Flow adaptive ensemble covariance localization

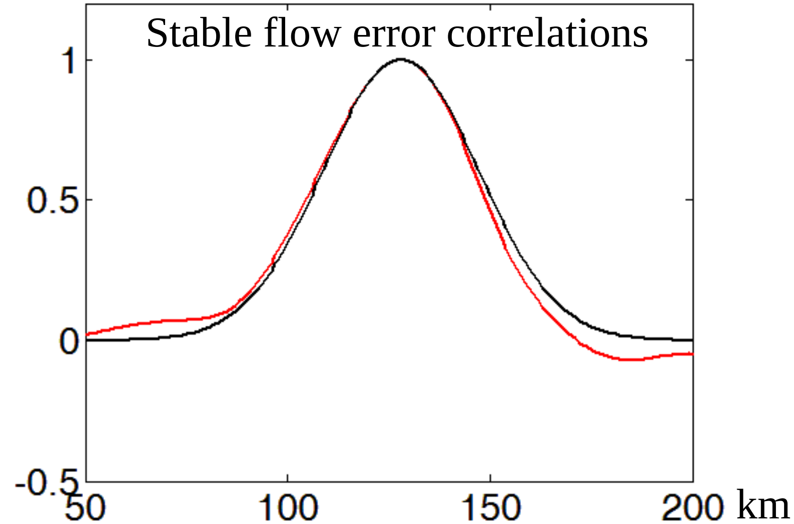
Craig H. Bishop, Daniel Hodyss
Naval Research Laboratory
Monterey, CA, USA

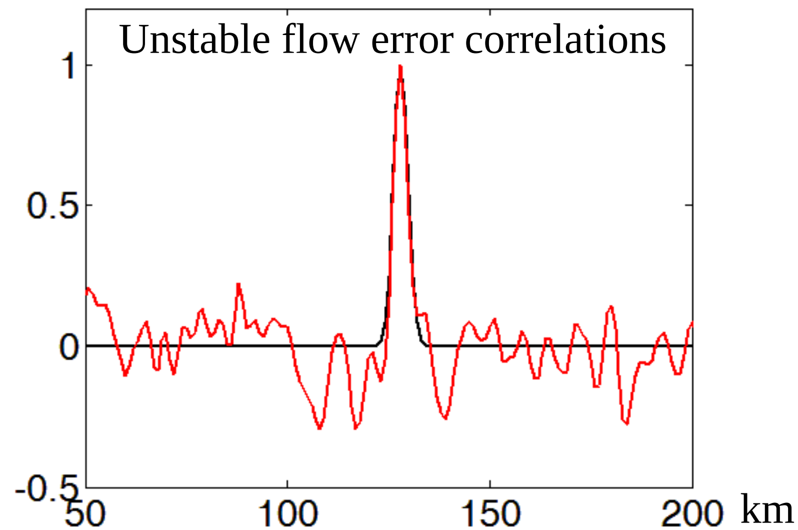
November 8, 2008

Overview

- Potential benefits of adaptive localization
- Building adaptive localizations from ensembles
- Mobile adaptive localization implies statistical TLM **M** and initial $\mathbf{P}^f_0$
- Preliminary tests of quality of **M** and $\mathbf{P}^f_0$ models given by adaptive localization
- Conclusions

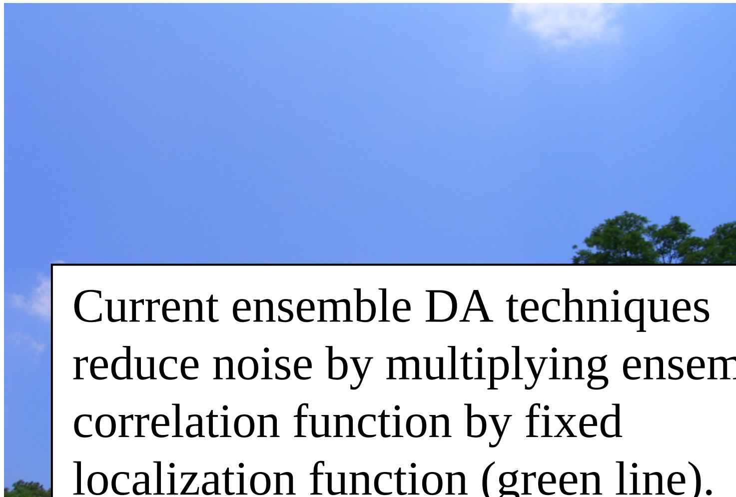
Small Ensembles and Spurious Correlations

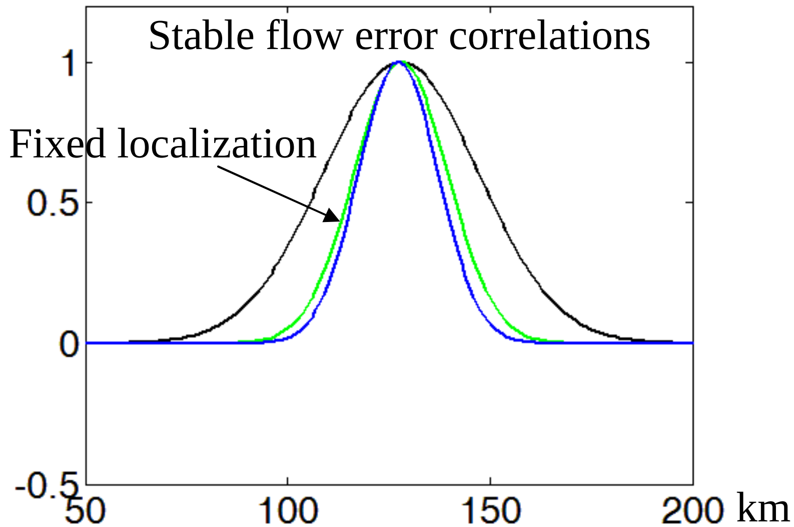




Ensembles give flow dependent, but ***noisy*** correlations

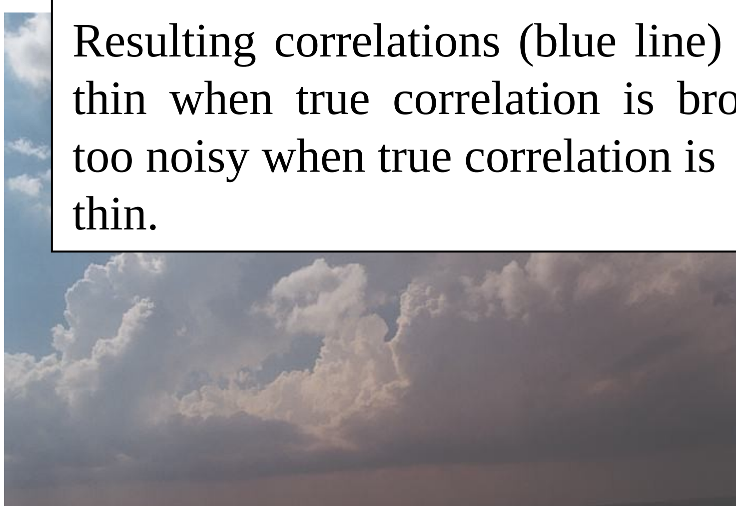
Small Ensembles and Spurious Correlations

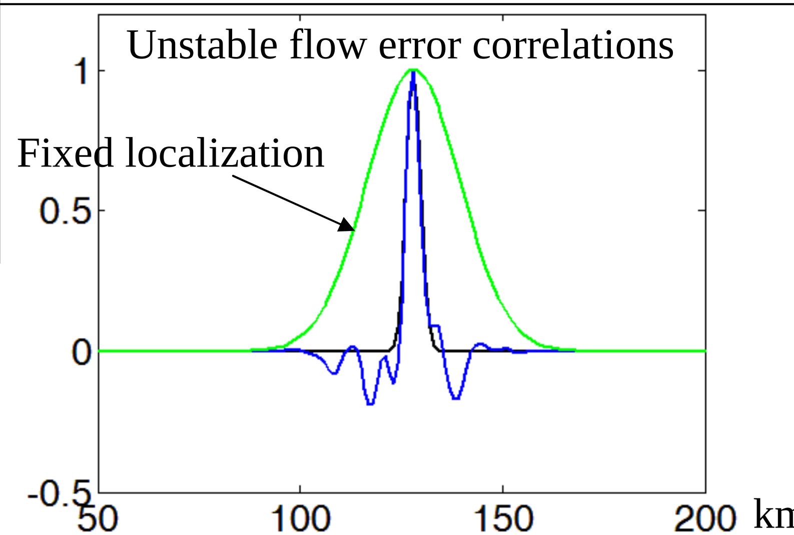




Current ensemble DA techniques reduce noise by multiplying ensemble correlation function by fixed localization function (green line).

Resulting correlations (blue line) are too thin when true correlation is broad and too noisy when true correlation is thin.

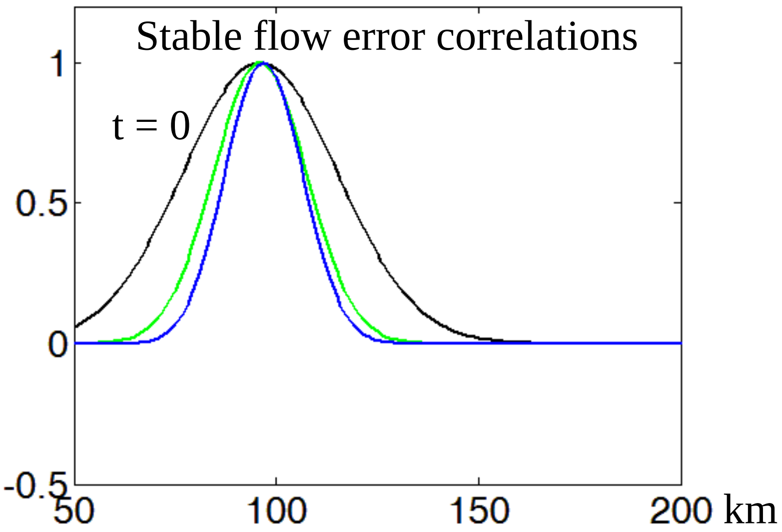


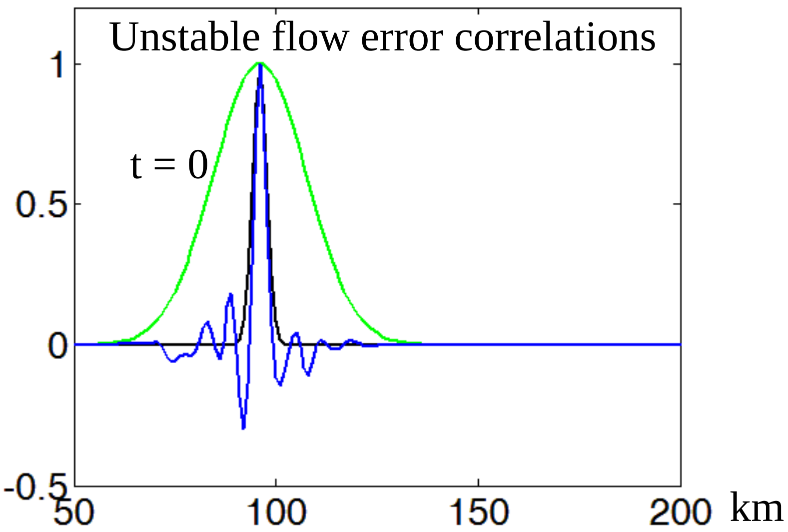


Today's fixed localization functions limit adaptivity

Small Ensembles and Spurious Correlations

- Current ensemble localization functions poorly represent propagating error correlations.

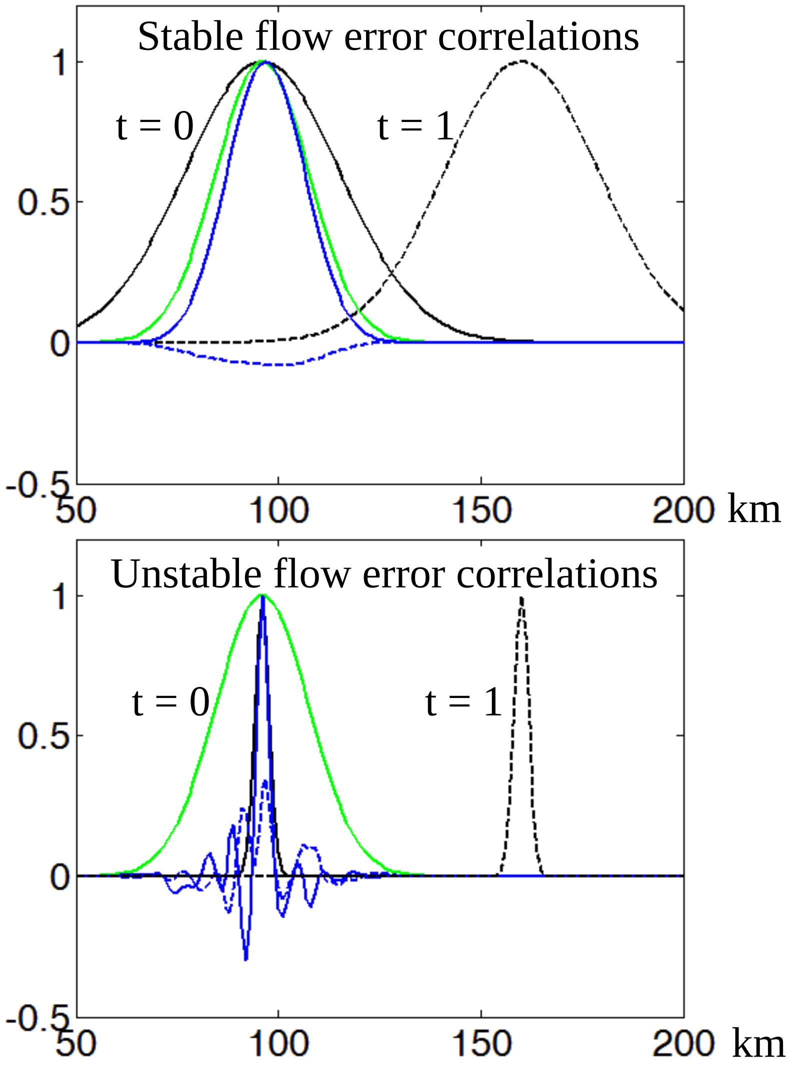




Today's fixed localization functions limit ensemble-based 4D DA

Small Ensembles and Spurious Correlations

- Current ensemble localization functions poorly represent propagating error correlations.



Today’s fixed localization functions limit ensemble-based 4D DA

Need for prior information

Given a true correlation c_{ij}^t, we can generate the pdf $\rho_L\left(c_{ij}^K \mid c_{ij}^t\right)$ of K-member sample correlations c_{ij}^K from theory. Can do this for any c_{ij}^t where $-1 \le c_{ij}^t \le 1$.

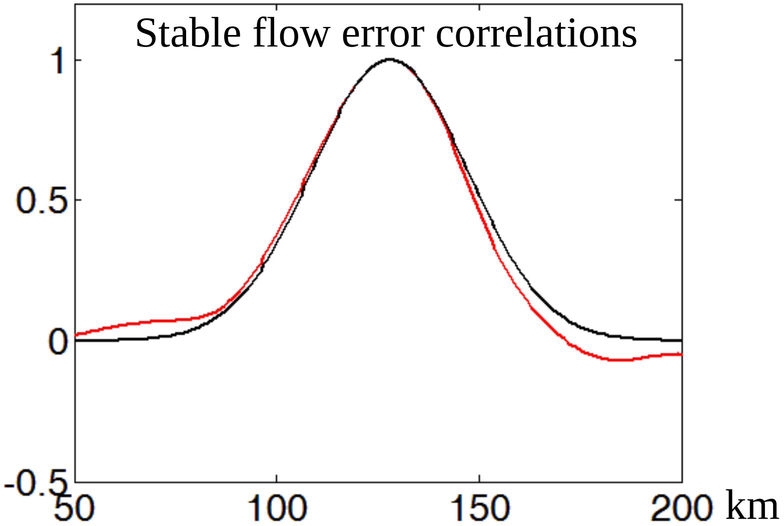
But to get the desired pdf $\rho\left(c_{ij}^t \mid c_{ij}^K\right)$ from the single realization of c_{ij}^K obtained from an ensemble, we need Bayes' theorem

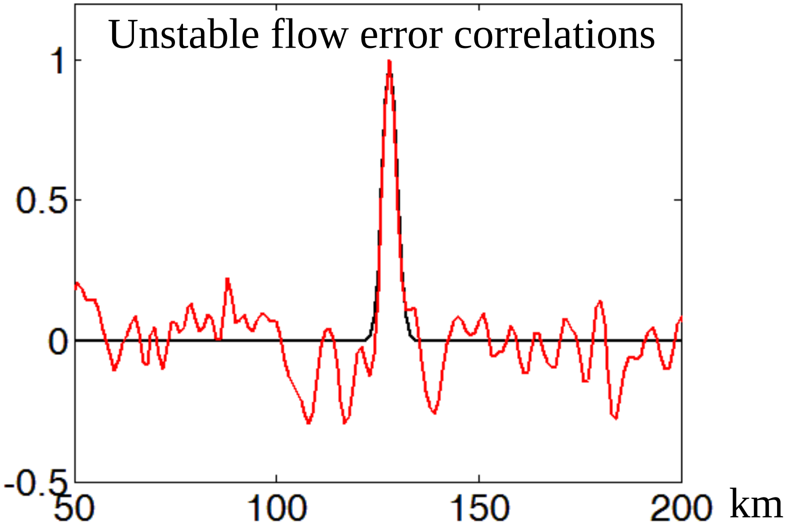
$$\rho\left(c_{ij}^t \mid c_{ij}^K\right) = \frac{\rho_L\left(c_{ij}^K \mid c_{ij}^t\right)\rho\left(c_{ij}^t\right)}{\int_{c_{ij}^t=-1}^{c_{ij}^t=1} \rho_L\left(c_{ij}^K \mid c_{ij}^t\right)\rho\left(c_{ij}^t\right) d\left(c_{ij}^t\right)}$$

where $\rho\left(c_{ij}^t\right)$ gives pdf of true correlations based on *prior* information.

Localization can be viewed as the process of guessing the mean of the posterior $\rho\left(c_{ij}^t \mid c_{ij}^K\right)$ based on a guess of the prior $\rho\left(c_{ij}^t\right)$.

Small Ensembles and Spurious Correlations





Ensembles give flow dependent, but ***noisy*** correlations

Prior correlation distribution might be a function of:

Non-adaptive localization

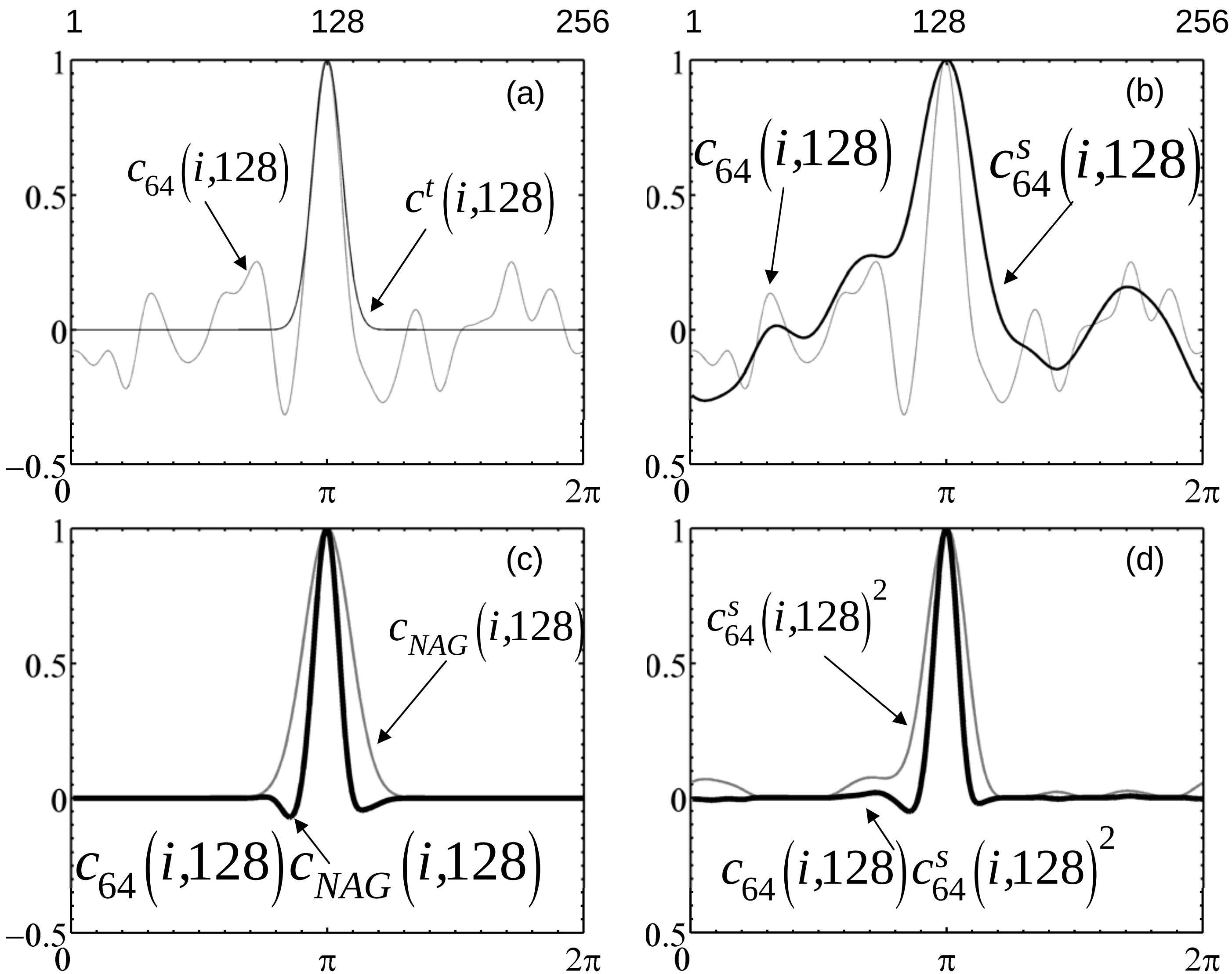
- distance from sample correlation of 1
- latitude, longitude, height
- Rossby radius, Richardson number
- estimated geostrophic coupling (for multi-variate case)
- estimated group velocity of errors (for variables separated in time)
- anisotropy of nearby features (e.g. fronts)

Adaptive localization

How about trying to extract this info from the ensemble?

Smoothed ENsemble COrrelations Raised to a Power (SENCORP)

(Bishop and Hodyss, 2007, QJRMS)



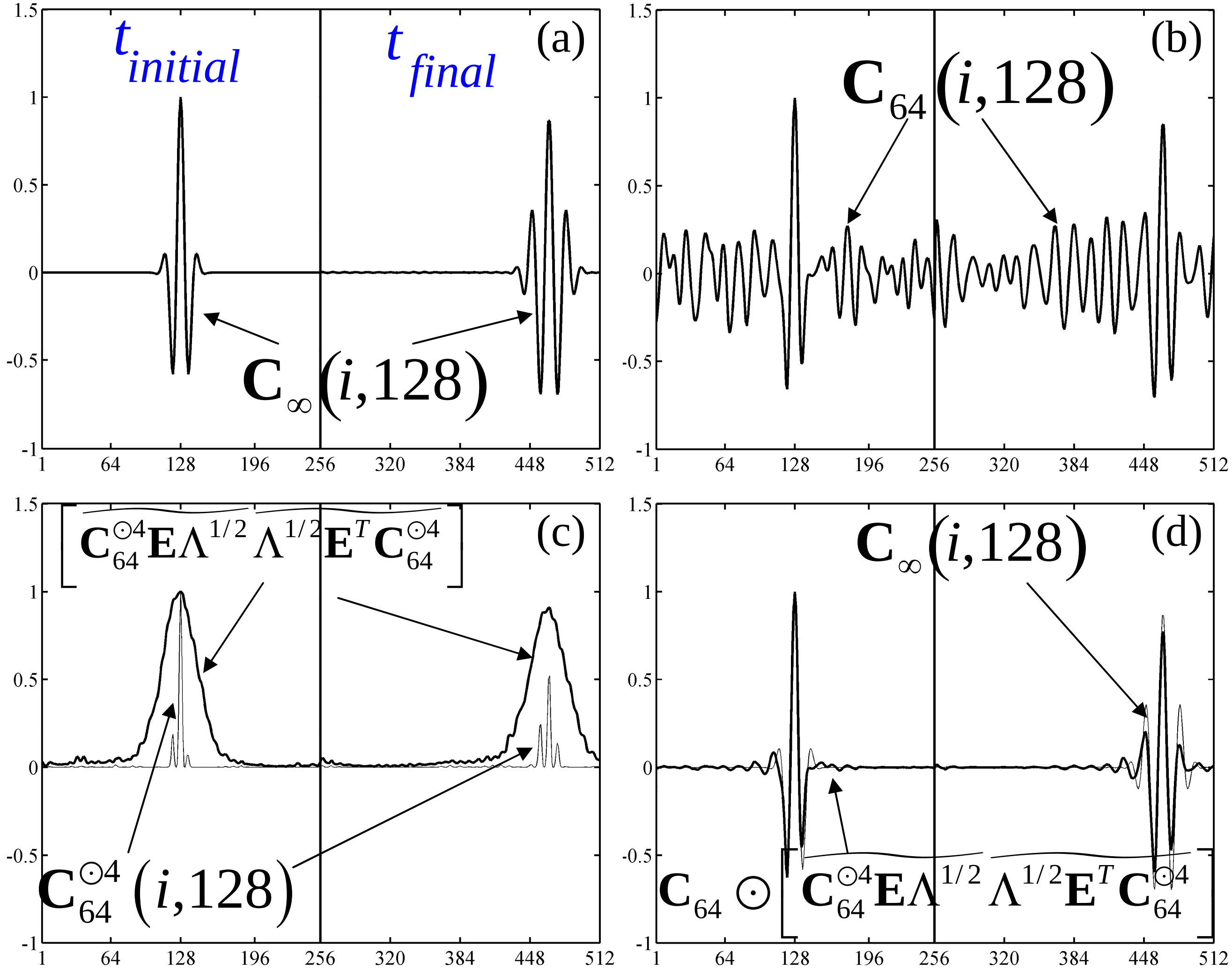
Bishop and Hodyss, 2007, QJRMS.

In matrix form,

$$\mathbf{P}^f = \mathbf{P}_K^f \odot \mathbf{C}_s \odot \mathbf{C}_s$$

Ensemble COrrelations Raised to A Power (ECO-RAP)

Bishop and Hodyss, 2008ab, Tellus



K = 64 member ensemble

$\mathbf{C}_K$ = Ensemble correlation matrix

$\mathbf{C}_K^{\odot n}$ = n elementwise products of ensemble correlation

$\mathbf{E}\Lambda\mathbf{E}^T$ = Non-adaptive localization matrix

Mobile adaptive localization provides a statistical TLM/PFM

$$\mathbf{x}^a(18)-\mathbf{x}^f(18)=M\left[\mathbf{x}^a(6)\right]-M\left[\mathbf{x}^f(6)\right]\approx\mathbf{M}\left[\mathbf{x}^a(6)-\mathbf{x}^f(6)\right]$$

Statistically, the Best Linear Unbiased Estimate of $\mathbf{M}$ is

$$\mathbf{M}=\left\langle\left[\mathbf{x}^a(18)-\mathbf{x}^f(18)\right]\left[\mathbf{x}^a(6)-\mathbf{x}^f(6)\right]^T\right\rangle\left\langle\left[\mathbf{x}^a(6)-\mathbf{x}^f(6)\right]\left[\mathbf{x}^a(6)-\mathbf{x}^f(6)\right]^T\right\rangle^{-1}$$

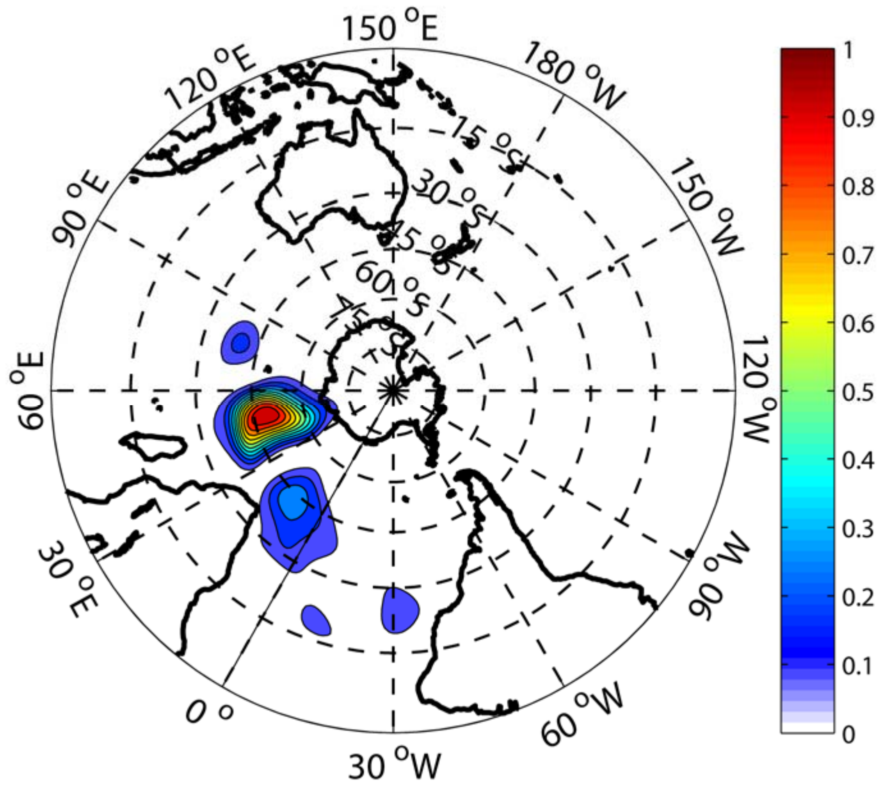
$$\approx\left\langle\varepsilon_{18}^f\varepsilon_6^{fT}\right\rangle\left\langle\varepsilon_6^f\varepsilon_6^{fT}\right\rangle^{-1}$$

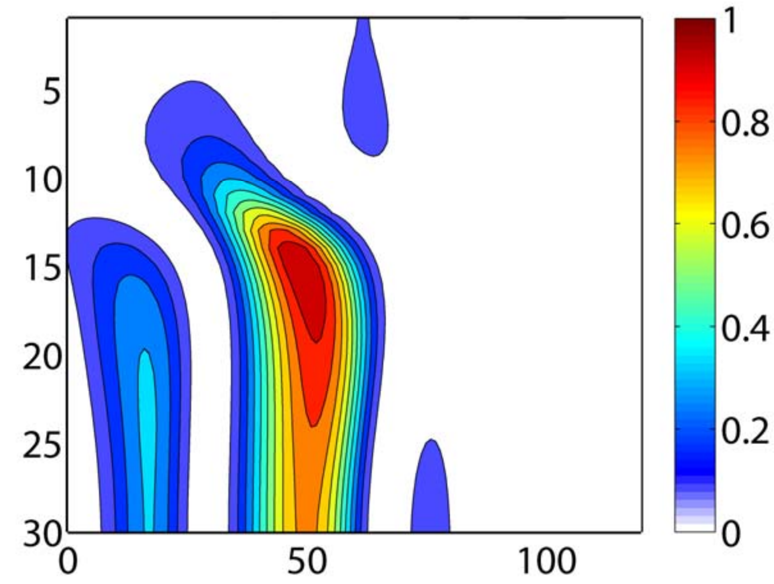
and hence $\mathbf{M}$ is readily derived from any four-dimensional $\underline{\mathbf{P}}^f$.

(Compare with "statistical 4D-VAR" TLM/PFM discussed

in Lorenc and Payne, 2007, QJRMS)

Example of a column of the localization $\underline{\mathbf{C}}_s \odot \underline{\mathbf{C}}_s$ with $K = 128$ 06Z

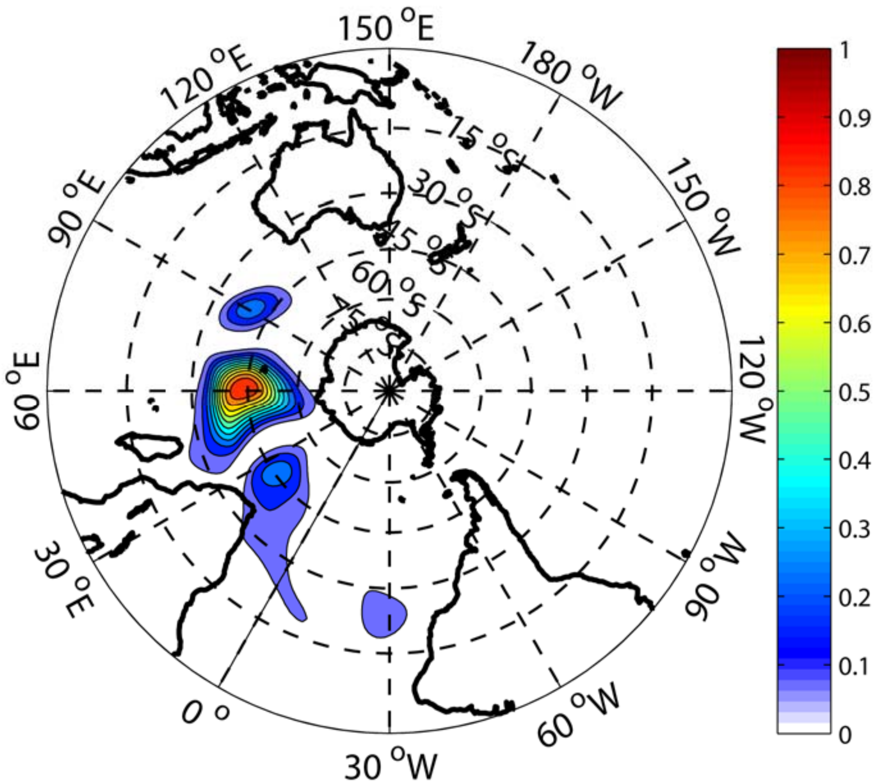


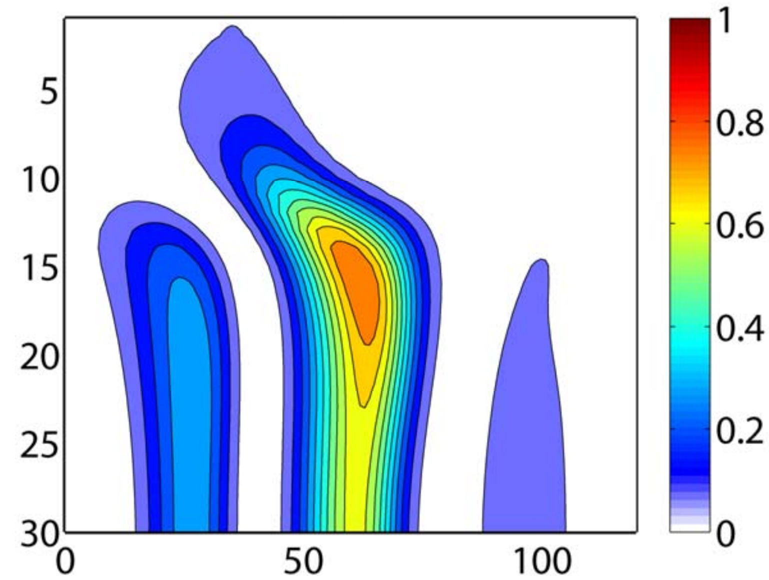


Ensemble based localization moves about 1000 km in 12 hrs. This is >=half-width of a typical LETKF observation volume (~900km).

Example of a column of the localization $\underline{\mathbf{C}}_s \odot \underline{\mathbf{C}}_s$ with $K = 128$

18Z

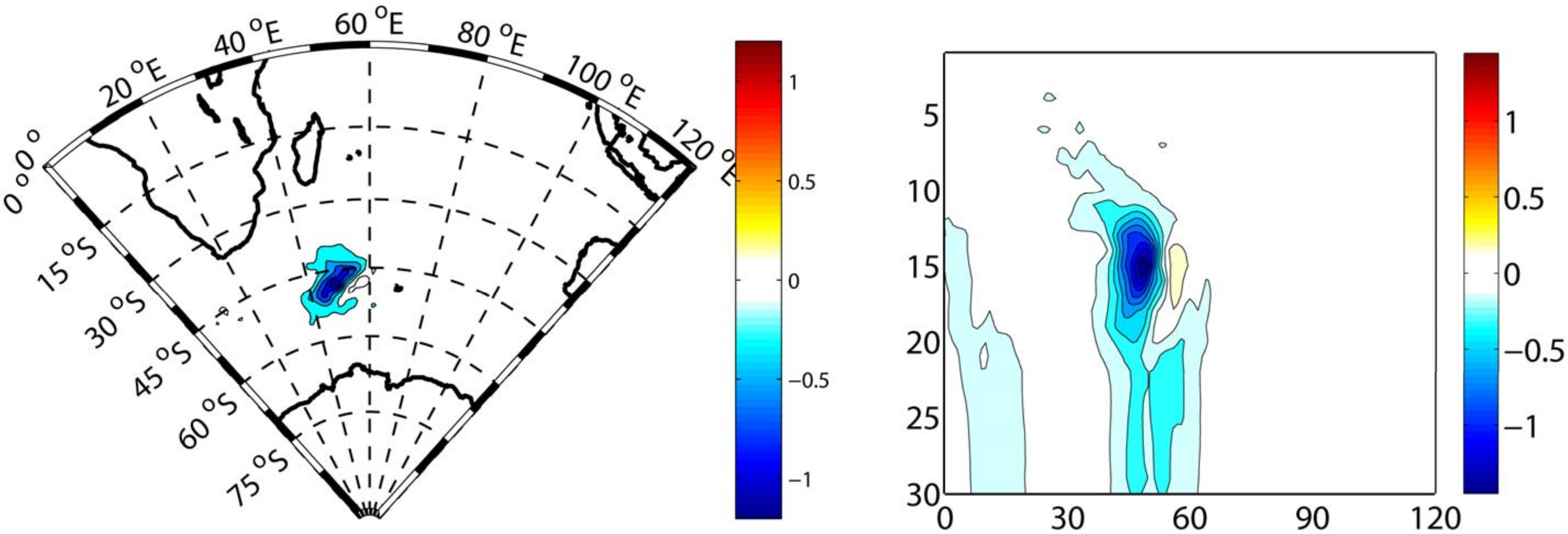




Ensemble based localization moves about 1000 km in 12 hrs. This is >=half-width of a typical LETKF observation volume (~900km).

Example of a column of $\underline{\mathbf{P}}_K^f \odot \underline{\mathbf{C}}_s \odot \underline{\mathbf{C}}_s$ with $K = 128$

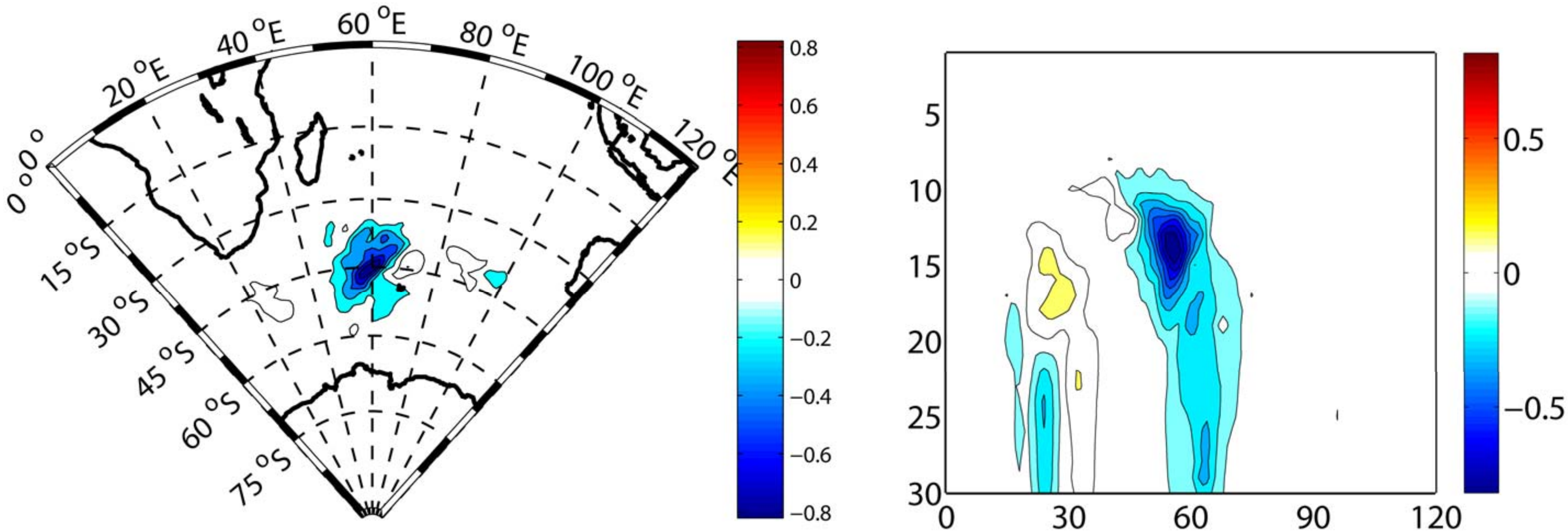
<*vv*> Increment 06Z



Statistical TLM implied by mobile adaptively localized covariance propagates single observation increment 1000 km in 12 hrs.

Example of a column of $\underline{\mathbf{P}}_K^f \odot \underline{\mathbf{C}}_s \odot \underline{\mathbf{C}}_s$ with $K = 128$

<*vv*> Increment 18Z



Statistical TLM implied by mobile adaptively localized covariance propagates single observation increment 1000 km in 12 hrs.

Mobile adaptive localization provides an initial time covariance $\mathbf{P}^f_0$ and a statistical **M**. Can the quality of such models of $\mathbf{P}^f_0$ and **M** match the quality of their non-ensemble counterparts?

Experiment 1: Test of Initial adaptively localized covariance $\mathbf{P}^f_0$

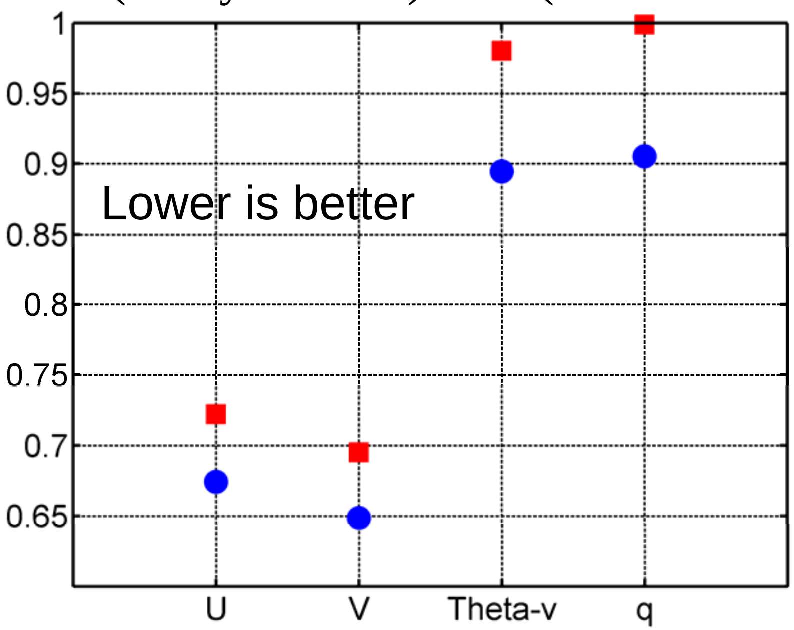
- Forecast error is difference between a T119L30 6Z-18Z forecast (the first guess) and a T119L30 30-42 hr forecast valid at the same time (the simulated truth).
- In this case, 440,000 evenly spaced obs of u,v, and T are simulated. (Every 3rd level in z, every 2nd grid point in x and y, no obs poleward of 80). (8 million variables).
- All obs are taken at 6Z (the beginning of DA window)
- Rms ob error for u and v is 2 ms^-1
- Rms ob error for T is 2 K
- 128 Ensemble Transform (ET) ensemble members=>localized ensemble has rank <1,056,768
- Compare $\underline{\mathbf{P}}^f = \underline{\mathbf{P}}_K^f \odot \underline{\mathbf{C}}_s \odot \underline{\mathbf{C}}_s$ with NAVDAS. Arguably, NAVDAS has the advantage because forecast error is the result of 4 NAVDAS analysis corrections (using real obs) made over the preceding 24 hrs.

Comparison of NAVDAS and $\underline{\mathbf{P}}^f = \underline{\mathbf{P}}_K^f \odot \underline{\mathbf{C}}_s \odot \underline{\mathbf{C}}_s$ analysis error

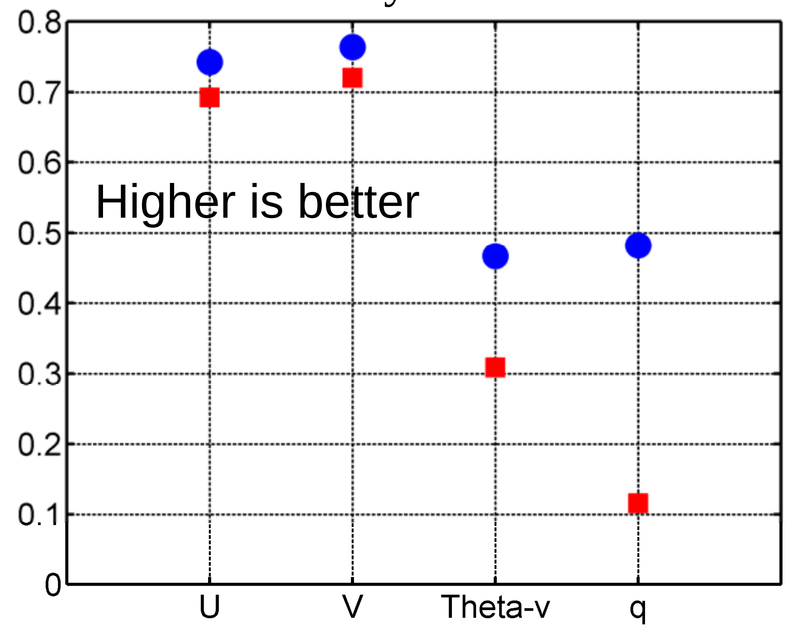
Red square → NAVDAS

Blue Circle → $\underline{\mathbf{P}}^f = \underline{\mathbf{P}}_K^f \odot \underline{\mathbf{C}}_s \odot \underline{\mathbf{C}}_s$

RMS(Analysis Error)/RMS(Forecast Error)



Anomaly Correlation

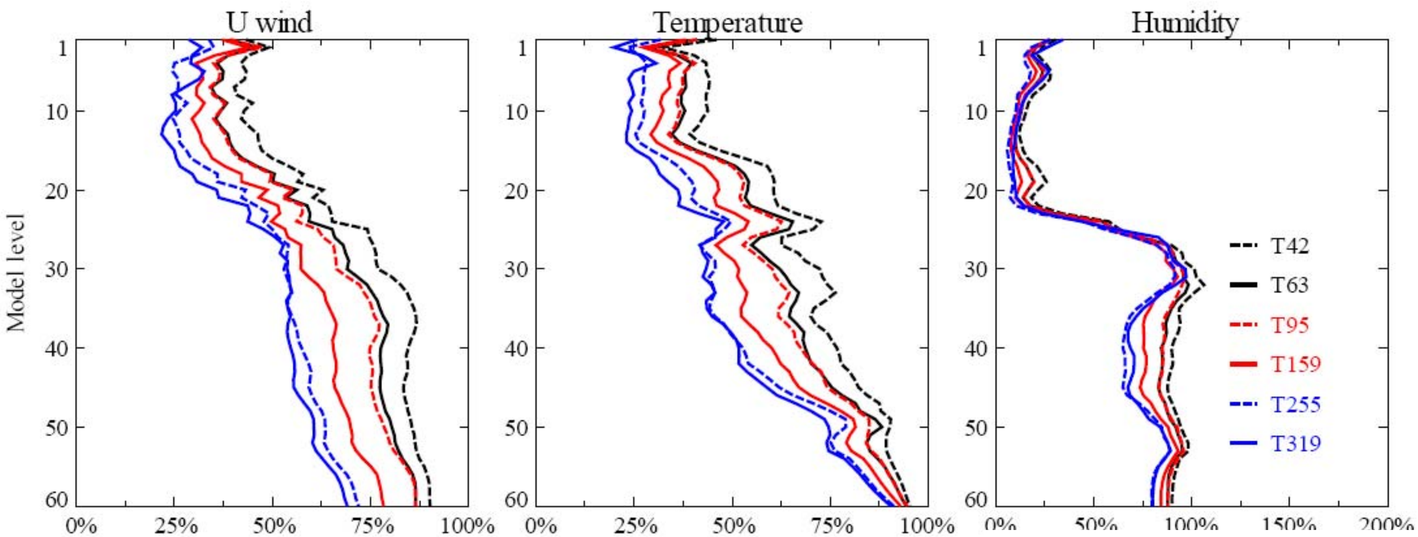


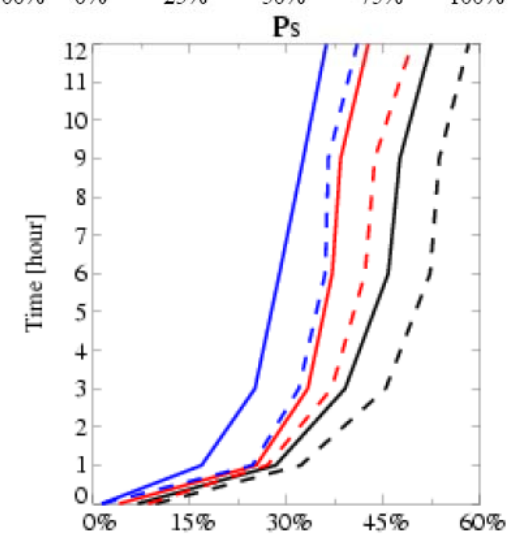
Anomaly correlation is between analysis correction and the “perfect” correction that would have eliminated all initial condition error.

Adaptively localized ensemble covariance produced smaller initial condition errors than covariance model used in operational 3D-PSAS/NAVDAS scheme

P^f_0 is OK. What about the TLM implied by 4D $\underline{\underline{P}}^f = \underline{\underline{P}}^f_K \odot \underline{\underline{C}}_s \odot \underline{\underline{C}}_s$?

$$r = \frac{\left\| M(x_i + S^{-1}\delta x_i) - M(x_i) - \mathbf{M}_i(\delta x_i) \right\|}{\left\| M(x_i + S^{-1}\delta x_i) - M(x_i) \right\|} = \frac{noise}{signal} = \frac{n}{s}$$





Will the 4D-Var approach be defeated by nonlinearity?

E. Andersson, M. Fisher, E. Hólm, L. Isaksen, G. Radnóti, and Y. Trémolet

Figure 3: Relative error of the tangent linear model for various resolutions with respect to T511 nonlinear, diabatic model after 12h for the 3 dimensional variables and for the whole 4D-Var window for surface pressure. Diagnostics are computed on the T255 resolution grid. From Radnóti et al. (2005).

Summary of rough TLM comparison (see conference web-site for more detail)

r-values for statistical TLM implied by $\underline{\mathbf{P}}^{f} = \underline{\mathbf{P}}_{K}^{f} \odot \underline{\mathbf{C}}_{s} \odot \underline{\mathbf{C}}_{s}$ smaller than those for ECMWF TLM at low levels but are larger at upper levels. Advantage of statistical TLM more apparent in specific humidity and temperature than in wind.

Comparison is not clean because amplitude of our correction was nearly twice as large as ECMWF correction and ECMWF non-linear model is at higher resolution than TLM whereas our non-linear model is same resolution as TLM.

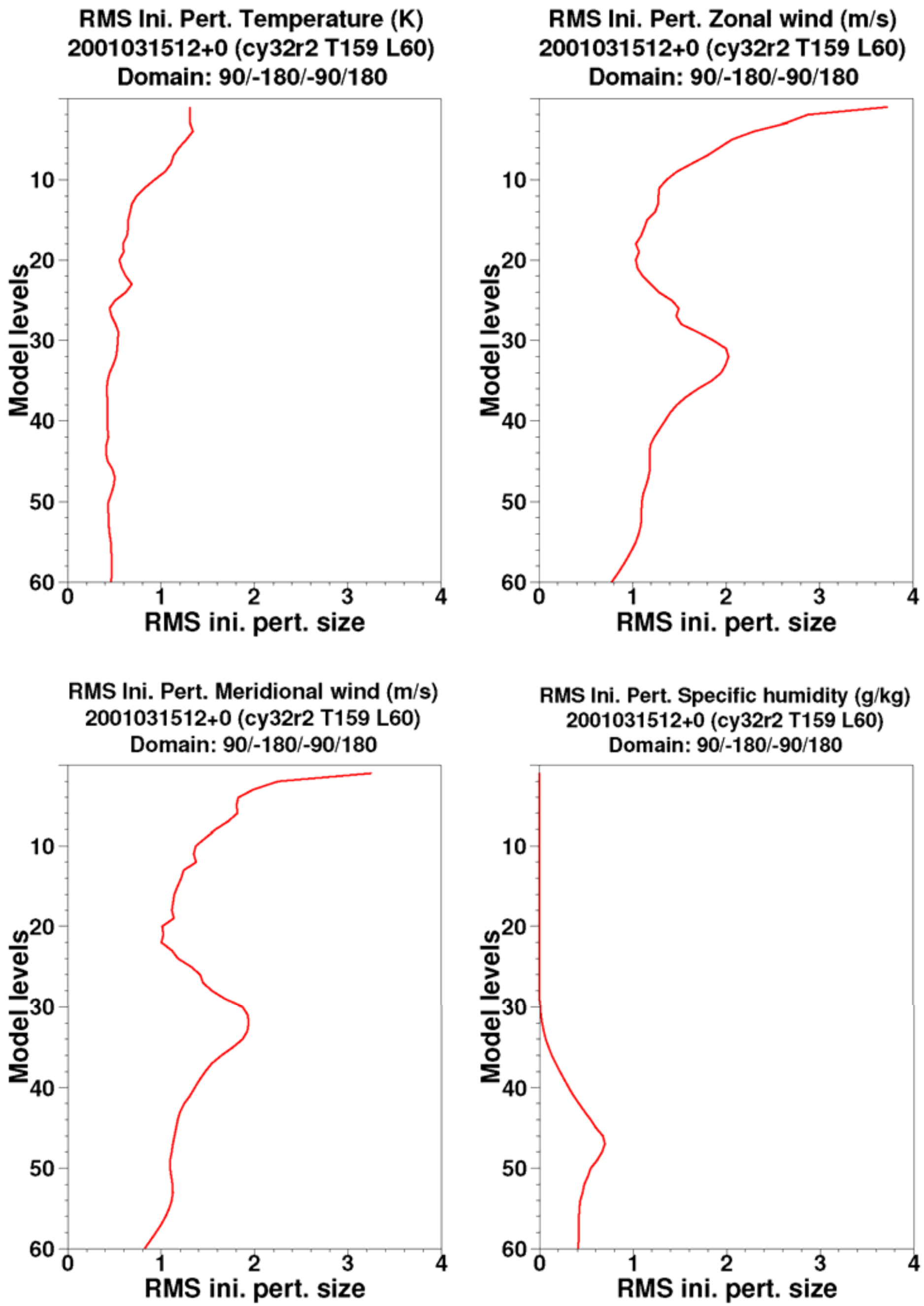
Comparison with NOGAPS TLM indicates much higher level of superiority for statistical TLM but comparison was not clean beacause NOGAPS TLM test was at higher resolution (T239L30) than statistical TLM test (T119L30).

Conclusions

- Useful adaptive localization functions can be built directly from the ensemble.
- To the extent that error covariance functions vary in space, adaptive localization improves the accuracy of covariance estimates.
- Useful statistical TLM's can be derived from *mobile* adaptive localization.
- Preliminary tests suggest that the quality of models of $\mathbf{P}^f_0$ and **M** based on adaptively localized covariances is not significantly worse than the quality of their non-ensemble counterparts.
- Global-4D-VAR with statistical TLM/PFM is enabled by application of square root theorem to adaptive localization (see conference web-site).

Questions

- Hi-res 4D-VAR can lead to the need to model covariances of errors in both the linear and non-linear regime. Might a statistical TLM enabled by flow-adaptive ensemble covariance localization provide a better TLM for this problem than a conventional TLM?
- How well would the outer loop of 4D-VAR converge if our statistical TLM was used instead of a conventional TLM/PFM?
- Would the ensemble based statistical TLM have an advantage for weak constraint 4D-VAR because of the ease with which the effects of model error could be incorporated in an ensemble?
- Best method for initialization of ensemble?



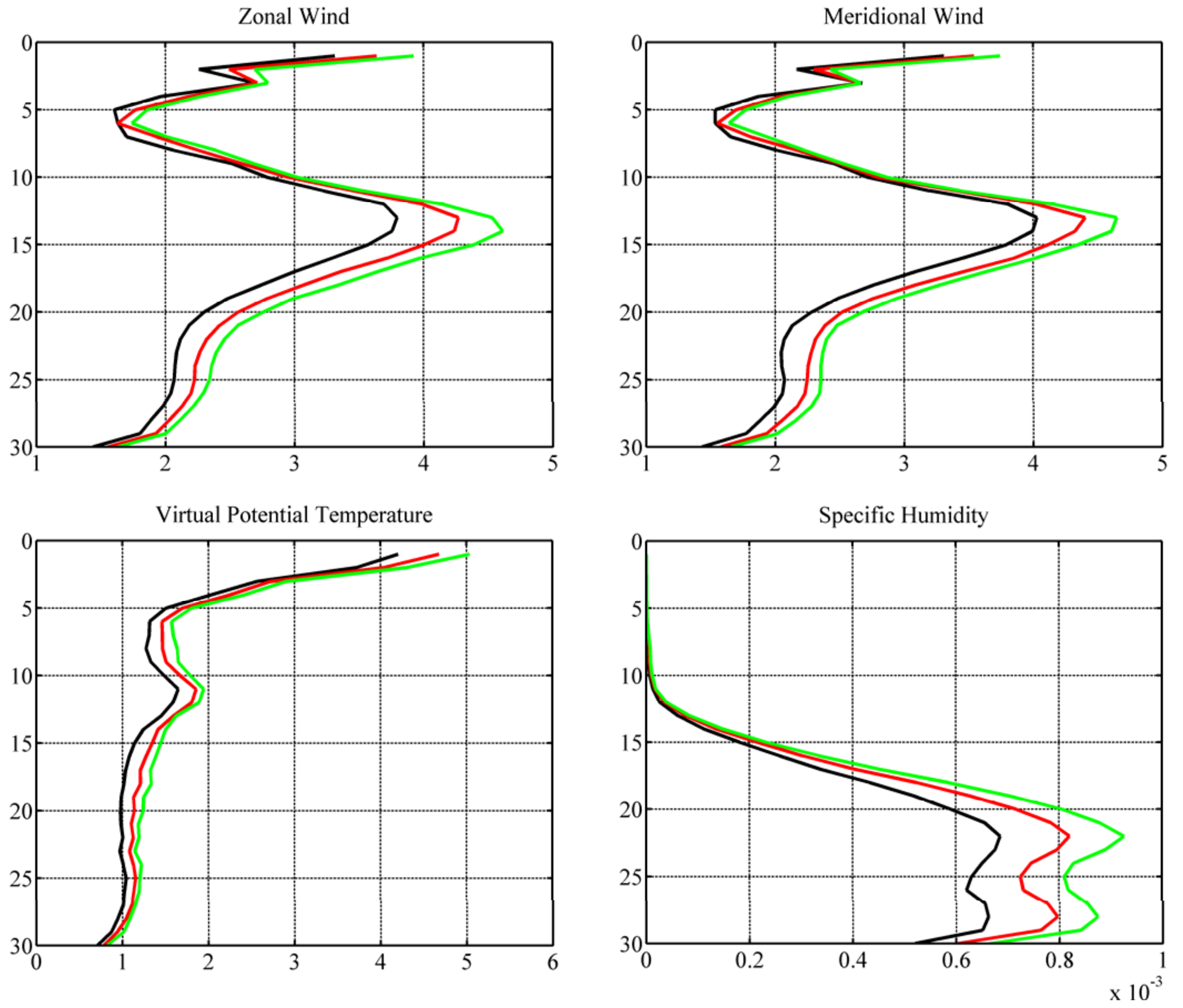
RMS magnitude of EC analysis correction is shown on left.

TLMs ignore non-linear terms.

Non-linear terms increase as size of analysis correction increases

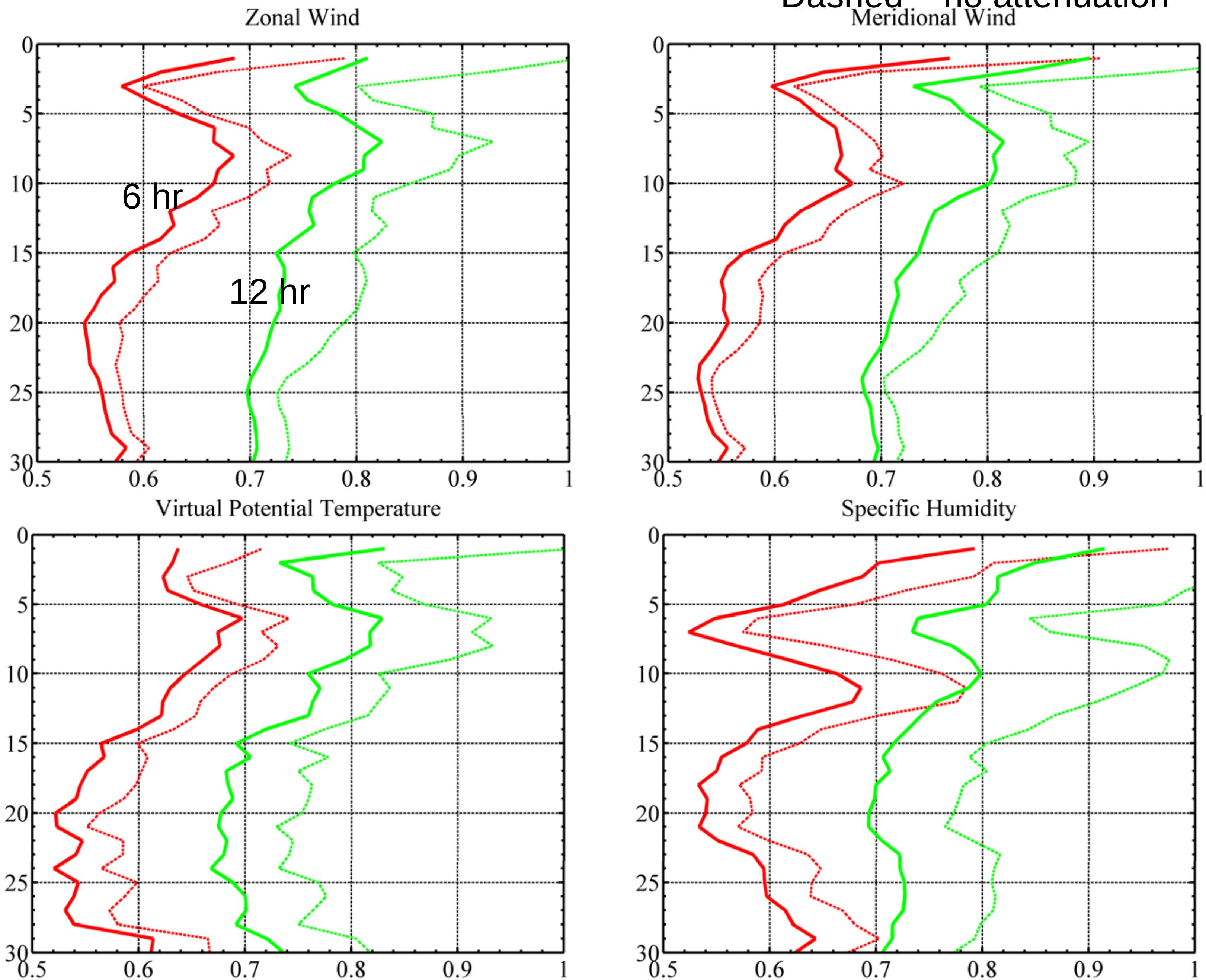
(Thanks to Philippe Lopez, ECMWF, for providing these figures)

RMS Amplitude of our correction [almost 2x(ECMWF correction)]



TLM implied by $\mathbf{P}^{f} = \mathbf{P}_{K}^{f} \odot \mathbf{C}_{s} \odot \mathbf{C}_{s}$ "r" Parameter

Solid – attenuation
Dashed – no attenuation



Comparison of TLM quality at surface

	ECMWF r=n/s	DAMES r=n/s	ECMWF size	DAMES size
u	0.68	0.7	2 m/s	3.75 m/s
T/theta_v	0.91	0.74	0.5 K	1 K
specific humidity	0.8	0.71	0.4 g/kg	0.5 g/kg

Square root theorem provides efficient representation (Bishop and Hodyss, 2008b, Tellus)

Raw covariance $\underline{\mathbf{P}}_K^f = \sum_{k=1}^{K} \underline{\mathbf{z}}_k \underline{\mathbf{z}}_k^T$, smooth ens correlation $\underline{\mathbf{C}}^s = \sum_{j=1}^{K} \underline{\mathbf{z}}_j^s \underline{\mathbf{z}}_j^{sT}$, hence

$$\left(\underline{\mathbf{P}}_K^f \odot \underline{\mathbf{C}}^s \odot \underline{\mathbf{C}}^s\right)_{mn} = \sum_{k=1}^{K} z_{mk} z_{nk} \sum_{j=1}^{K} z_{mj}^s z_{nj}^s \sum_{i=1}^{K} z_{mi}^s z_{ni}^s = \sum_{k=1}^{K} \sum_{j=1}^{K} \sum_{i=1}^{K} \left(z_{mk} z_{mj}^s z_{mi}^s\right)\left(z_{nk} z_{nj}^s z_{ni}^s\right)$$

Hence,

$$\underline{\mathbf{P}}_K^f \odot \underline{\mathbf{C}}^s \odot \underline{\mathbf{C}}^s = \sum_{k=1}^{K} \sum_{j=1}^{K} \sum_{i=1}^{K} \underbrace{\left(\underline{\mathbf{z}}_k \odot \underline{\mathbf{z}}_j^s \odot \underline{\mathbf{z}}_i^s\right)}\left(\underline{\mathbf{z}}_k \odot \underline{\mathbf{z}}_j^s \odot \underline{\mathbf{z}}_i^s\right)^T = \underbrace{\underline{\mathbf{Z}}_D}\, \underline{\mathbf{Z}}_D^T$$

Modulated ensemble

Modulated ensemble member

Root is huge ensemble of modulated ensemble members!

***K*²(*K*+1)/2 can be linearly independent (*K=128* => a possible 1,056,768 linear independent members)**

Data Assimilation using Modulated EnsembleS (DAMES)

Both incremental and non-incremental 4D-VAR are possible.

Non-incremental weak constraint 4D-VAR is as follows.

Maximal liklihood state $\mathbf{x}^a$ is the $\mathbf{x}$ value that minimizes

$$J(\underline{\mathbf{x}}) = \frac{1}{2}\left\{\left[\underline{\mathbf{y}} - \underline{H}(\underline{\mathbf{x}})\right]^T \underline{\mathbf{R}}^{-1}\left[\underline{\mathbf{y}} - \underline{H}(\underline{\mathbf{x}})\right] + \left(\underline{\mathbf{x}} - \underline{\mathbf{x}}^f\right)^T \left(\underline{\mathbf{P}}^f\right)^{-1}\left(\underline{\mathbf{x}} - \underline{\mathbf{x}}^f\right)\right\}$$

where $\underline{\mathbf{P}}^f = \underline{\mathbf{Z}}_D \underline{\mathbf{Z}}_D^T$ is a 4-dimensional forecast error covariance matrix and $\underline{\mathbf{Z}}_D$ is its square root. Each column of $\underline{\mathbf{Z}}_D$ is a concatenation of ensemble perts at different times.

Letting $\left(\underline{\mathbf{x}} - \underline{\mathbf{x}}^f\right) = \underline{\mathbf{Z}}_D \mathbf{a}$ and $\underline{\mathbf{y}}' = \left[\underline{\mathbf{y}} - \underline{H}\left(\underline{\mathbf{x}}^f\right)\right]$, J is minimized when

$$\frac{\partial J}{\partial \mathbf{a}} = -\left[\underline{\mathbf{H}}\underline{\mathbf{Z}}_D\right]^T \underline{\mathbf{R}}^{-1}\left[\underline{\mathbf{y}}'\right] + \left\{\left[\underline{\mathbf{H}}\underline{\mathbf{Z}}_D\right]^T \underline{\mathbf{R}}^{-1}\left[\underline{\mathbf{H}}\underline{\mathbf{Z}}_D\right] + \mathbf{I}\right\}\mathbf{a}$$

$$= -\left[\underline{\mathbf{R}}^{-1/2}\underline{\mathbf{H}}\underline{\mathbf{Z}}_D\right]^T \left[\underline{\mathbf{R}}^{-1/2}\underline{\mathbf{y}}'\right] + \left\{\left[\underline{\mathbf{R}}^{-1/2}\underline{\mathbf{H}}\underline{\mathbf{Z}}_D\right]^T \left[\underline{\mathbf{R}}^{-1/2}\underline{\mathbf{H}}\underline{\mathbf{Z}}_D\right] + \mathbf{I}\right\}\mathbf{a} = 0. \text{ Hence}$$

$$\underline{\mathbf{x}}^a - \underline{\mathbf{x}}^f = \underline{\mathbf{Z}}_D \left\{\left[\underline{\mathbf{R}}^{-1/2}\underline{\mathbf{H}}\underline{\mathbf{Z}}_D\right]^T \left[\underline{\mathbf{R}}^{-1/2}\underline{\mathbf{H}}\underline{\mathbf{Z}}_D\right] + \mathbf{I}\right\}^{-1} \left[\underline{\mathbf{R}}^{-1/2}\underline{\mathbf{H}}\underline{\mathbf{Z}}_D\right]^T \left[\underline{\mathbf{R}}^{-1/2}\left[\underline{\mathbf{y}} - H\left(\underline{\mathbf{x}}^f\right)\right]\right]$$

Observation space form is also straightforward (El Akkraoui et al., 2008, QJRMS) .