P}_{ENS}^{-1} = \boldsymbol{P}_f^{\,1/2}(\boldsymbol{I} + \boldsymbol{Z}^T \boldsymbol{Z})^{-1} \boldsymbol{P}_f^{\,T/2}$$

Impact of Hessian preconditioning on the accuracy of minimization solution

In each iteration search for the solution of linear equation:

$$Fd = -g \qquad \Rightarrow \qquad d = F^{-1} g$$

Matrix C is often neglected $\longrightarrow$ $d_1 = -g$ $\qquad F = I$

but the correct solution is $\longrightarrow$ $d_{opt} = -\left[I + C\right]^{-1} g$ $\qquad F = I + C$

What is the error of solution if C is neglected?

$$\left\|d_1 - d_{opt}\right\| = \left\|g - [I+C]^{-1} g\right\| = \left\|[I+C][I+C]^{-1} g - [I+C]^{-1} g\right\| \le \|C\| \left\|[I+C]^{-1} g\right\|$$

$$\boxed{\frac{\left\|d_1 - d_{opt}\right\|}{\left\|d_{opt}\right\|} \le \|C\| \qquad \|C\|_2 \sim O(10^2)}$$

$$C = P_f^{T/2} H^T R^{-1} H P_f^{1/2}$$

Solution error can be significant if sub-optimal Hessian preconditioning is used

Components of hybrid variational-ensemble DA

(1) Flow-dependent error covariance (improve variational DA)

- Ensemble DA produces flow-dependent forecast error covariance, while variational DA has static forecast error covariance (i.e. ***B*** matrix)

(2) Full-rank error covariance (improve ensemble DA)

- Ensemble DA has insufficient number of DOF, while variational DA has all DOFs

(3) Nonlinear minimization (improve ensemble DA)

- Ensemble DA solve linear KF equation, while variational DA has nonlinear minimization

(4) Optimal Hessian preconditioning (improve variational DA)

- Ensemble DA has an implicit optimal preconditioning (e.g., Kalman gain), while variational DA can employ only an approximate preconditioning

(5) Uncertainty feedback between analysis and forecast (improve variational DA)

- forecast uncertainty impacts the analysis uncertainty, and vice versa

Several or **all** these components can be incorporated in a hybrid DA system

Hybridization considerations

- Combining two (or more) DA algorithms can be advantageous
- Need to understand well each of the algorithms incorporated
- Important to include as many as possible requirements
- Straightforward hybridization may be simpler, but is not optimal
- Selective hybridization may be more complex, but has a potential to be optimal: selectively include only the desired components
- There is no hybrid DA method that includes all five requirements

Can we include *all five* components in a hybrid DA system?

Consider Maximum Likelihood Ensemble Filter (MLEF)

(Zupanski 2005, MWR; Zupanski et al. 2008, QJRMS)

Analysis:

Standard KF:

- Analysis is equivalent to minimizing a quadratic cost function (posterior pdf)
- Uncertainty is given by the inverse Hessian

MLEF: Generalize KF to include *nonlinear observation* operators:

- Minimize arbitrary nonlinear cost function
- Use inverse Hessian at the minimum as uncertainty estimate

Forecast:

Standard KF:

- Initial guess is the forecast from previous analysis $x^f = \mathbf{M} x^a$
- Forecast uncertainty is an evolution of analysis uncertainty by a linear model

$$P_f^{1/2} = MP_a^{1/2} \quad \Rightarrow \quad \begin{bmatrix} p_1^f & \cdots & p_n^f \end{bmatrix} = \begin{bmatrix} \mathbf{M} p_1^a & \cdots & \mathbf{M} p_n^a \end{bmatrix}$$

MLEF: Generalize KF by evolving the state and analysis uncertainty by a *nonlinear* model

- Initial guess is the forecast from previous analysis $x^f = \mathcal{M}(x^a)$
- Forecast uncertainty is an evolution of analysis uncertainty by a nonlinear model

$$\begin{bmatrix} p_1^f & \cdots & p_n^f \end{bmatrix} = \begin{bmatrix} \mathcal{M}(x^a + p_1^a) - \mathcal{M}(x^a) & \cdots & \mathcal{M}(x^a + p_n^a) - \mathcal{M}(x^a) \end{bmatrix}$$

Full rank static error covariance and optimal Hessian preconditioning

Q: What needs to be done in order to maintain optimal Hessian preconditioning not only for the ensemble, but also for the static covariance component?

A: Square root forecast error covariance should have a small number of columns in order to allow the computation of

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

Consequence:

- ❑ **Not feasible to have a full rank static error covariance and optimal Hessian preconditioning due to prohibitive cost of computing perturbed observation operator and SVD**

Alternative:

- ❑ **Is it possible to define a *sufficient*-rank static error covariance instead, "sufficient" defined as an acceptable approximation to the full-rank static error covariance?**

Sufficient-rank static error covariance

Assume that a full rank static error covariance square root is defined

$$P^{1/2}$$

1. **Construct an orthonormal reduced rank matrix** Q**, and**
2. **Define a sufficient-rank static covariance** P_{SR} **as**

$$P_{SR}^{1/2} = P^{1/2} Q$$

Define "sufficient" using a measure of distance μ **between full rank and reduced rank matrices**

$$\mu = \frac{\left\| P - P_{SR} \right\|}{\left\| P \right\|} = \frac{\left\| P^{1/2} P^{T/2} - P^{1/2} Q Q^T P^{T/2} \right\|}{\left\| P \right\|} \le \frac{\left\| P \right\| \left\| I - QQ^T \right\|}{\left\| P \right\|} = \left\| I - QQ^T \right\|$$

- μ **is between 0 and 1: smaller** μ **implies better approximation**
- **Distance between full-rank and sufficient rank covariance is bounded by the norm of orthogonal projection onto** $ker(QQ^T)$

Orthonormal matrix Q

1. **Define local covariance as sub-matrix of $P^{1/2}$ over local domain defined by typical decorrelation length**

2. Compute SVD of local matrix $P_L^{1/2}=U\Sigma U^T$ and truncate to rank *M*

$$\varepsilon = \frac{\left\| P_L^{1/2} - P_{LM}^{1/2} \right\|_2}{\left\| P_L^{1/2} \right\|_2} = \frac{\sigma_{M+1}}{\sigma_1} << 1 \qquad \sigma_1 \geq \cdots \geq \sigma_M \geq \cdots \geq \sigma_L > 0$$

3. Build *Q* **as a block-circulant matrix of local singular vectors** $\{u_i\}$

$$Q = \begin{pmatrix} u_1 & u_2 & \cdots & u_M \\ u_M & u_1 & \cdots & u_{M-1} \\ \vdots & \vdots & \ddots & \vdots \\ u_2 & u_3 & \cdots & u_1 \end{pmatrix} \qquad Q^T Q = I$$

For $\mu << 1$ **and** $\varepsilon << 1$ one can build an acceptable sufficient-rank $P_{SR}^{1/2} = P^{1/2} Q$

Preliminary results

1. **Global domain** 147x147x33 (~700,000) **grid points**
2. **Local domain** 21x21x33 (decorrelation length10x10x7)
3. **Global covariance defined as a banded Toeplitz matrix**
4. **Compute** Q **using** M=49 **singular vectors**

6. Evaluate covariance structure as a response to single observation at:

a- central point of global (and local) domains

b- corner point of global domain

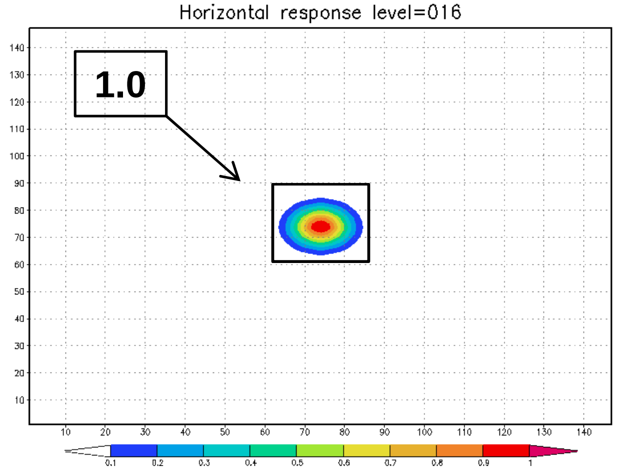
c- corner point of local domain

$$P = \begin{pmatrix} p_0 & \cdots & p_k & 0 & \cdots & 0 \\ \vdots & \ddots & & \ddots & \ddots & \vdots \\ p_k & & p_0 & & \ddots & 0 \\ 0 & \ddots & & \ddots & & p_k \\ \vdots & \ddots & \ddots & & \ddots & \vdots \\ 0 & \cdots & 0 & p_k & \cdots & p_0 \end{pmatrix}$$

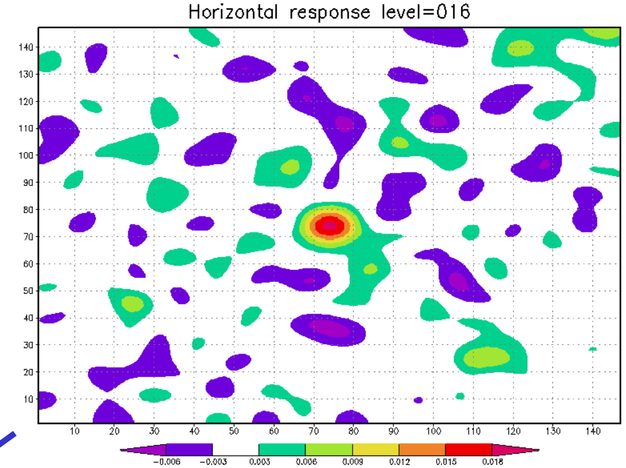
$$P \begin{pmatrix} 0 \\ \vdots \\ 1_i \\ \vdots \\ 0 \end{pmatrix} \qquad P^{1/2} Q Q^T P^{T/2} \begin{pmatrix} 0 \\ \vdots \\ 1_i \\ \vdots \\ 0 \end{pmatrix}$$

Processing sufficient rank matrix: central point

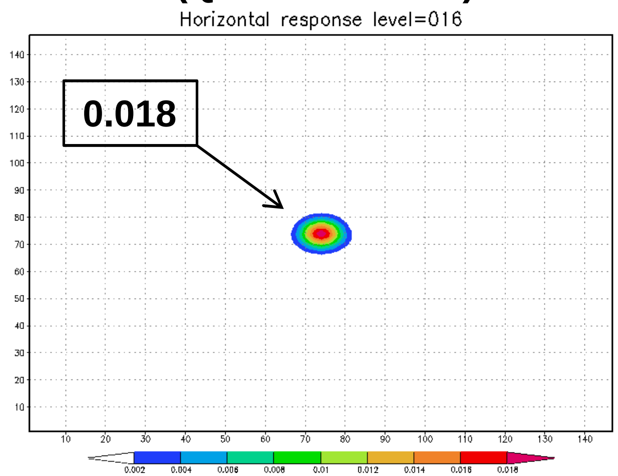
Horizontal response (truth)



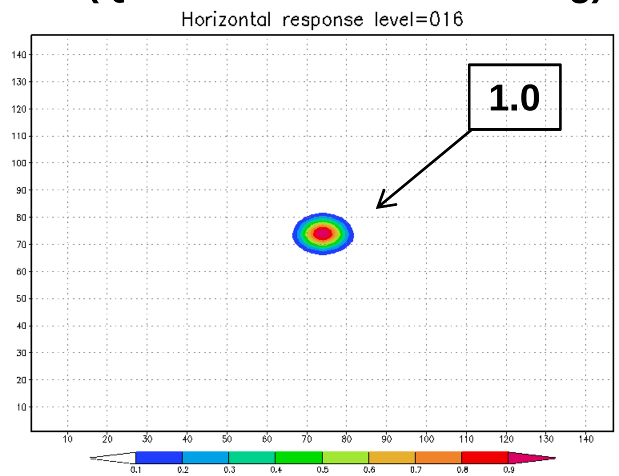
Horizontal response (Q)



Horizontal response (Q + localization)



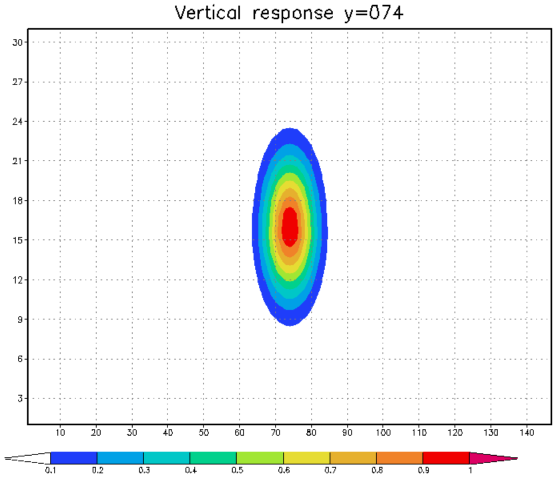
Horizontal response (Q + localization + re-scaling)



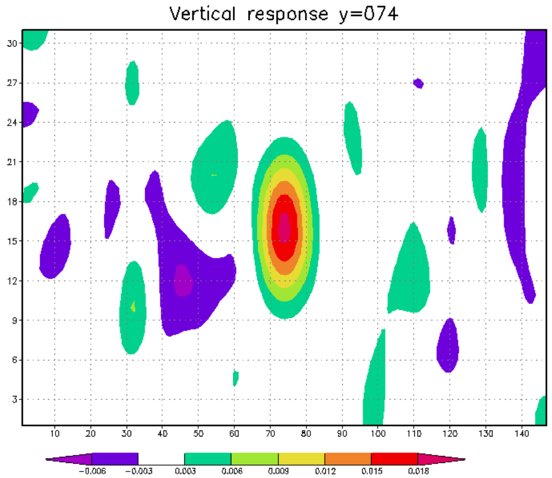
Sufficient rank covariance becomes acceptable after processing

Processing reduced rank matrix: central point

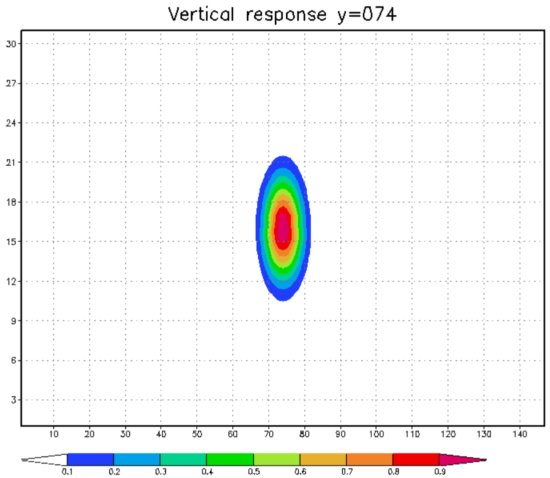
Vertical response (truth)



Vertical response (Q)



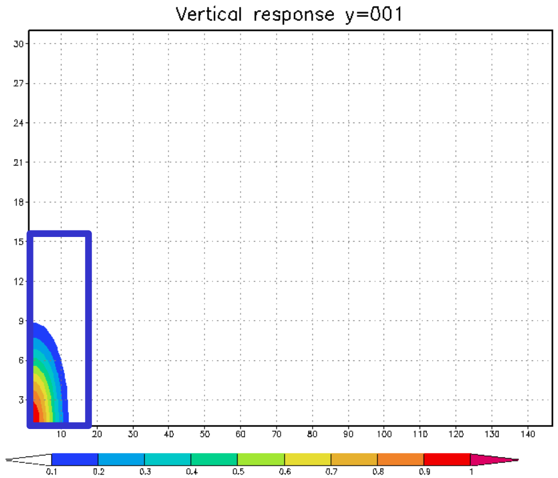
Vertical response (Q + localization + re-scaling)



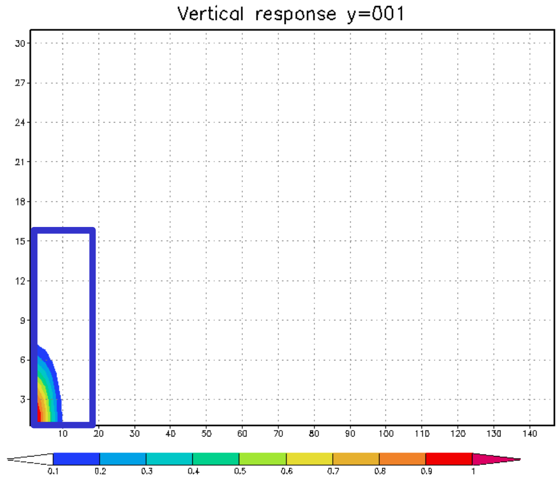
Vertical response becomes also acceptable after post-processing

Processing reduced rank matrix: corner point (b)

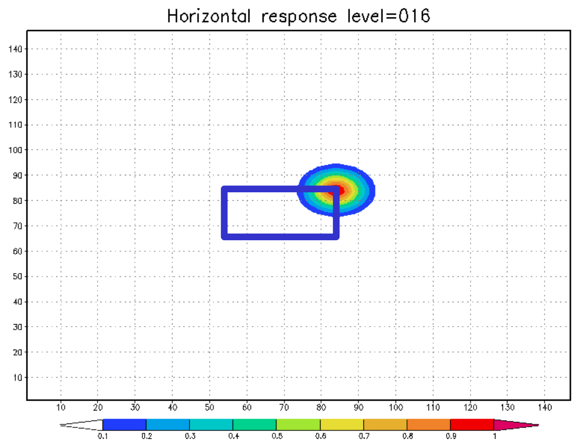
Global corner point (truth)



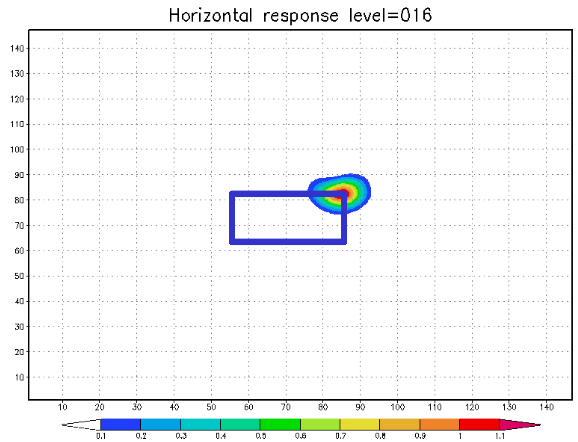
Global corner point (post-processed)



Local corner point (truth)



Local corner point (post-processed)



Cross-correlation exists even for corner points

Summary and future work

- **Sufficient rank achieved with relatively small number of additional columns**
  - important for computational reasons
- **Multivariate response**
  - evaluate the reduced-rank impact on cross-variable correlations
- **Implement hybrid covariance**
  - augment ensemble covariance by adding static covariance columns
  - define orthogonally complement subspaces
- **Tests of the new hybrid method in realistic systems**
  - NASA WRF, NOAA WRF-NMM, NOAA hurricane WRF, WRF-CHEM,
  - assimilation of NOAA operational observations
  - assimilation of cloud- and precipitation-affected MW and IR radiances
  - assimilation of lightning flash rates