

Review of “Covariance Localization” in Ensemble Filters

Tom Hamill

NOAA Earth System Research Lab, Boulder, CO

tom.hamill@noaa.gov

Canonical ensemble Kalman filter update equations

$$\mathbf{x}_i^a = \mathbf{x}_i^b + \mathbf{K} (\mathbf{y}_i - H\mathbf{x}_i^b)$$

H = (possibly nonlinear)
operator from model to
observation space

$$\mathbf{K} = \mathbf{P}^b H^T (H \mathbf{P}^b H^T + \mathbf{R})^{-1}$$

\mathbf{x} = state vector
($_i$ for *ith* member)

$$\mathbf{P}^b = \mathbf{X} \mathbf{X}^T$$

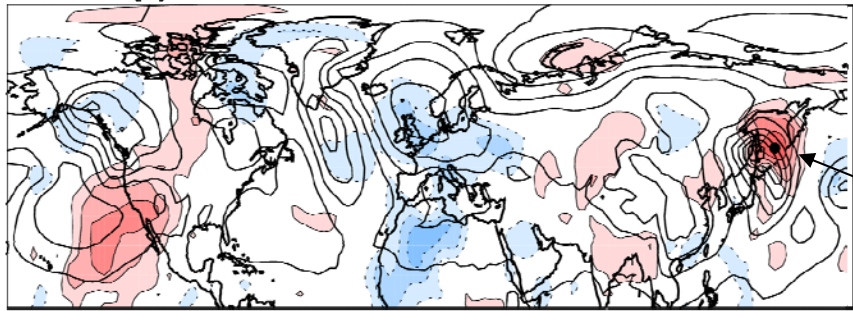
$$\mathbf{y}_i = \mathbf{y} + \mathbf{y}_i'$$

$$\mathbf{X} = \left(\mathbf{x}_1^b - \overline{\mathbf{x}^b}, \mathbf{K}, \mathbf{x}_n^b - \overline{\mathbf{x}^b} \right)$$

$$\mathbf{y}_i' \sim N(0, \mathbf{R})$$

- Notes:
- (1) An ensemble of parallel data assimilation cycles is conducted, here assimilating *perturbed observations*.
 - (2) Background-error covariances are estimated using the ensemble.

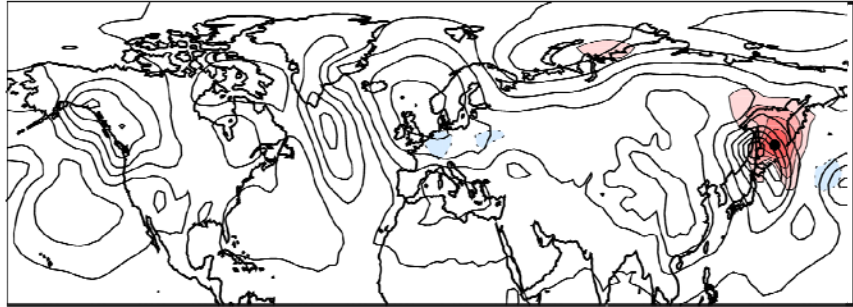
(a) Correlations in P^b , 25-member ensemble



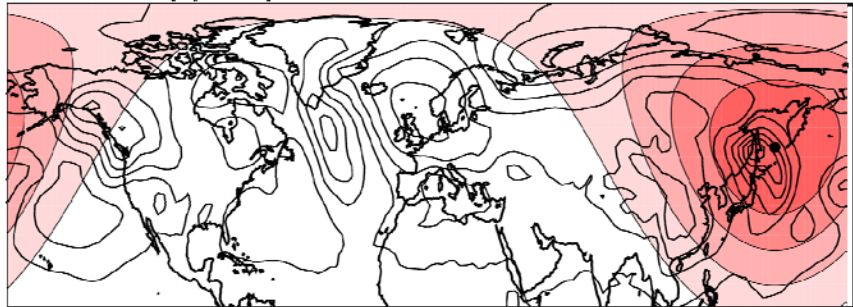
obs
location

Example of covariance localization

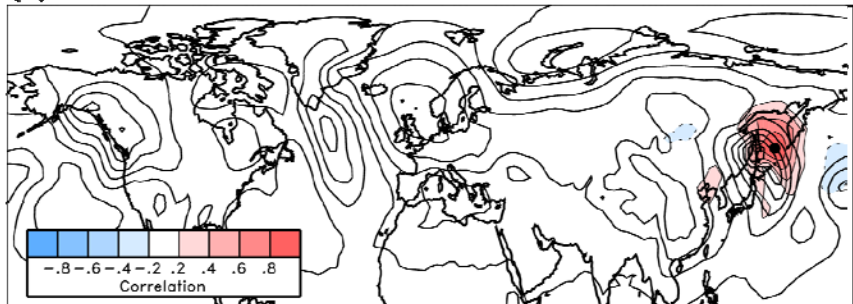
(b) Correlations in P^b , 200-member ensemble



(c) Gaspari & Cohn correlation function



(d) Correlations in P^b after localization, 25-member ensemble



Background-error correlations estimated from 25 members of a 200-member ensemble exhibit a large amount of structure that does not appear to have any physical meaning. Without correction, an observation at the dotted location would produce increments across the globe.

Proposed solution is element-wise multiplication of the ensemble estimates (a) with a smooth correlation function (c) to produce (d), which now resembles the large-ensemble estimate (b). This has been dubbed “covariance localization.”

How covariance localization is commonly implemented in EnKF

$$\hat{\mathbf{K}} = (\rho_s \circ \hat{\mathbf{P}}^b) H^T \left[H (\rho_s \circ \hat{\mathbf{P}}^b) H^T + \mathbf{R} \right]^{-1}$$

which is approximately

$$\hat{\mathbf{K}} = \left[\rho_s \circ (\hat{\mathbf{P}}^b H^T) \right] \left[\rho_s \circ (H \hat{\mathbf{P}}^b H^T) + \mathbf{R} \right]^{-1}$$

ρ_s is typically a compactly supported function similar in shape to a Gaussian, 1.0 at observation location. See Gaspari and Cohn, 1999, *QJRMS*, **125**, 723-757, eq. 4.10

define $\overline{H\mathbf{x}^b} = \frac{1}{m} \sum_{i=1}^m H\mathbf{x}_i^b$

$$\hat{\mathbf{P}}^b H^T = \frac{1}{m-1} \sum_{i=1}^m (\mathbf{x}_i^b - \overline{\mathbf{x}^b}) (H\mathbf{x}_i^b - \overline{H\mathbf{x}^b})^T$$

$$H \hat{\mathbf{P}}^b H^T = \frac{1}{m-1} \sum_{i=1}^m (H\mathbf{x}_i^b - \overline{H\mathbf{x}^b}) (H\mathbf{x}_i^b - \overline{H\mathbf{x}^b})^T$$

Why localization ? Relative sampling error's dependence on ensemble size, correlation.

True background-error covariance of model with two grid points:

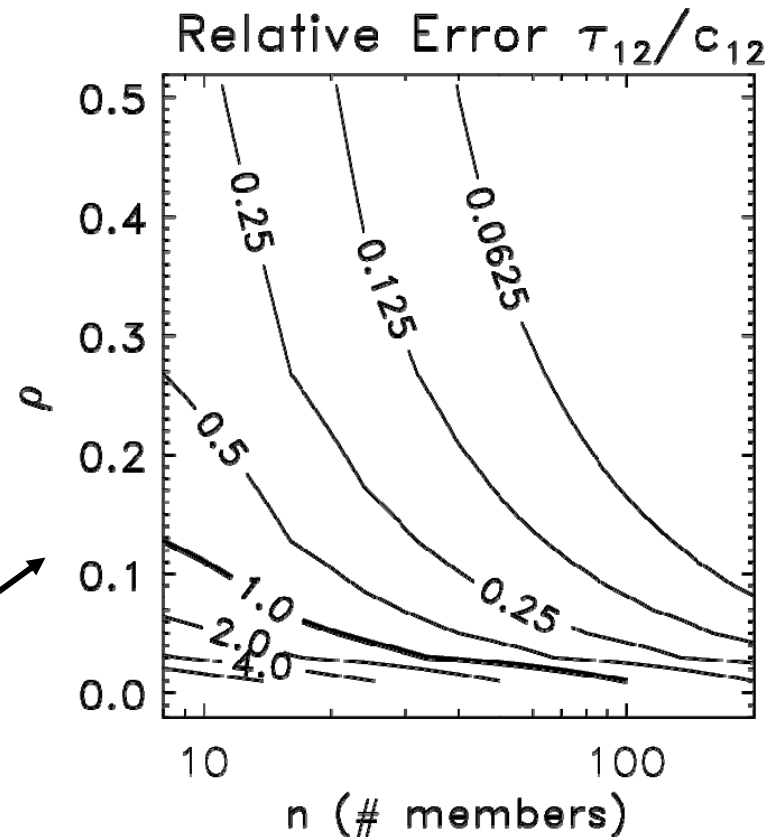
$$\mathbf{P}^b = \begin{pmatrix} \sigma_1^2 & c_{12} \\ c_{12} & \sigma_2^2 \end{pmatrix}$$

and obs at grid point 1 with error covariance σ

Suppose there is inaccurate estimate of background-error covariance matrix

$$\hat{\mathbf{P}}^b = \begin{pmatrix} \sigma_1^2 & c_{12} + \varepsilon_C \\ c_{12} + \varepsilon_C & \sigma_2^2 \end{pmatrix}, \quad \varepsilon_C \sim N(0, \tau_{12})$$

How do errors change with correlation and ensemble size? **Relative error is large when correlation and/or ensemble size is small.**



Spreads and errors as function of localization length scale

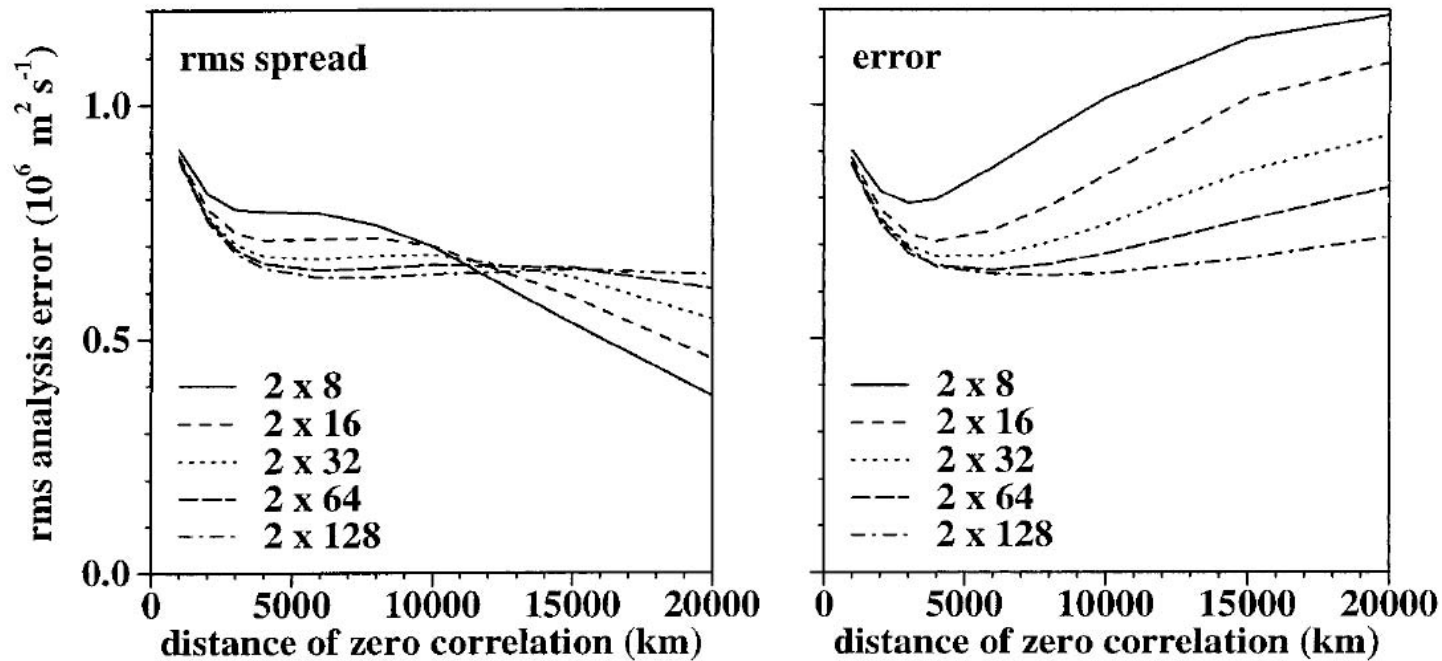
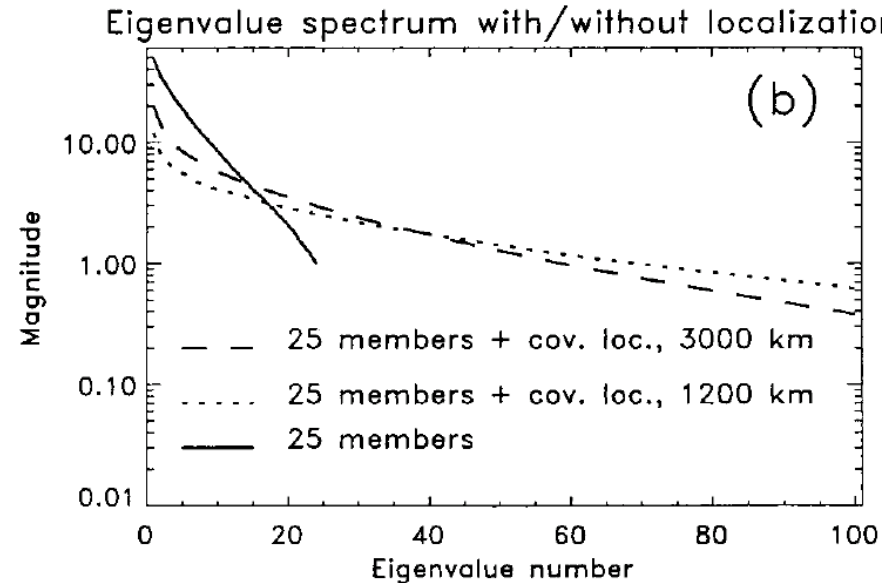
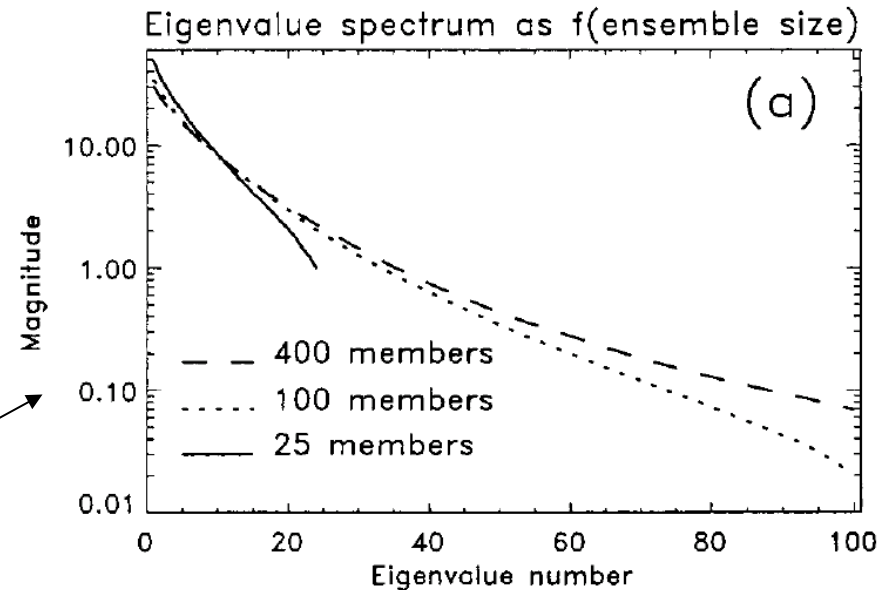


FIG. 4. Analysis error as a function of r_1 (km), the distance beyond which ρ is zero, for various ensemble sizes. (left) The rms spread in the ensemble, and (right) the rms error of the ensemble mean.

- (1) Small ensembles minimize error with short length scale, large ensemble with longer length scale. That is, less localization needed for large ensemble.
- (2) The longer the localization length scale, the less spread in the posterior ensemble.

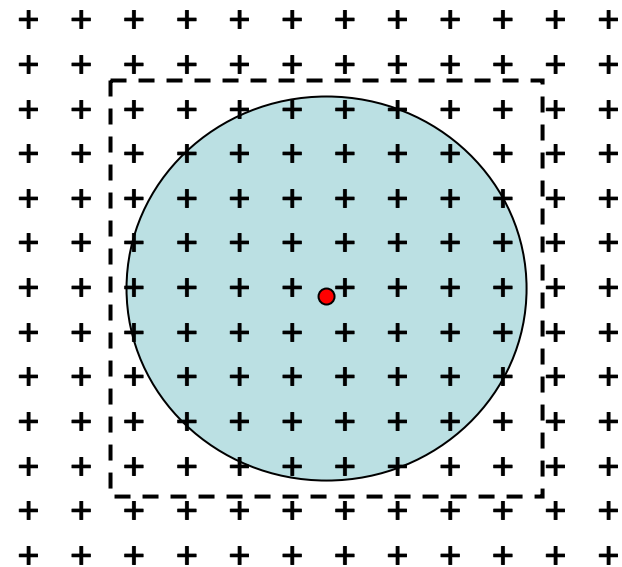
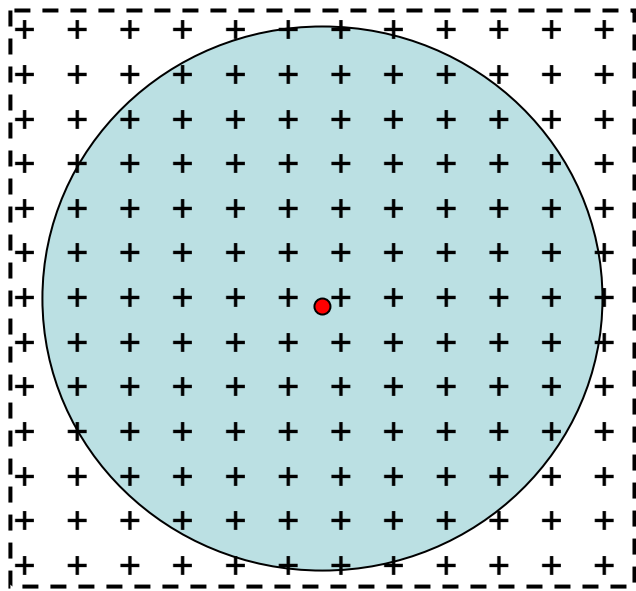
Localization adds effective degrees of freedom to background-error covariance model

- Experiments with toy 2-layer primitive equation model
- With few members, eigenvalue spectrum of background error covariance is very steep, and with n -member ensemble, only $n-1$ degrees of freedom; with more members, less steep spectrum.
- Covariance localization in ensemble increases the effective degrees of freedom, makes eigenvalue spectrum flatter; more ways the ensemble can adjust to the data.
- At the end of the data assimilation, though, you still have posterior ensemble with only $n-1$ degrees of freedom.



Increasing localization scale increases CPU run time

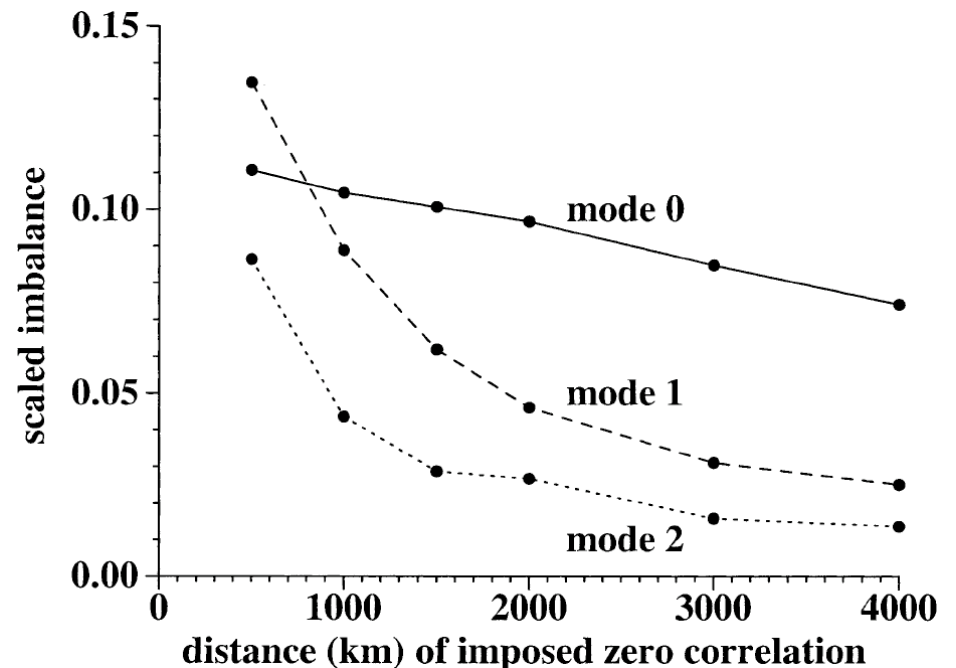
- Ensemble filters commonly serially process the observations.



+ signs are grid points. Red dot: observation location. Blue disc: area spanned by localization function. Dashed line: box enclosing all possible grid points that must be updated given this localization scale.

Covariance localization introduces imbalances

- Scaled imbalance (imbalance / increment magnitude) for 3 gravest vertical modes, pair of 32-member ensembles.
- Imbalance measured by comparing digitally filtered state to unfiltered.
- Larger scaled imbalance at shorter correlation length scales.
- The imbalance may limit the growth of spread, making background-error covariance model estimated from ensemble less accurate.



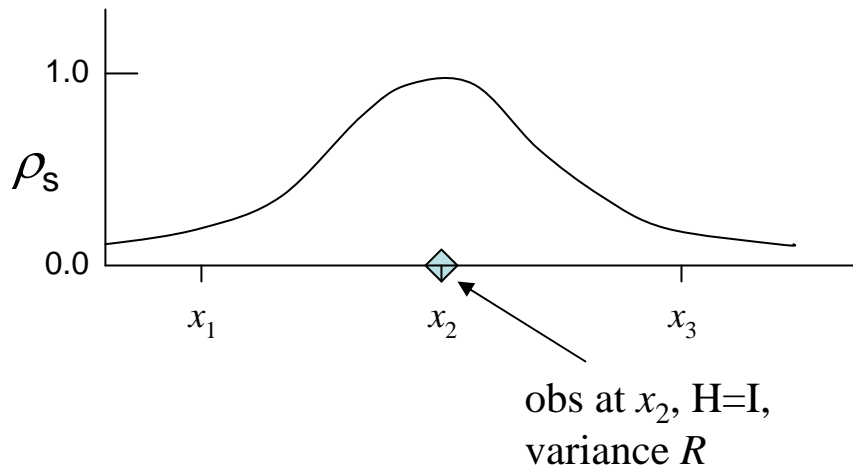
Synthesis

“Localization scale is a tuning parameter that trades off computational effort, sampling error, and imbalance error.”

(Ref: Zhou et al., 2008, *MWR*, **136**, 678-698.)

Some alternatives
to standard
covariance localization

Localization by inversely scaling observation-error covariances



$$\hat{\mathbf{K}} = (\rho_s \circ \hat{\mathbf{P}}^b) H^T [H (\rho_s \circ \hat{\mathbf{P}}^b) H^T + \mathbf{R}]^{-1}$$

$$\hat{K}(x_1) = \frac{\rho_s(x_1) \text{cov}(x_1, x_2)}{\rho_s(x_2) \sigma^2(x_2) + R} = \frac{\text{cov}(x_1, x_2)}{\frac{\rho_s(x_2)}{\rho_s(x_1)} (\sigma^2(x_2) + R)}$$

$$\hat{K}(x_1) \cong \frac{\text{cov}(x_1, x_2)}{\sigma^2(x_2) + \frac{R}{\rho_s(x_1)}}$$

something akin to covariance localization can be obtained by inversely scaling the observation-error variance rather than modifying the background-error covariances. Pioneered in LETKF.

“Hierarchical” ensemble filter

- Suppose m groups of n -member ensembles. Then m estimates of background-error covariance relationships between obs location, model grid point (*).
- Kalman filter update: effectively a linear regression analysis, compute increment in state variable x , given increments for observation variable, y^o . Hence m sample values of β (i.e., $K_{(*)}$) are available.

$$x_{(*)}^a - x_{(*)}^b = \beta(y^b - y^o)$$

- Assume correct β is a random draw from same distribution the m samples drawn from.
- “*Regression confidence factor*” α chosen to minimize

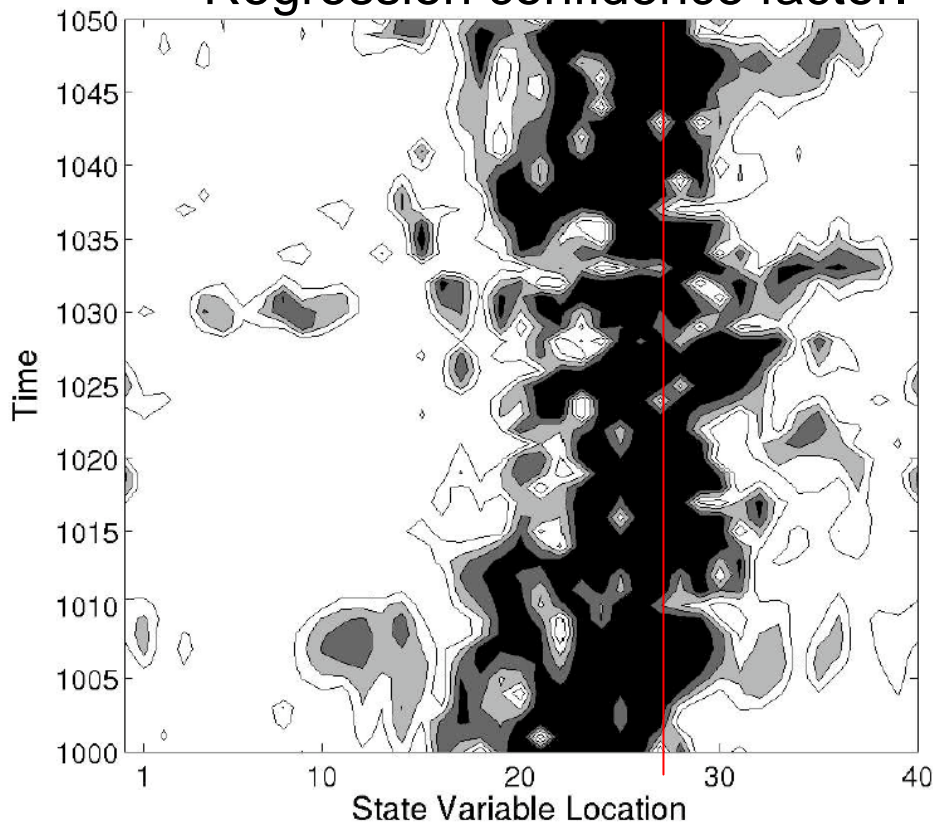
$$\sqrt{\sum_{j=1}^m \sum_{i=1, i \neq j}^m (\alpha \beta_i - \beta_j)^2}$$

- An “on-the-fly” moderation of covariances; replace β with $\alpha \beta$.

Anderson's hierarchical ensemble filter (continued)

$$\alpha_{\min} = \left\{ \left[\frac{\left(\sum_{i=1}^m \beta_i \right)^2}{\sum_{i=1}^m \beta_i^2} \right] - 1 \right\} / (m-1)$$

Regression confidence factor.



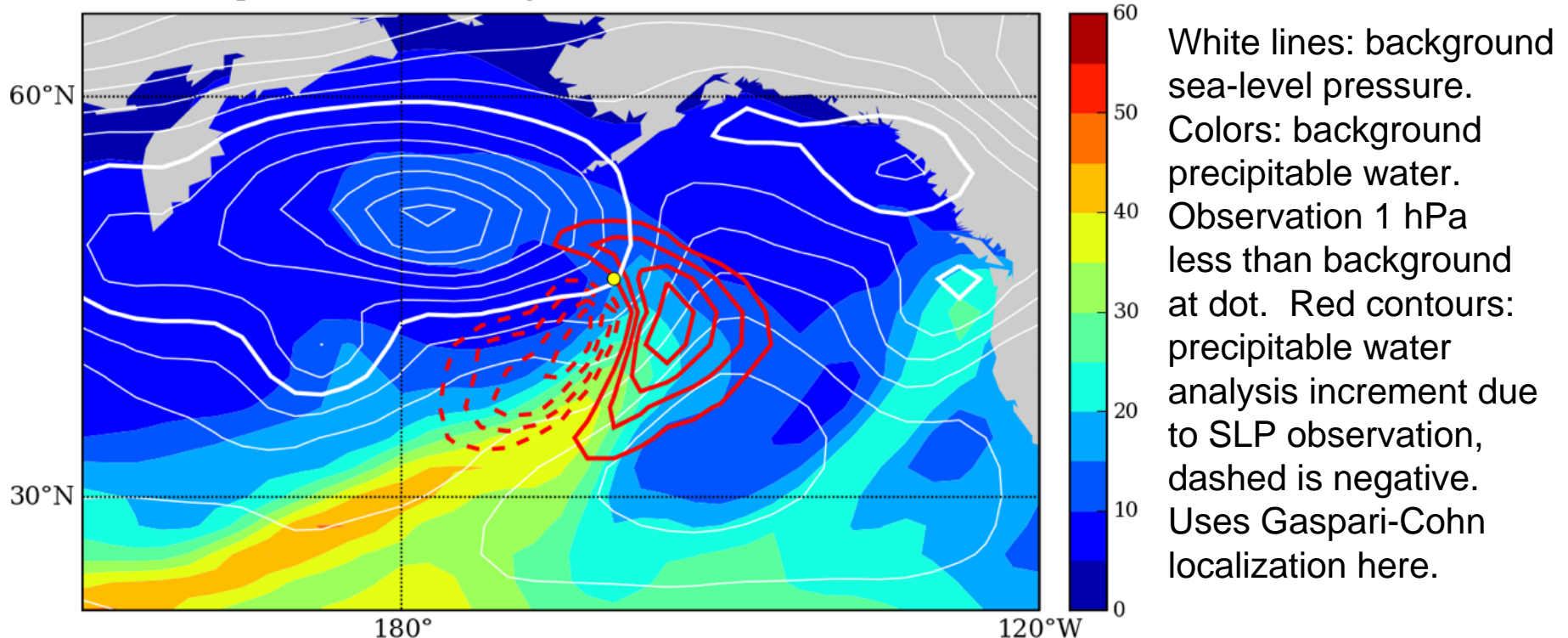
Hovmoller of regression confidence factor for Lorenz '96 model assimilations for observations at red line, 16-group ensemble, 14 members.

$\alpha_{\min} \rightarrow 0$ when spread of β is large relative to mean. $\alpha_{\min} \rightarrow 1$ when spread is small relative to mean.

Is the variability with time a feature or a bug?

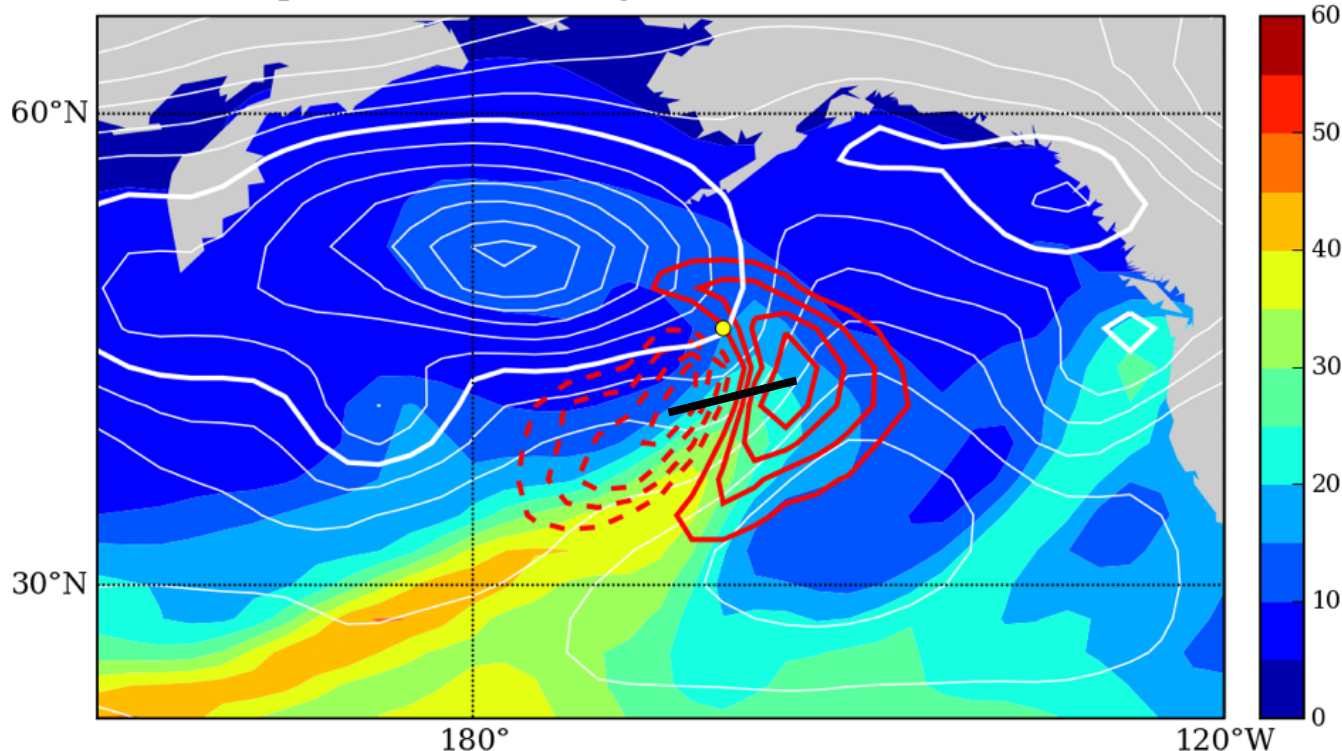
Example of where regression confidence factor may induce unrealistic increments

Precipitable Water Analysis Increment 2004013000



Example of where regression confidence factor may induce unrealistic increments

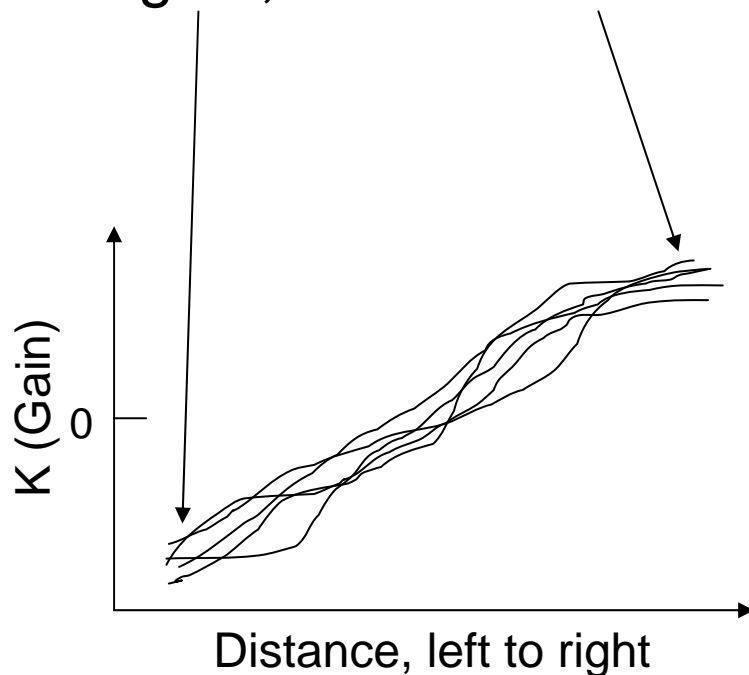
Precipitable Water Analysis Increment 2004013000



Imagine gain calculation along black line, and how it may vary if several parallel ensemble filters were run.

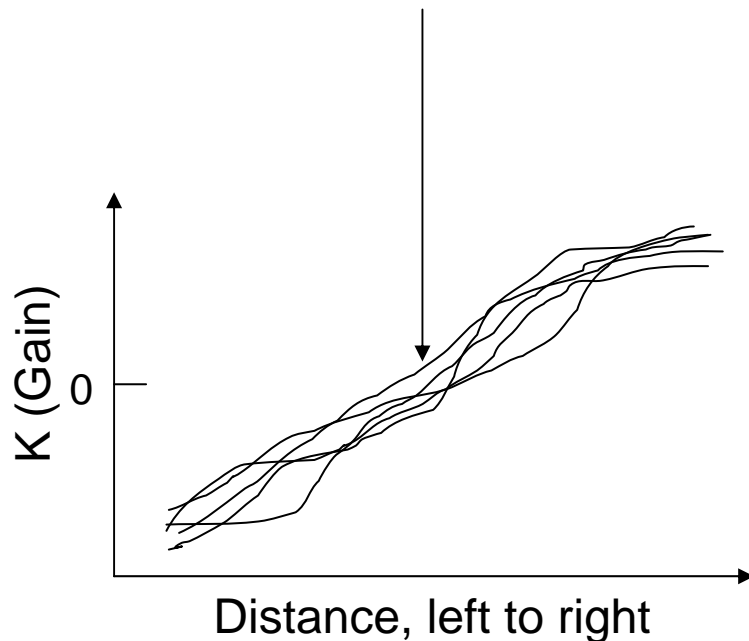
Gain calculations along cross-section

Here, the spread in gains is **small** relative to the magnitude of gain; hence $\alpha \approx 1$



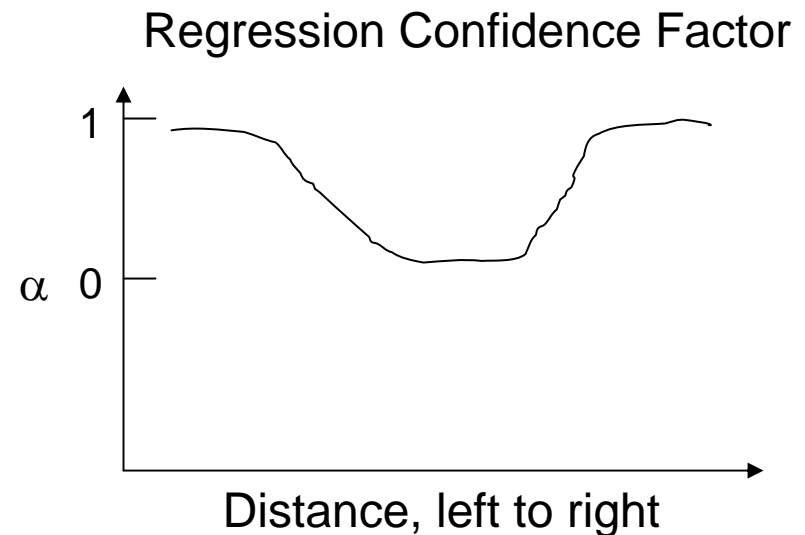
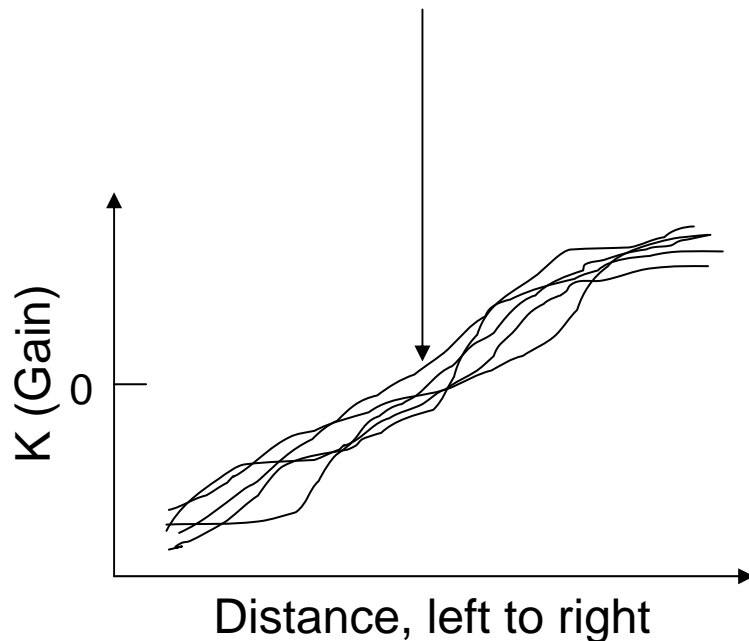
Gain calculations along cross-section

Here, the spread in gains is **large** relative to the magnitude of gain; hence $\alpha \approx 0$



Gain calculations along cross-section

Here, the spread in gains is large relative to the magnitude of gain; hence $\alpha \approx 0$



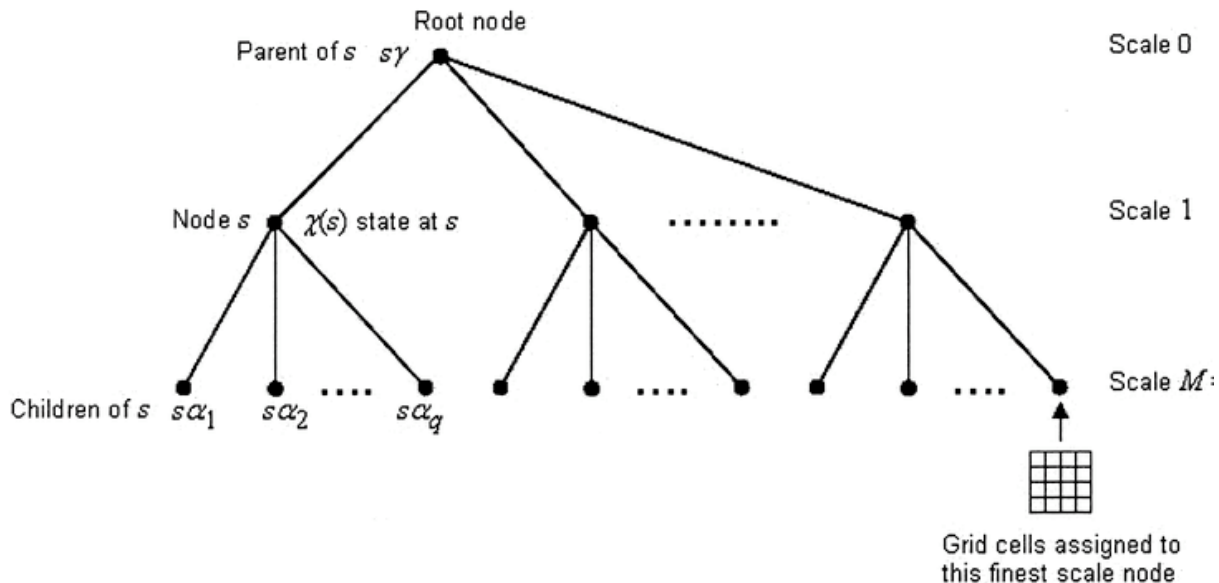
Gain calculations along cross-section

The product of α & K now produces a larger region in the middle with nearly zero gain, and hence near-zero increment. This may be, in some circumstances, non-physical.



Other localization ideas

- **Multiscale filter:** replaces at each update time, the sample covariance with a multiscale tree, composed of nodes distributed over a relatively small number of discrete scales
 - Ref: Zhou et al. 2008, *MWR*, **136**, 678-698.



Other localization ideas

- **Spectral localization:** Filter covariances after transformation to spectral space, or combine spectral and grid-point localization.
 - Ref: Buehner and Charron, 2007, *QJRMS*, **133**, 615-630.

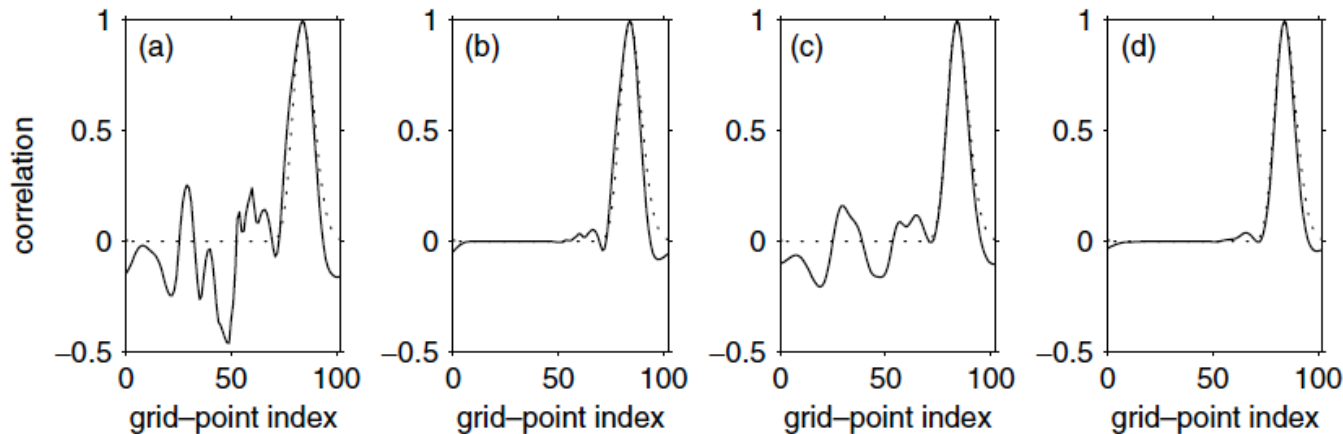


Figure 1. True correlation function (dotted lines) and estimated correlation function (solid lines) for a location near the domain edge where the true correlations are broad. (a) Original sample estimate. (b) Sample estimate with optimal amount of spatial localization (localization radius of 44 grid points). (c) Sample estimate with optimal amount of spectral localization (localization radius of 11). (d) Sample estimate with optimal amount of spatial and spectral localization.

“Successive Covariance Localization”

Fuqing Zhang, Penn State

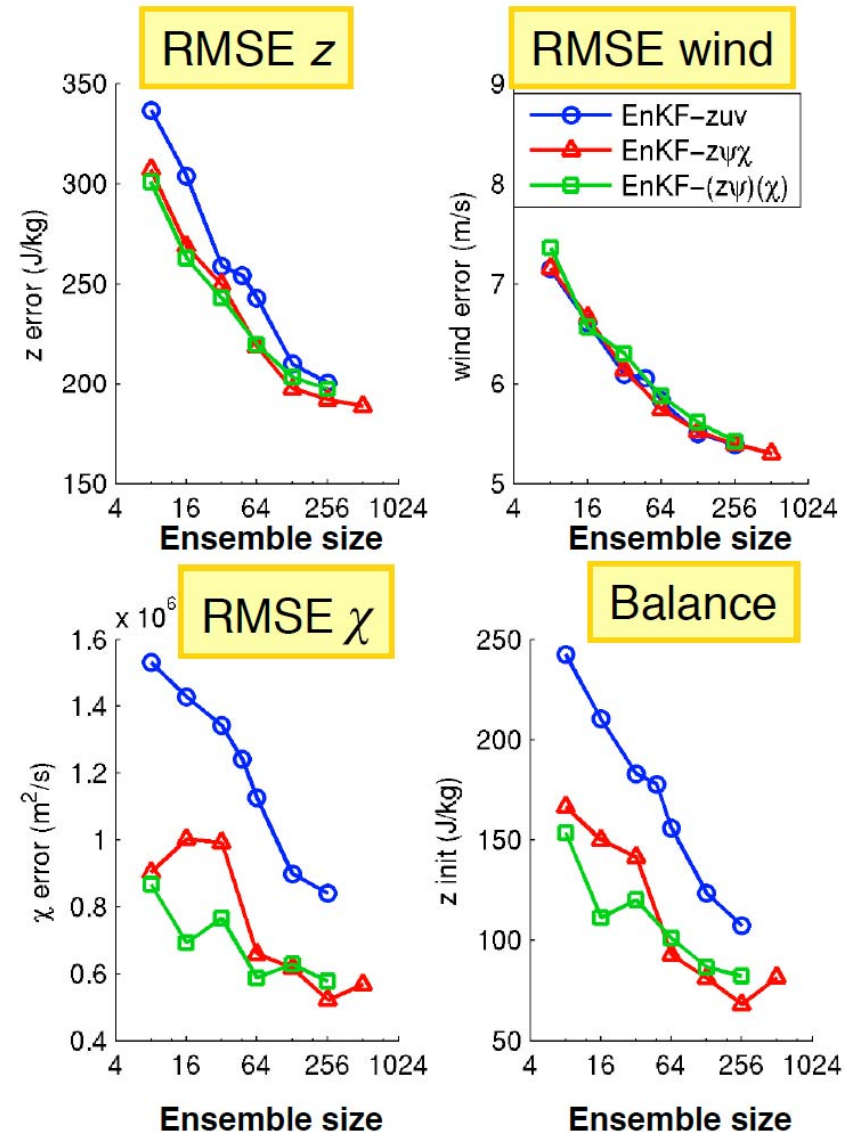
- Adjust ensemble to regular network of observations using a broad localization length scale.
 - Where observations are sparse, use of short localization length scale will result in imbalances, “holes” in analysis-error spread, increased analysis errors.
- Adjust modified ensemble further to high-density observations using a short localization length scale.
 - Where dense observations, use of long localization length scale will smooth out mesoscale detail, trust ensemble covariance model too much, and strip too much variance from ensemble.
- Effective compromise that works across both data-dense and data-sparse regions?

Change model variables?

- Localization in (z, ψ, χ) (streamfunction, velocity potential) instead of (z, u, v) ?

- Ref: J. Kepert, 2006, tinyurl.com/4j539t

- Figure to right: errors for normal localization, localization in (z, ψ, χ) space, and localization zeroing out $(z- \psi)$ & $(\psi-\chi)$



Ensemble Correlations Raised to a Power (ECO-RAP)

- Craig Bishop will present this idea in more detail in the next talk.

Conclusions

- Some form of covariance localization is a practical necessity for EnKF NWP applications.
- Because of drawbacks (e.g., imbalances introduced), active search for improved localization algorithms.
- Many interesting alternatives, but community not yet united behind one obviously superior approach. Alternatives tend to have their own set of strengths and weaknesses.