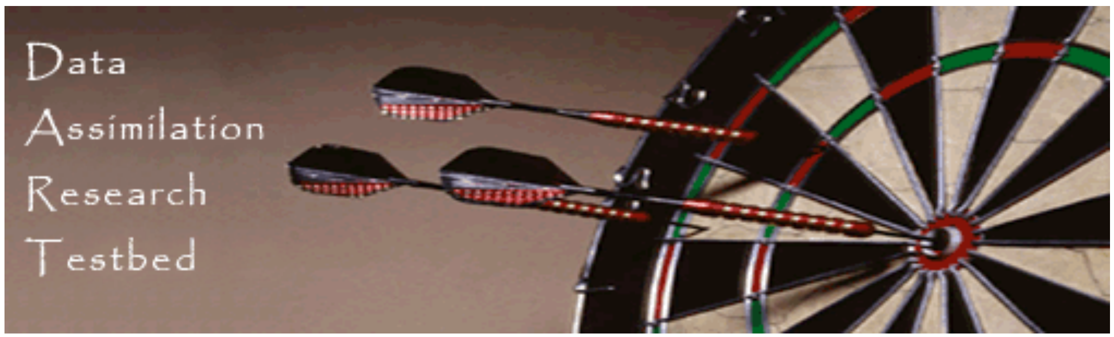
Localization and Correlation in Ensemble Kalman Filters

Jeff Anderson



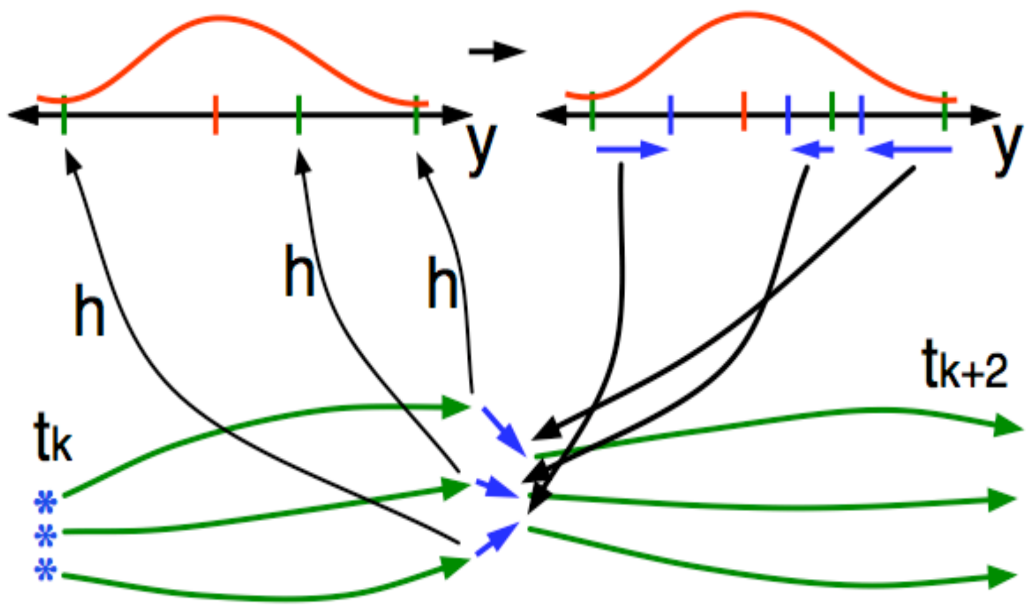
NCAR Data Assimilation Research Section

The National Center for Atmospheric Research is sponsored by the National Science Foundation.

NCAR

Sequential ensemble filter assimilation

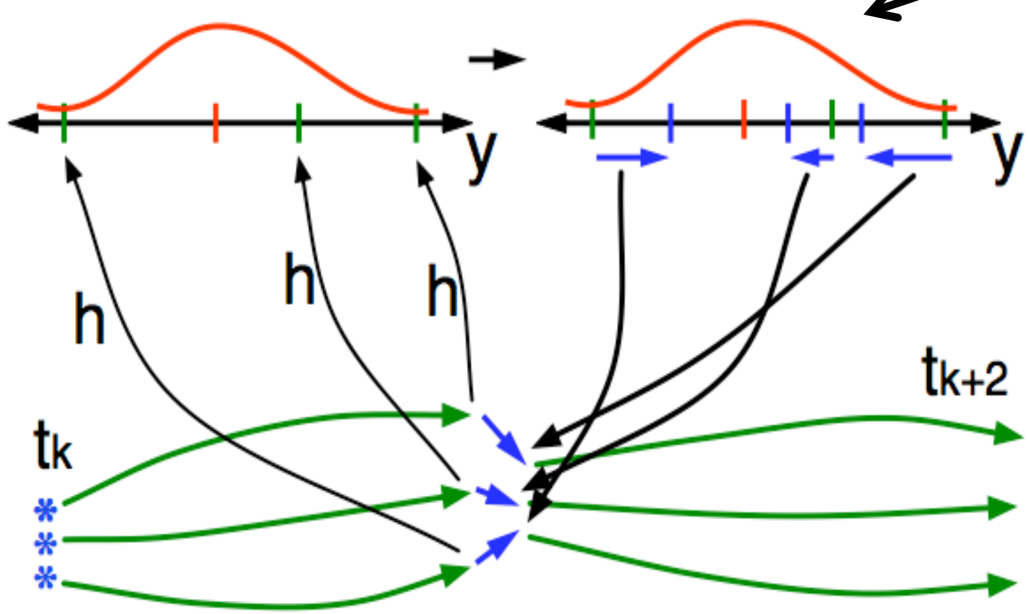
- Impact of single ob y on scalar state component x.
- Describes problem without loss of generality.



Sequential ensemble filter assimilation

- Impact of single ob y on scalar state component x.

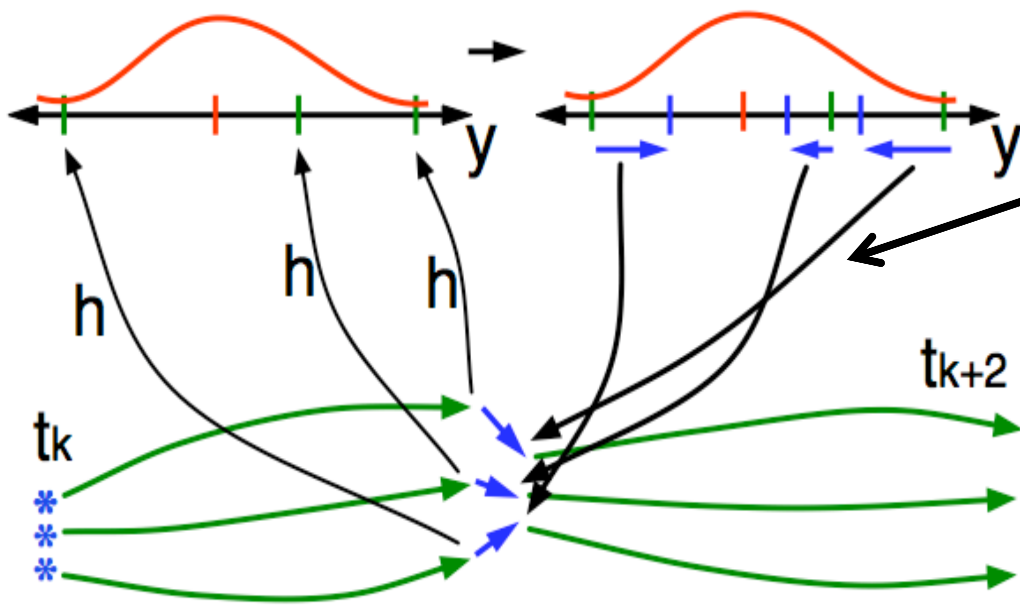
Compute increments for prior y ensemble.
(Details irrelevant for this talk)



Sequential ensemble filter assimilation

- Impact of single ob y on scalar state component x.

Regress y increments onto x.



$\Delta x_i = b\Delta y_i$, $i=1,\ldots N$

N is ensemble size.

b is regression coefficient.

Ensemble filters are optimal and exact when…

- Model is linear,
- Observation operators are linear,
- Observation error is gaussian,
- Ensemble size $N > N_{crit}$,
- Filter is EAKF (or similar deterministic).

Violating any of these may lead to suboptimal solution.

Most commonly observed errors are…

- Too little variance (often solved by inflation).
- Too much correlation (often solved by localization).

We will treat all errors as sampling errors.

Introduce sampling error correction that acts as a localization.

Working definition of localization

- Multiply regression coefficient by factor α

  $\Delta x_i = \alpha b \Delta y_i$, $\quad$ i=1,…N.

- α often a function of distance between observation y and state variable x.
- Gaspari Cohn compact pseudo-gaussian function.
- Traditionally, α<1, correlation never too small.

Using additional prior information

Need for localization implies that standard ensemble is suboptimal.

Need additional prior assumptions to improve it.

It’s nice to know what these assumptions are, but not always possible.

Sampling error correction as a localization

Let $\hat{b}$ be ensemble sample regression.

Assume that there is sampling error.

View $\hat{b}$ as draw from $Normal\left(\overline{b}, \sigma_b\right)$.

Find value of α that minimizes expected RMS error of $\overline{x}$ after regression.

This can be done analytically…

$$\alpha = \overline{b}^{\,2} \Big/ \left(\sigma_b^2 + \overline{b}^{\,2}\right) = Q^2 \Big/ \left(1 + Q^2\right)$$

where

$$Q = \overline{b} \Big/ \sigma_b$$

But, all I have is single estimate of $\hat{b}$.

Group filter (old work):

- Run several ensembles.
- Get set of $\hat{b}$ estimates.
- Estimate $\bar{b}$ and σ_b from this set.

- But, extra cost of group often too high.

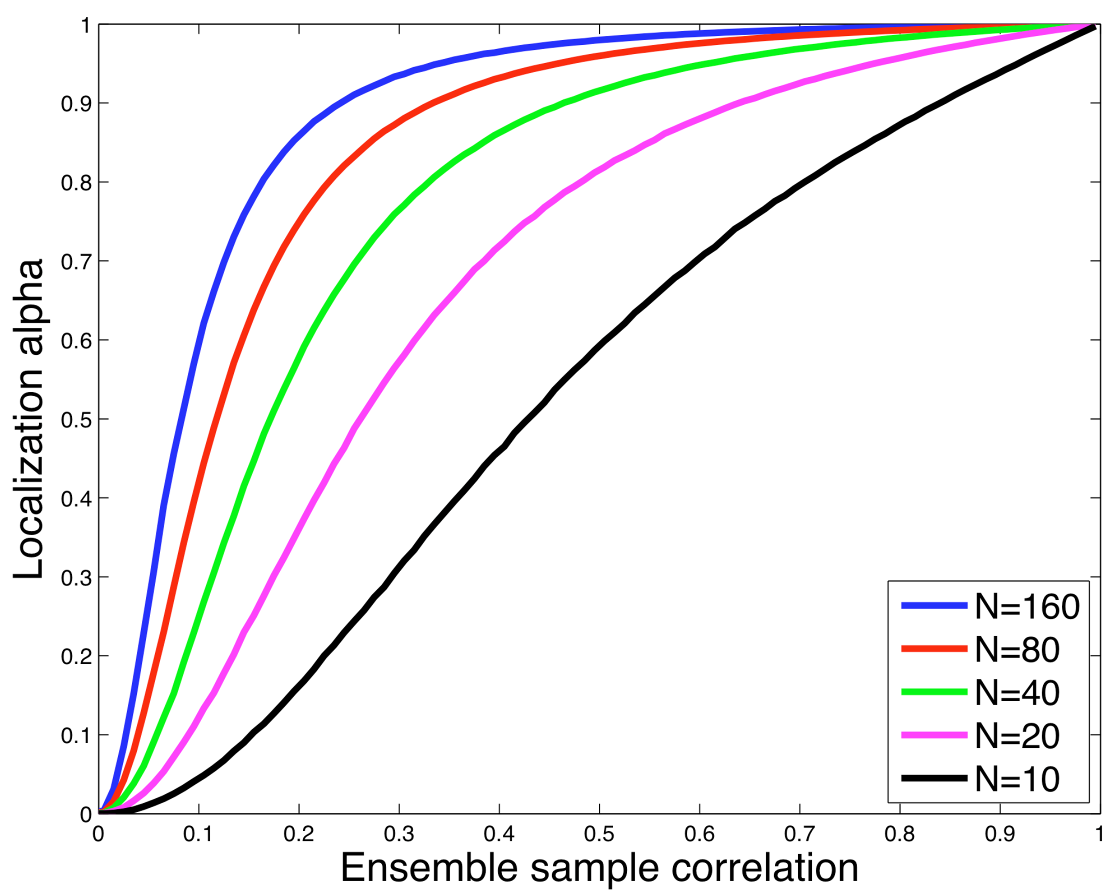
Solution: Use a prior distribution for correlation r

- Try U(-1, 1), all correlations equally likely.
- Given this prior and ensemble sample $\hat{r}$ use Monte Carlo to estimate…
  - Most likely value of ‘true’ correlation r,
  - Underlying values of $\overline{b}$ and σ_b ,
  - Optimal value of α.

Replace $\hat{r}$ with r , and localize with α.

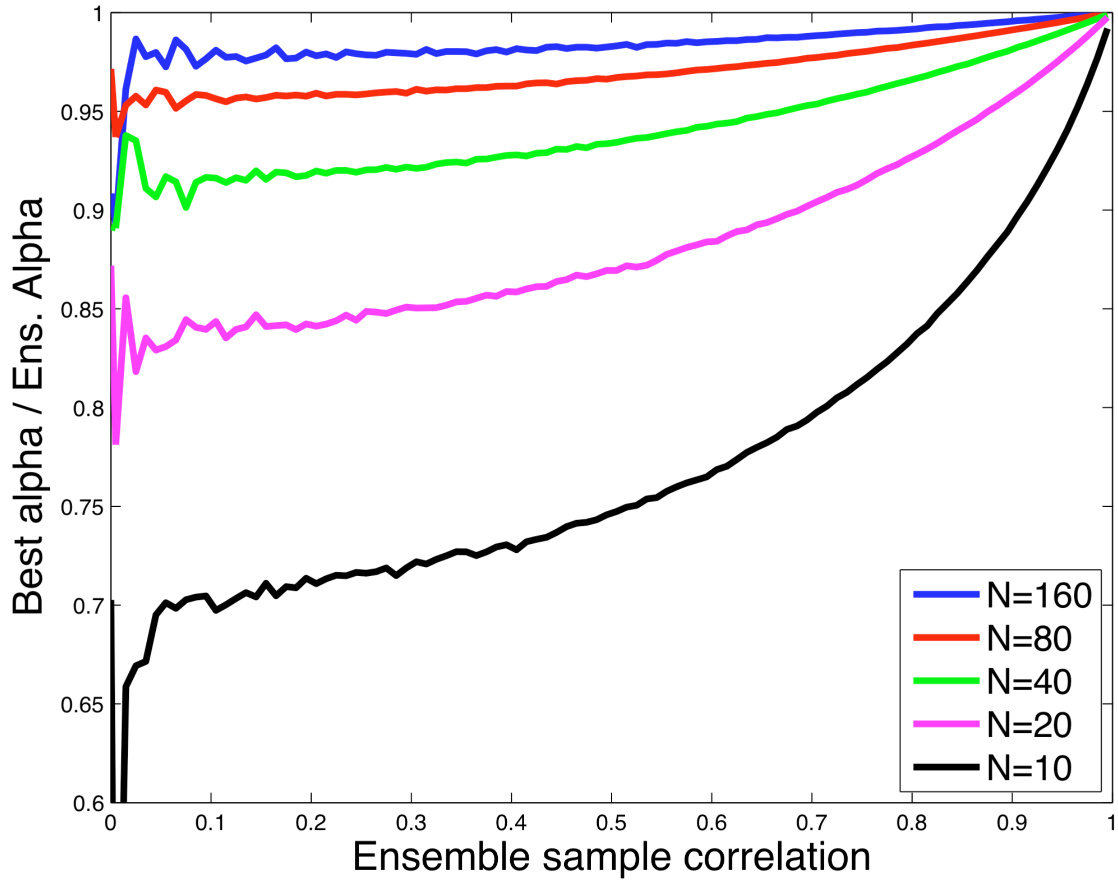
Only a function of $\hat{r}$ and N so we can precompute r and α.

Localization is function of N and sample correlation $\hat{r}$.



This is precomputed for table look-up.

Correction for r is also function of N and $\hat{r}$.



This is precomputed for table look-up.

Test I: Separable linear model, identity obs.

$x_{j,t+1} = (1 + c)x_{j,t}$ j = 1, 200.

Observation error variance is N(0, 1).

Has 200 degrees of freedom.

N>200 EAKF converges to optimal solution.

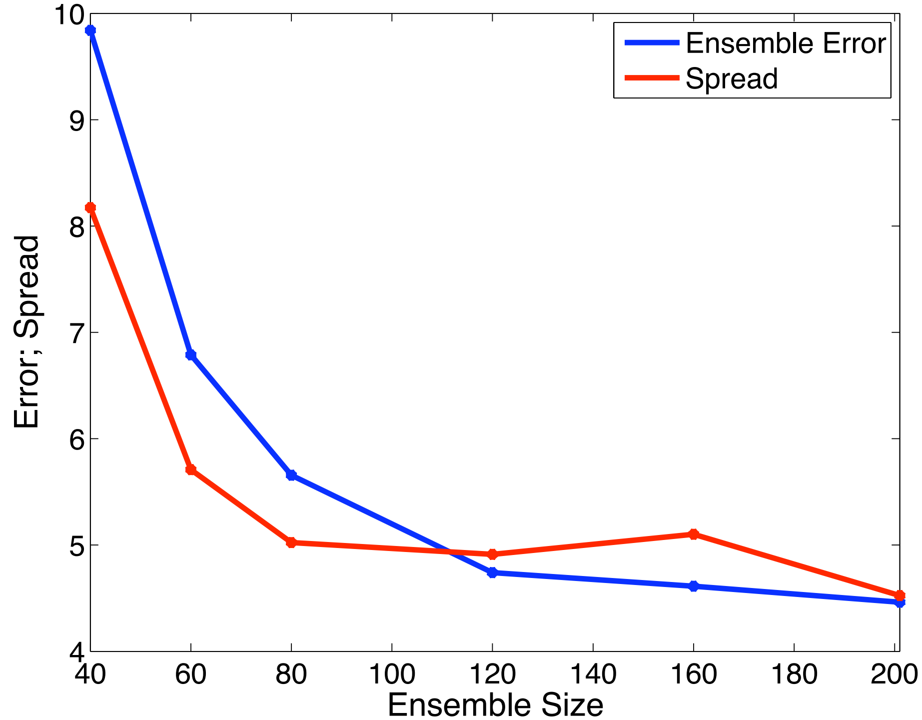
N<201 diverges without sampling error correction.

For N<201, optimal localization is delta function.

Compute time mean RMS and α for various N.

Note: Adaptive inflation used to stabilize.

Test I: Separable linear model, identity obs.

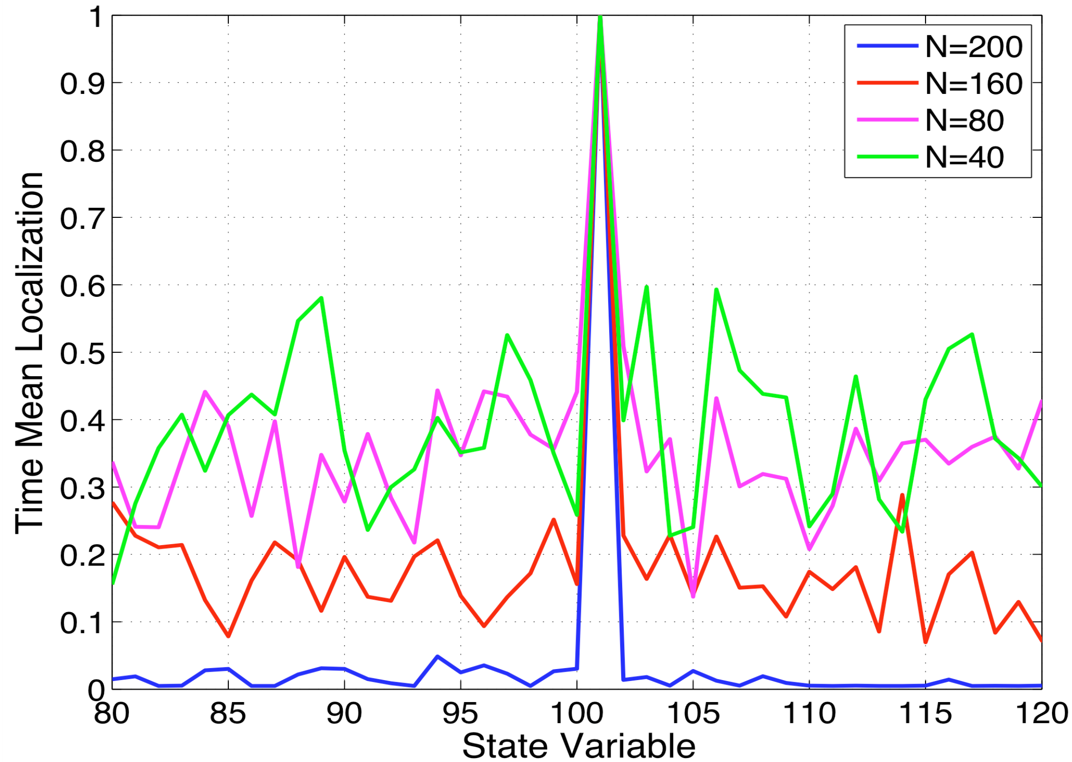


Time mean RMSE and spread as function of N.

N=201 is nearly optimal.

N<201 fails without sampling error correction localization.

Test I: Separable linear model, identity obs.



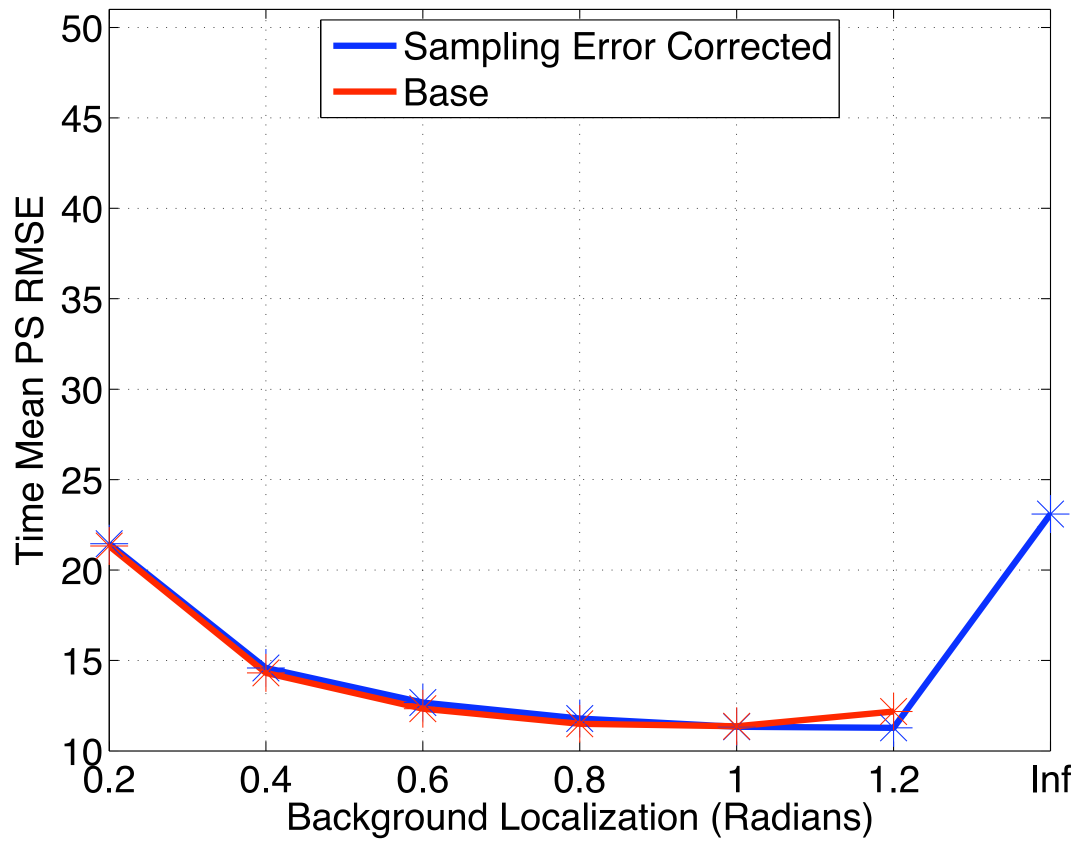
Time mean localization of observation 101.
Should be delta function.
Ability to detect noise degrades as N decreases.

Test II: Low-order dry dynamical core

5 levels, 30 latitudes, 60 longitudes.
Has baroclinic instability.
Observations at 300 randomly located ‘radiosondes’:
	Every 12 hours,
	Obs errors: PS 1 hPa, T 1K, U and V 1 m/s.
1000 assimilation days.

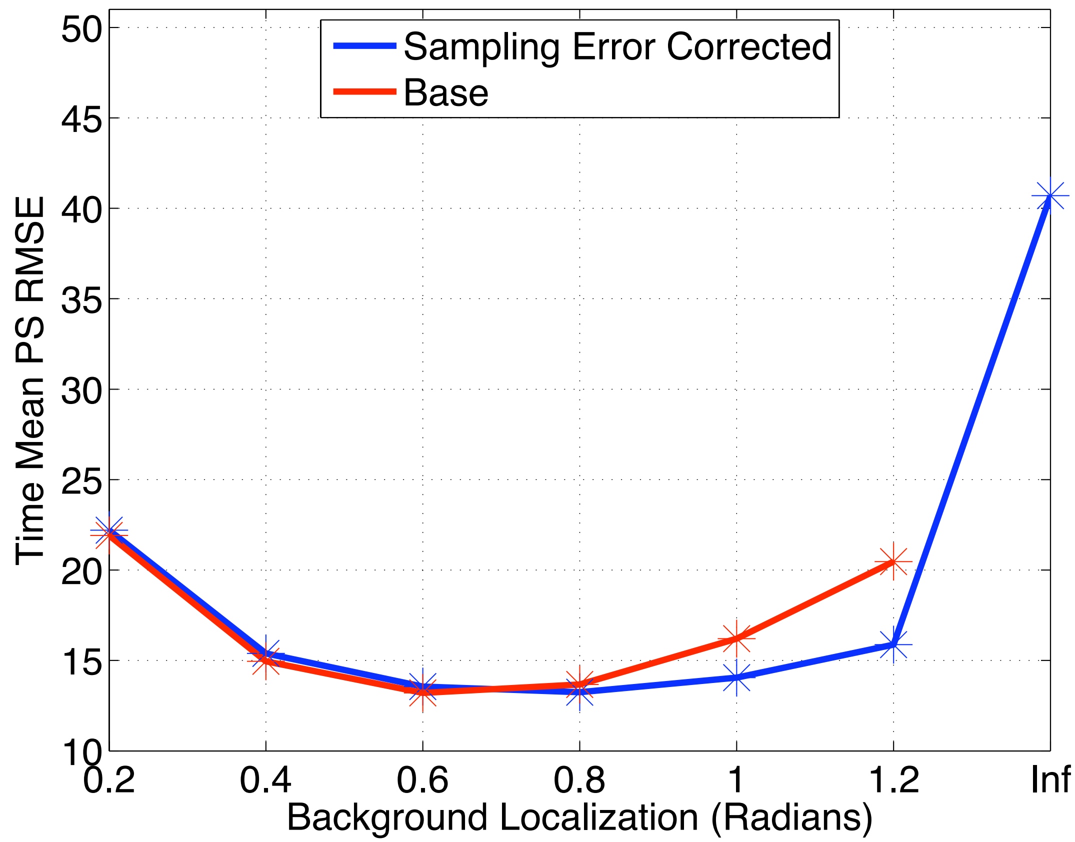
Adaptive spatially-varying inflation.
Background Gaspari Cohn localization.

Test II: Low-order dry dynamical core



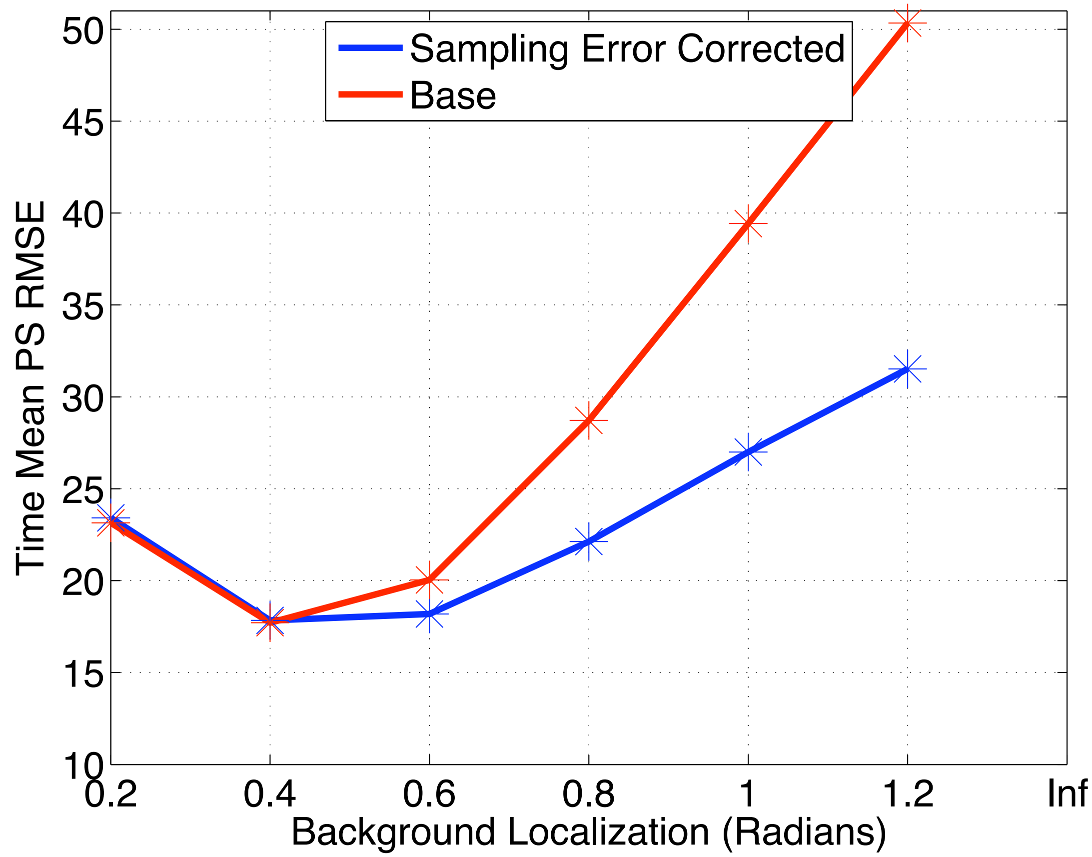
Time mean prior RMSE for PS (Pa) for N=80.
No sampling error correction with no GC diverges.

Test II: Low-order dry dynamical core



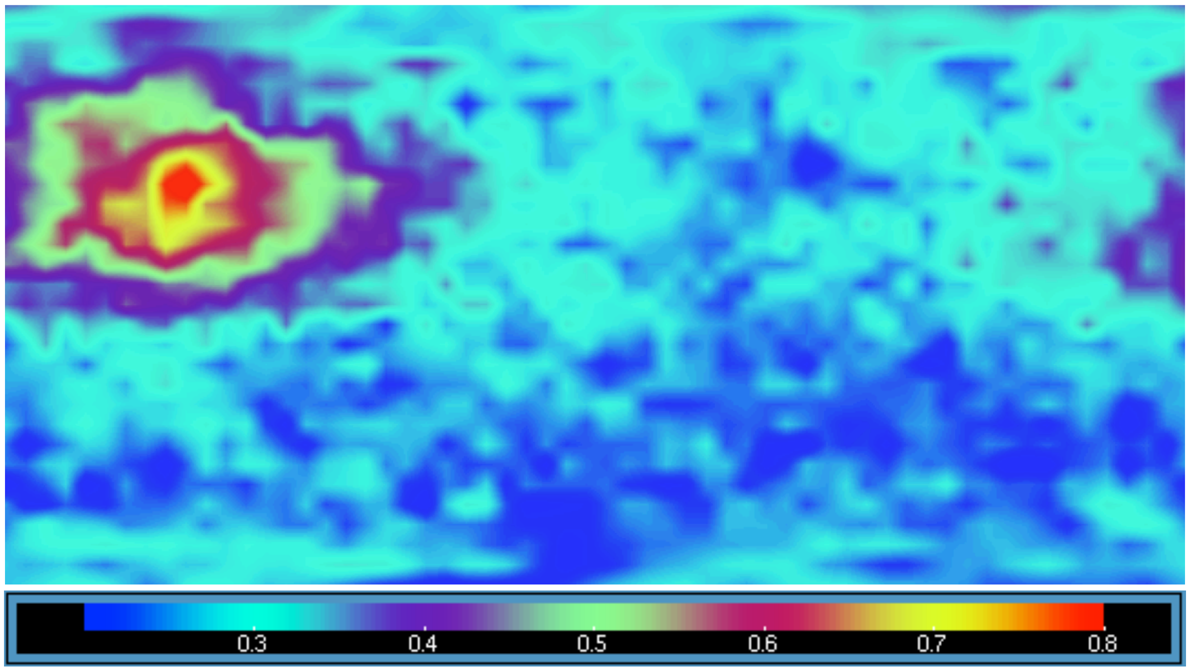
Time mean prior RMSE for PS (Pa) for N=40.
No sampling error correction with no GC diverges.

Test II: Low-order dry dynamical core



Time mean prior RMSE for PS (Pa) for N=20.
Sampling error correction better for large GC.

Test II: Low-order dry dynamical core

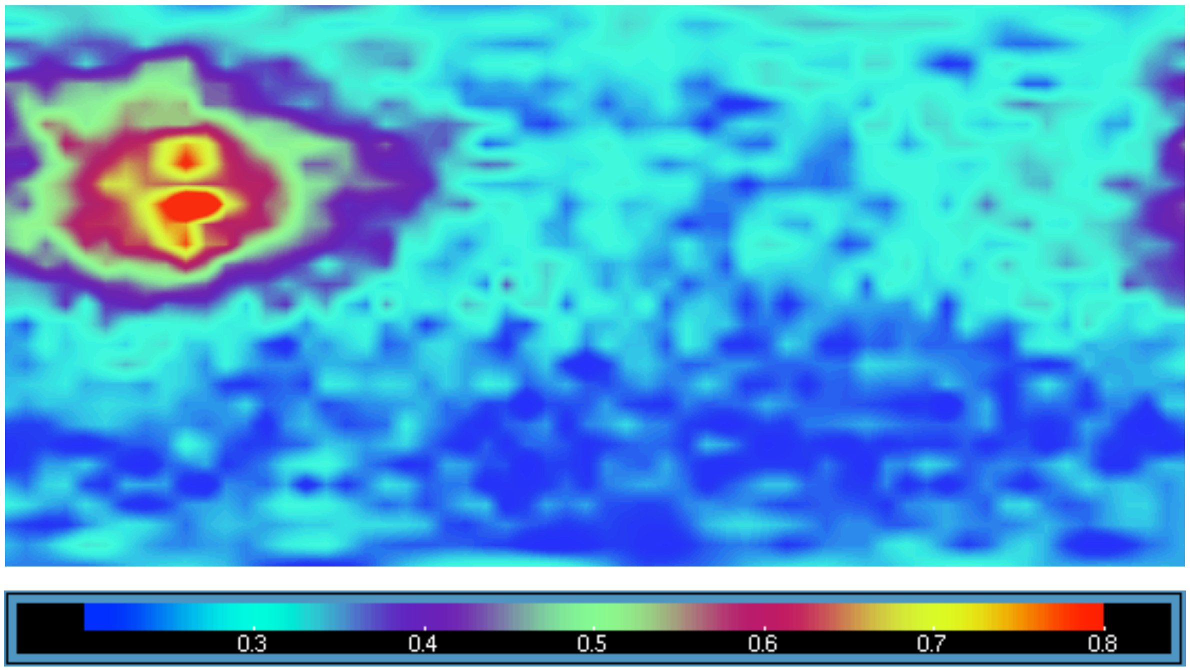


Time mean localization for observation of level 3 U
on level 3 T state variables.

N=80, no Gaspari Cohn.

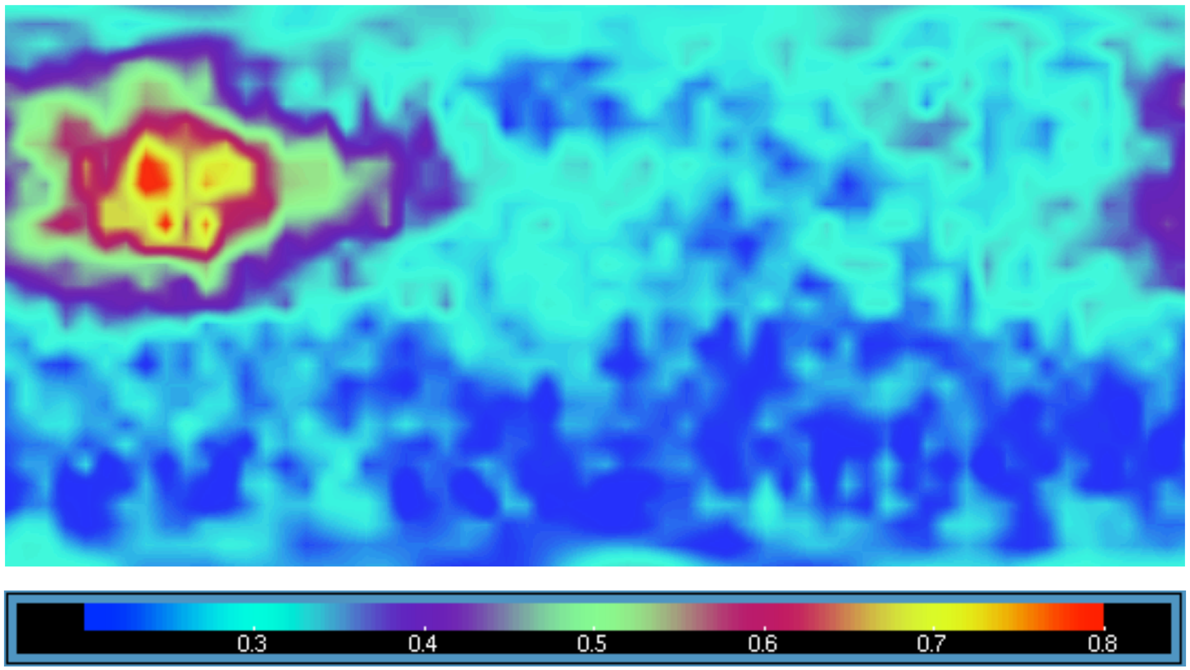
Nearly gaussian but,
maximum is not co-located with observation.

Test II: Low-order dry dynamical core



Time mean localization for observation of level 3 U
on level 3 U state variables.
N=80, no Gaspari Cohn.
Localization is tripole (similar to group filter results).

Test II: Low-order dry dynamical core

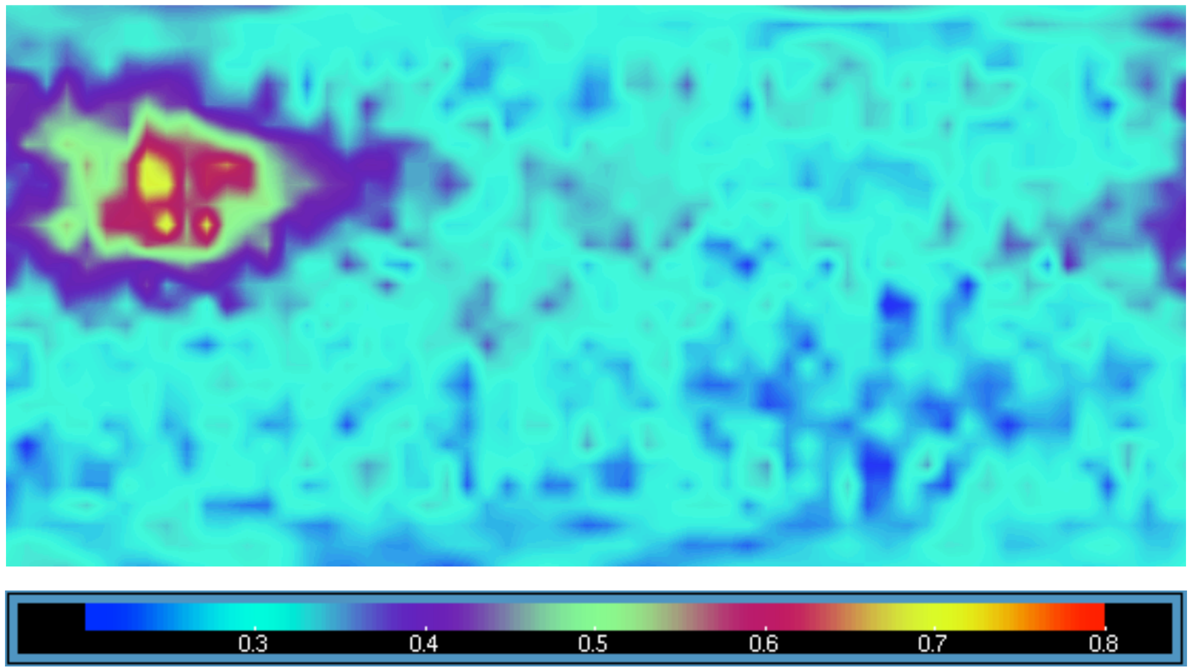


Time mean localization for observation of level 3 U on level 3 V state variables.

N=80, no Gaspari Cohn.

Localization is quadrupole (similar to group filter results).

Test II: Low-order dry dynamical core



Time mean localization for observation of level 3 U
on level 3 V state variables.
N=40, no Gaspari Cohn.
Maximum localization is smaller.
Background minimum localization is larger (0.3 vs. 0.2).

Conclusions

- Implemented sampling error correction dynamic localization.
- Localization is function of correlation and ensemble size.
- Improves performance for low-order examples.
- Can produce time mean estimates of localization.
- Could build statistical models of localization from output.
- Can reduce noise, increase balance potentially.
- Can be used in concert with other a priori localizations.
- Helps when localization is not known a priori.