ECO-RAP
A new adaptive error covariance localization tool for 4-dimensional ensemble data assimilation

Craig H. Bishop, Daniel Hodyss,

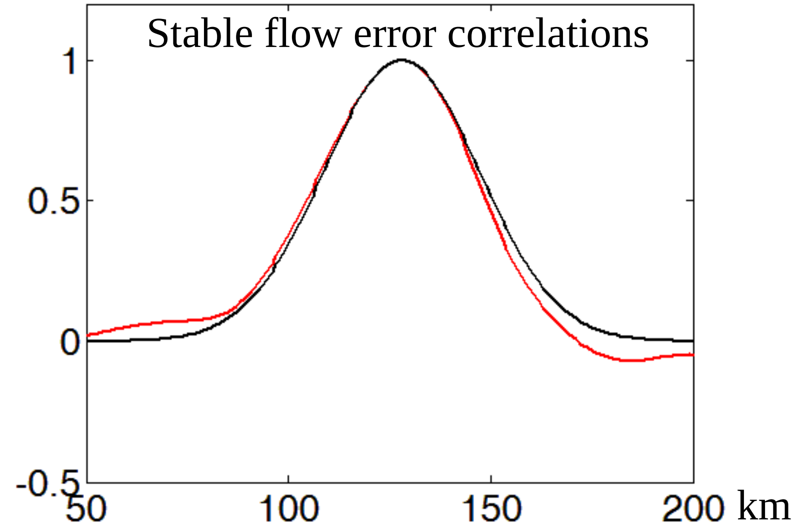
William. F. Campbell, and Justin G. Mclay

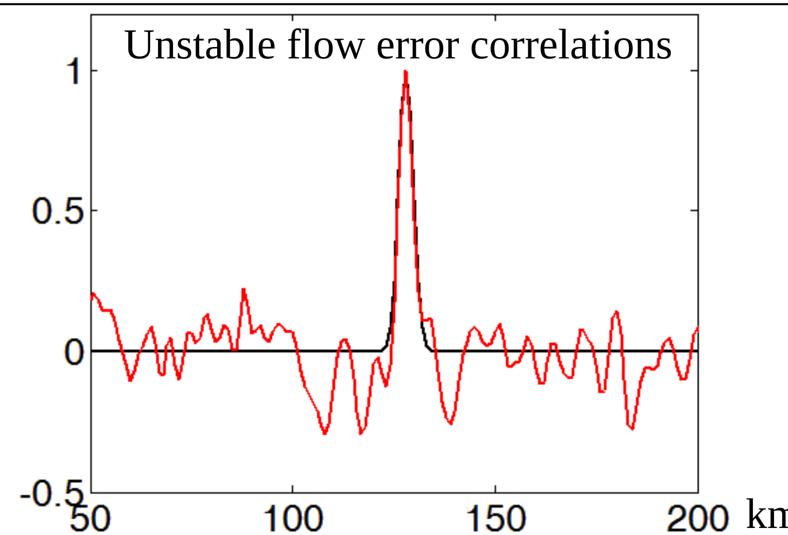
Naval Research Laboratory, Monterey, California

Outline

- Motivation
- How ECO-RAP works
- Idealized tests
- Review
- Computational considerations/speed-up
- Preliminary experiment with NWP model
- Conclusions

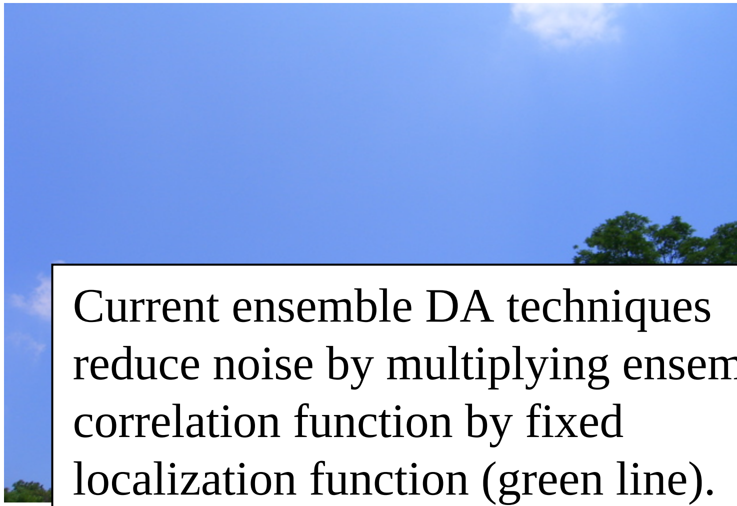
Small Ensembles and Spurious Correlations

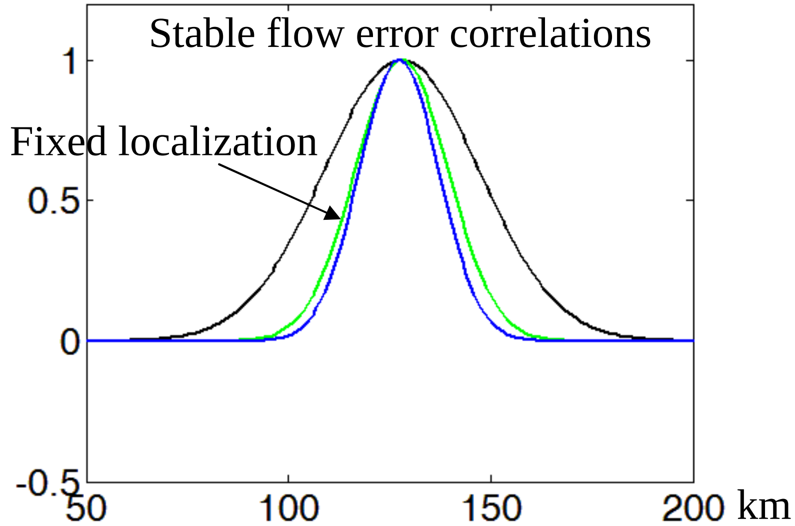




Ensembles give flow dependent, but ***noisy*** correlations

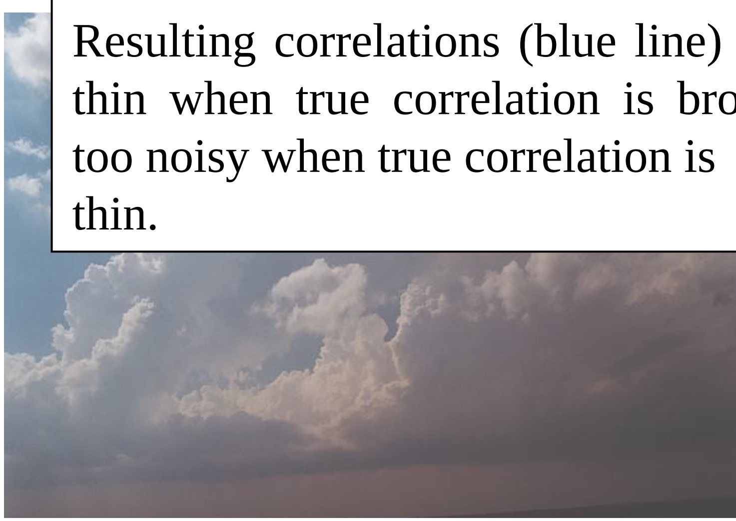
Small Ensembles and Spurious Correlations

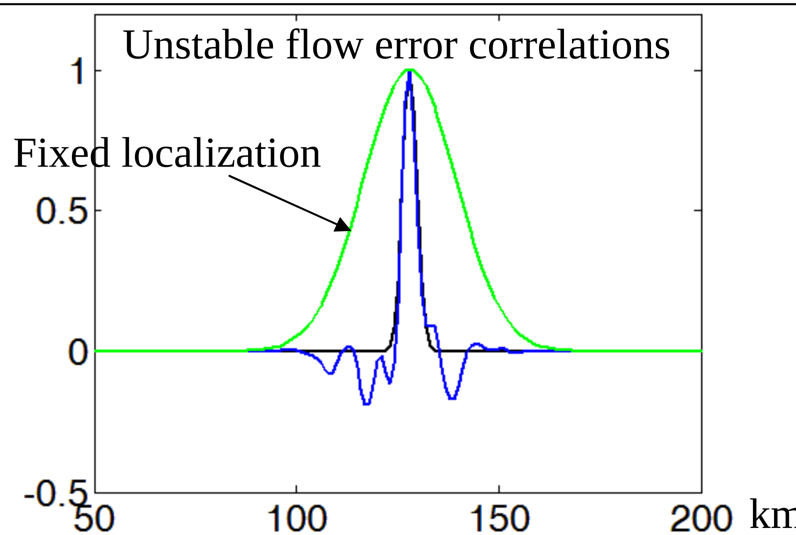




Current ensemble DA techniques reduce noise by multiplying ensemble correlation function by fixed localization function (green line).

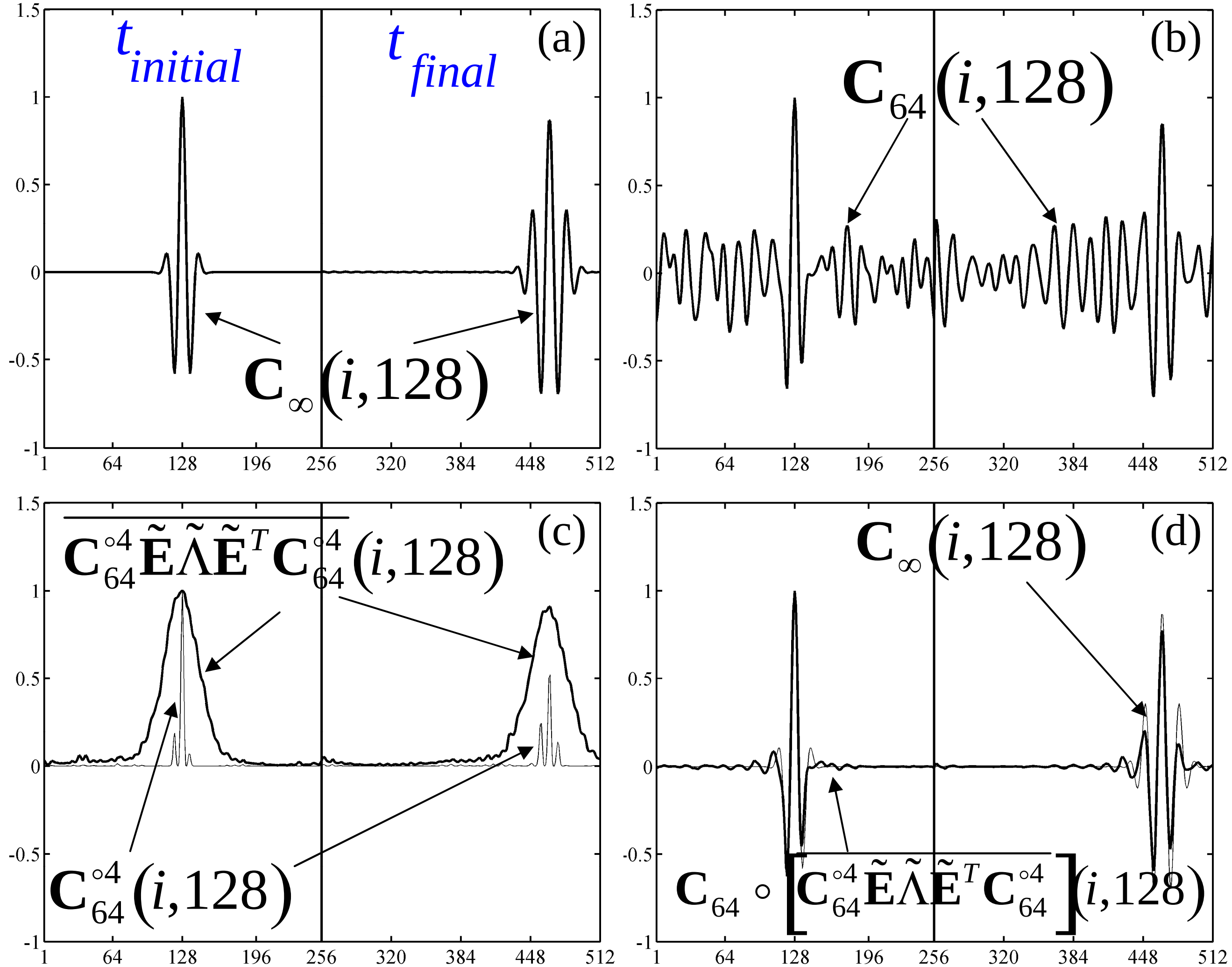
Resulting correlations (blue line) are too thin when true correlation is broad and too noisy when true correlation is thin.





Today's fixed localization functions limit adaptivity

CALECO



$K = 64$ member ensemble

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

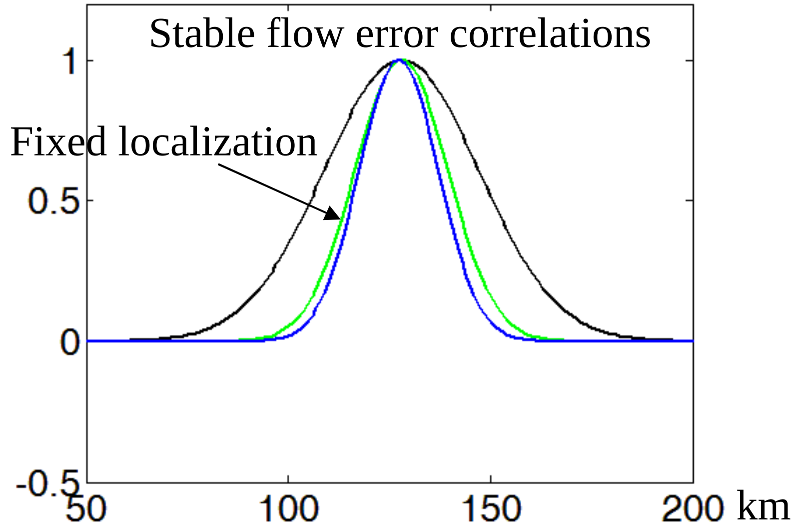
$\mathbf{C}_K^{\circ n} =$ n elementwise products of ensemble correlation

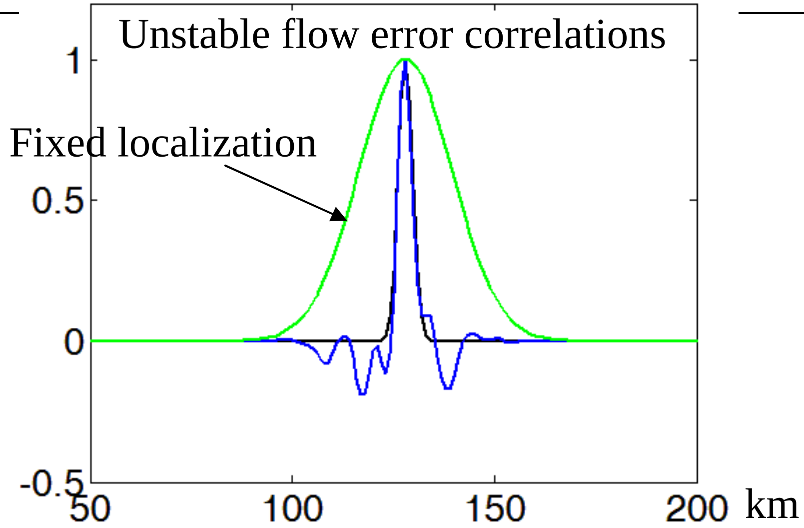
$\tilde{\mathbf{E}}\tilde{\Lambda}\tilde{\mathbf{E}}^{T} =$ Non-adaptive localization matrix

ECO-RAP: Ensemble COrrelations RAised to a Power

- Ensemble correlations contain propagation and length scale information.
- Ensemble correlations corresponding to large true correlations are bigger than those corresponding to true zero correlations. (Variance of spurious is 1/K).
- Raising ensemble correlations to a power attenuates small values more than large values.
- Sandwiching *non-adaptive* localization matrix between ensemble correlation matrices raised to a power yields *adaptive* localization matrix.

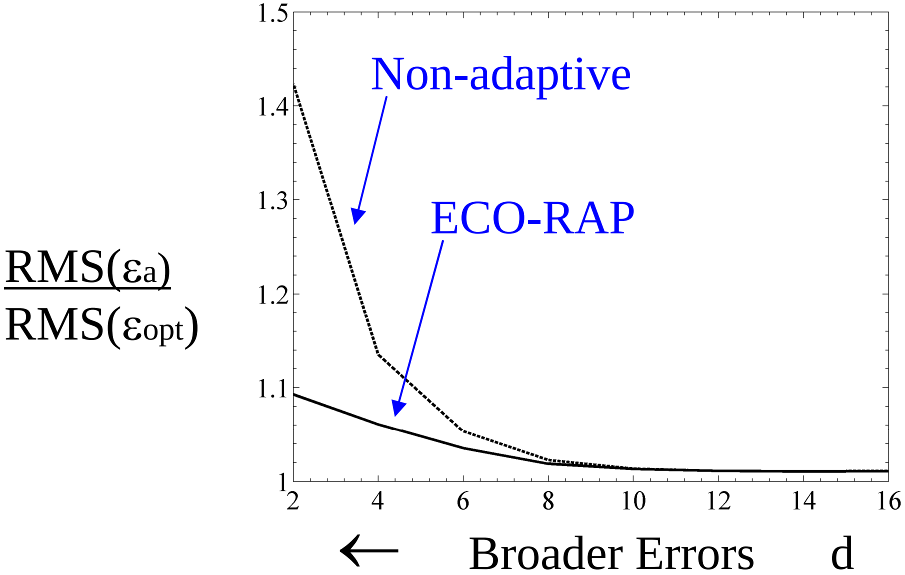
Length Scale Variability Experiment





Tune for short error length scales then test on broad

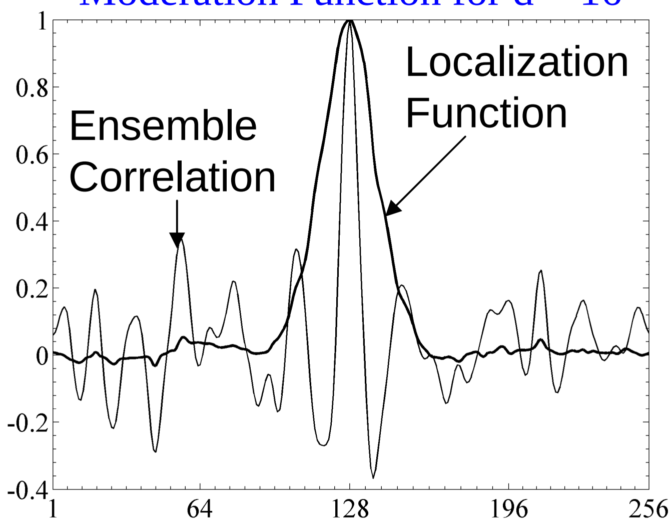
Length Scale Variability Experiment



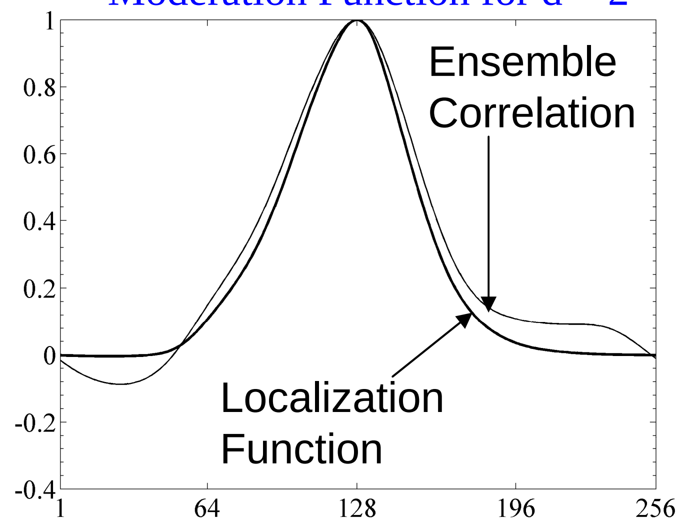
- Tune ECO-RAP and Non-Adaptive localization methods for lowest analysis error at scale d = 16 with 156 obs

- Compare the performance of the two schemes when the true error correlation length scale is broader than that for d=16

Moderation Function for d = 16



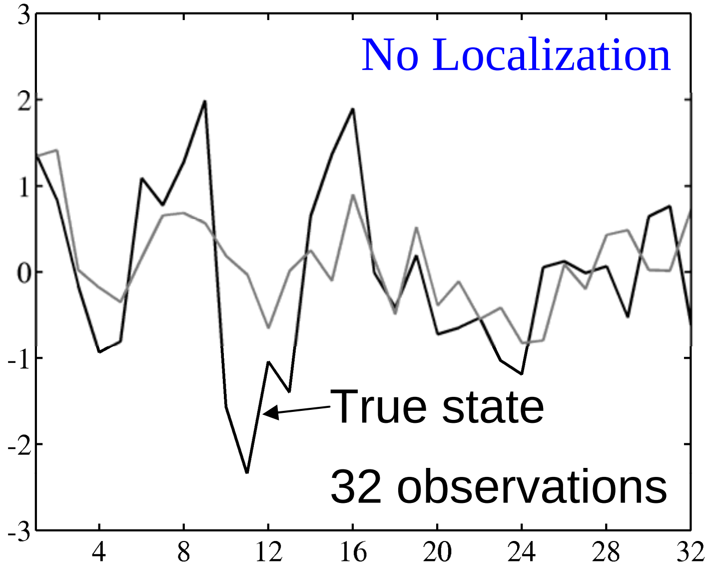
Moderation Function for d = 2

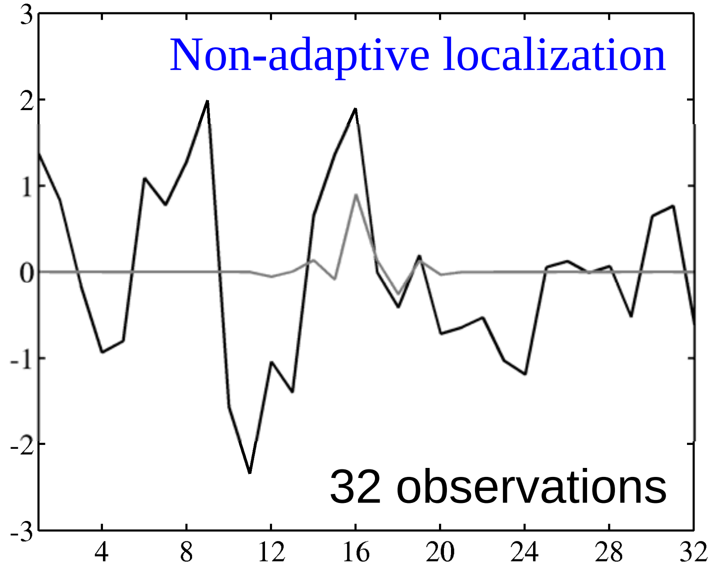


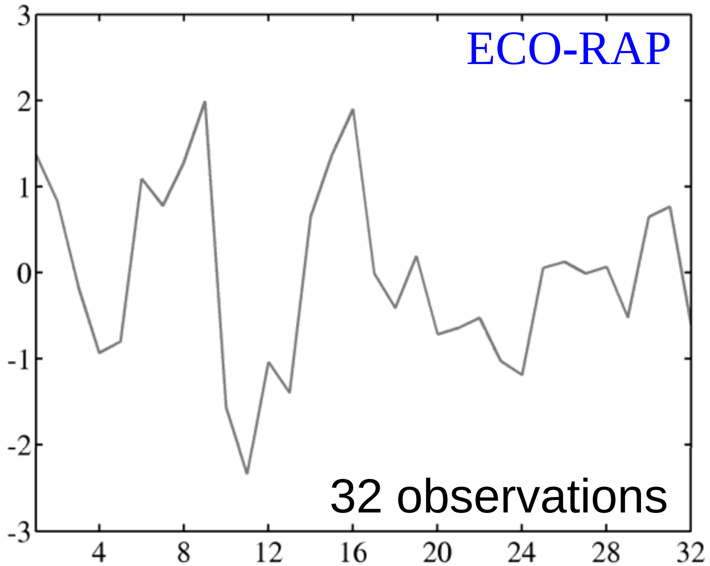
Single ob. 4D Data Assimilation Test (16 members)

- 32 variables in periodic domain
- Truth moves to the right one grid point per time step
- One ob. per time step at variable 16 (very small ob error variance)
- After 32 time steps use all 32 collocated observations to estimate the initial state

Single ob. 4D Data Assimilation Test (16 members)





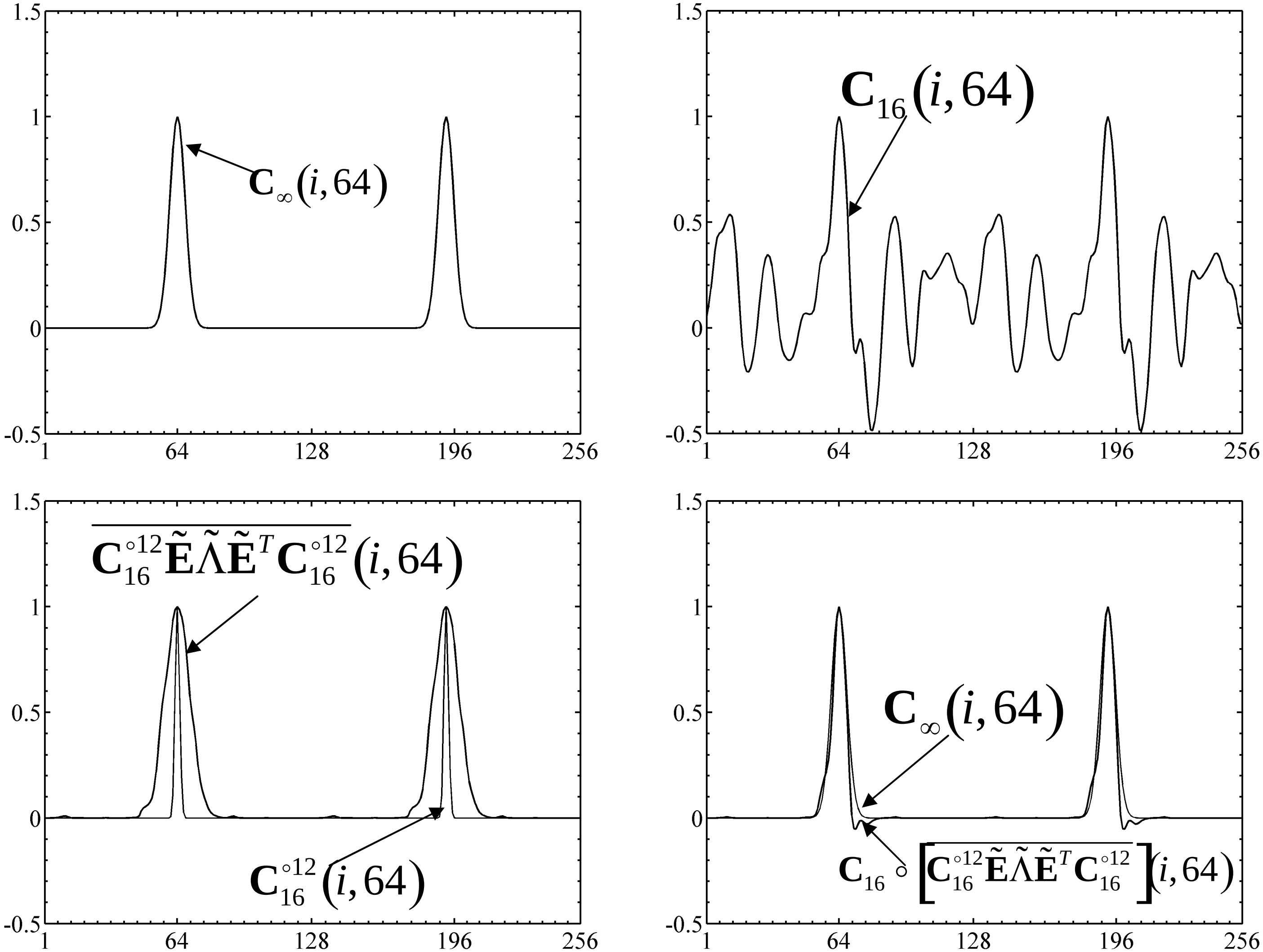


- No localization produces an inaccurate estimate everywhere
- Non-adaptive localization can only use observations close to the analysis time
- ECO-RAP recovers the true state

Localization or Moderation?

1) Are the variables whose forecast errors correlate with another variable confined to the geographic neighbourhood of that variable?
2) What is "local" about forecast errors due to a misspecification of the albedo of stratus clouds?
3) What is "local" about errors associated with a sudden stratospheric warming event?
4) ECO-RAP can moderate spurious correlations even when the answer to (1) is "No".

Localization or Moderation?



K = 16 member ensemble

$\mathbf{C}_K =$ Ensemble correlation (K members)

$\mathbf{C}_K^{\circ n} =$ n Hadamard products of ensemble correlation

$\tilde{\mathbf{E}}\tilde{\Lambda}\tilde{\mathbf{E}}^T =$ a spectral smoother

Review

- ECO-RAP is a new flow-adaptive localization method for ensemble DA.
- It raises ensemble correlations to powers (Hadamard products) to selectively reduce spurious correlations.
- Broad localization functions are obtained by sandwiching non-adaptive localization matrices between correlation matrices raised to a power.
- ECO-RAP adapts to changes in the propagation and scale characteristics of errors.
- ECO-RAP is as good as non-adaptive localization when error distribution is invariant.

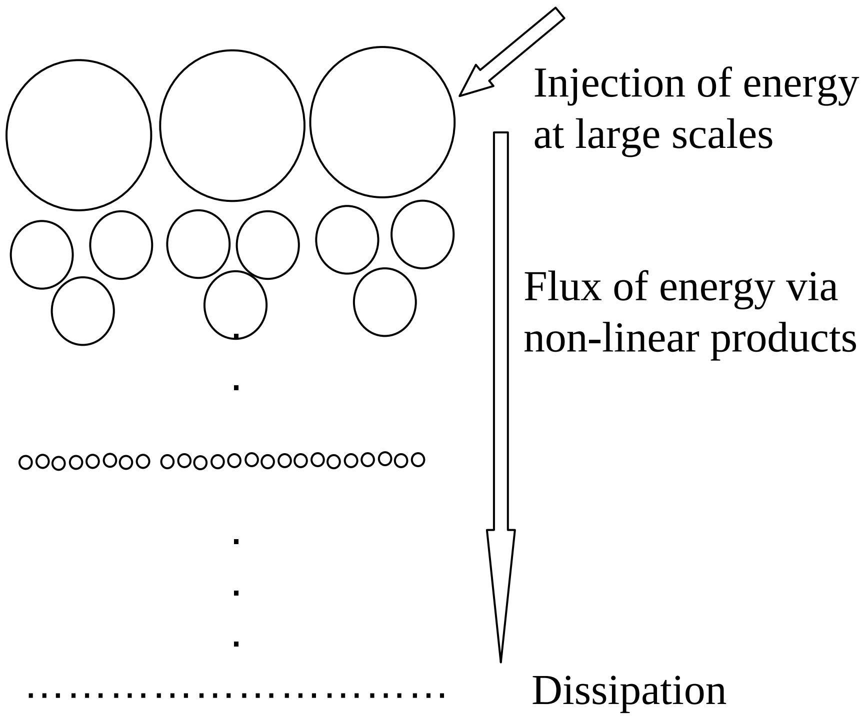
Computational Issues

- N=number of model variables
- $\mathbf{C}_{\text{ECO-RAP}} = \overline{\mathbf{C}_K^{\circ n}\tilde{\mathbf{E}}\tilde{\Lambda}\tilde{\mathbf{E}}^T\mathbf{C}_K^{\circ n}}$ has N^2 elements
- So does the Covariances Adaptively Localized with ECO-rap (CALECO) matrix

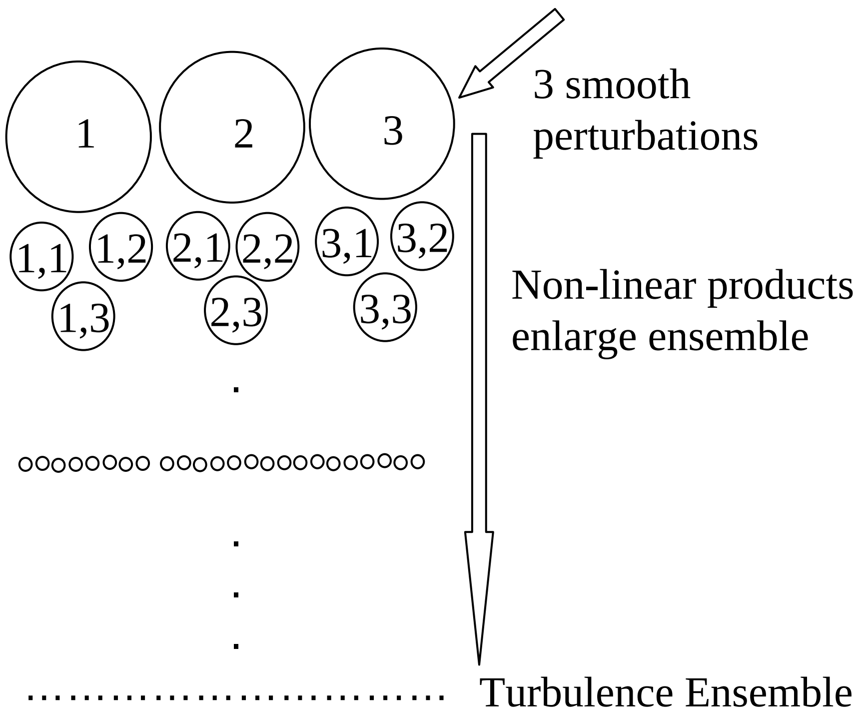
$$\mathbf{P}_{CALECO}^{f} = \mathbf{P}_K^f \circ \left[\overline{\mathbf{C}_K^{\circ n}\tilde{\mathbf{E}}\tilde{\Lambda}\tilde{\mathbf{E}}^T\mathbf{C}_K^{\circ n}} \right]$$

Turbulence inspired recipe for huge ‘turbulence’ ensemble

Turbulent energy cascade following Richardson’s (1922) ideas.



Recipe for creation of huge ensemble from small ensemble



If covariance of turbulence ensemble was CALECO then computer memory would only need to store “energy containing eddies”.
Is there a turbulence ensemble whose covariance is CALECO?

Turbulence ensemble for CALECO

It may be shown that if

$$\mathbf{A} = \mathbf{U}\mathbf{U}^T \text{ and } \mathbf{B} = \mathbf{V}\mathbf{V}^T \text{ then } \mathbf{A} \circ \mathbf{B} = \left[\mathbf{U} \otimes \mathbf{V}\right]\left[\mathbf{U} \otimes \mathbf{V}\right]^T,$$

where $\mathbf{U} \otimes \mathbf{V}$ indicates the matrix whose columns list all possible non-linear products of the columns of $\mathbf{U}$ and $\mathbf{V}$.

Consequently, covariance of turbulence ensemble one obtains by taking all possible non-linear products of "energy containing eddies" $\mathbf{U}$ and $\mathbf{V}$ is the element-wise product of the covariances of $\mathbf{U}$ and $\mathbf{V}$. Hence, since

$$\mathbf{P}_K^f = \mathbf{Z}_K \mathbf{Z}_K^T \text{ and } \mathbf{C}_{\text{ECO-RAP}} = \overline{\mathbf{C}_K^{\circ n} \tilde{\mathbf{E}} \tilde{\Lambda} \tilde{\mathbf{E}}^T \mathbf{C}_K^{\circ n}} = \left[\overline{\mathbf{C}_K^{\circ n} \tilde{\mathbf{E}} \tilde{\Lambda}^{1/2}}\right]\left[\overline{\mathbf{C}_K^{\circ n} \tilde{\mathbf{E}} \tilde{\Lambda}^{1/2}}\right]^T$$

It follows that the energy containing eddies for the turbulence ensemble whose covariance is CALECO are

$$\mathbf{Z}_K \text{ and } \left[\overline{\mathbf{C}_K^{\circ n} \tilde{\mathbf{E}} \tilde{\Lambda}^{1/2}}\right]$$

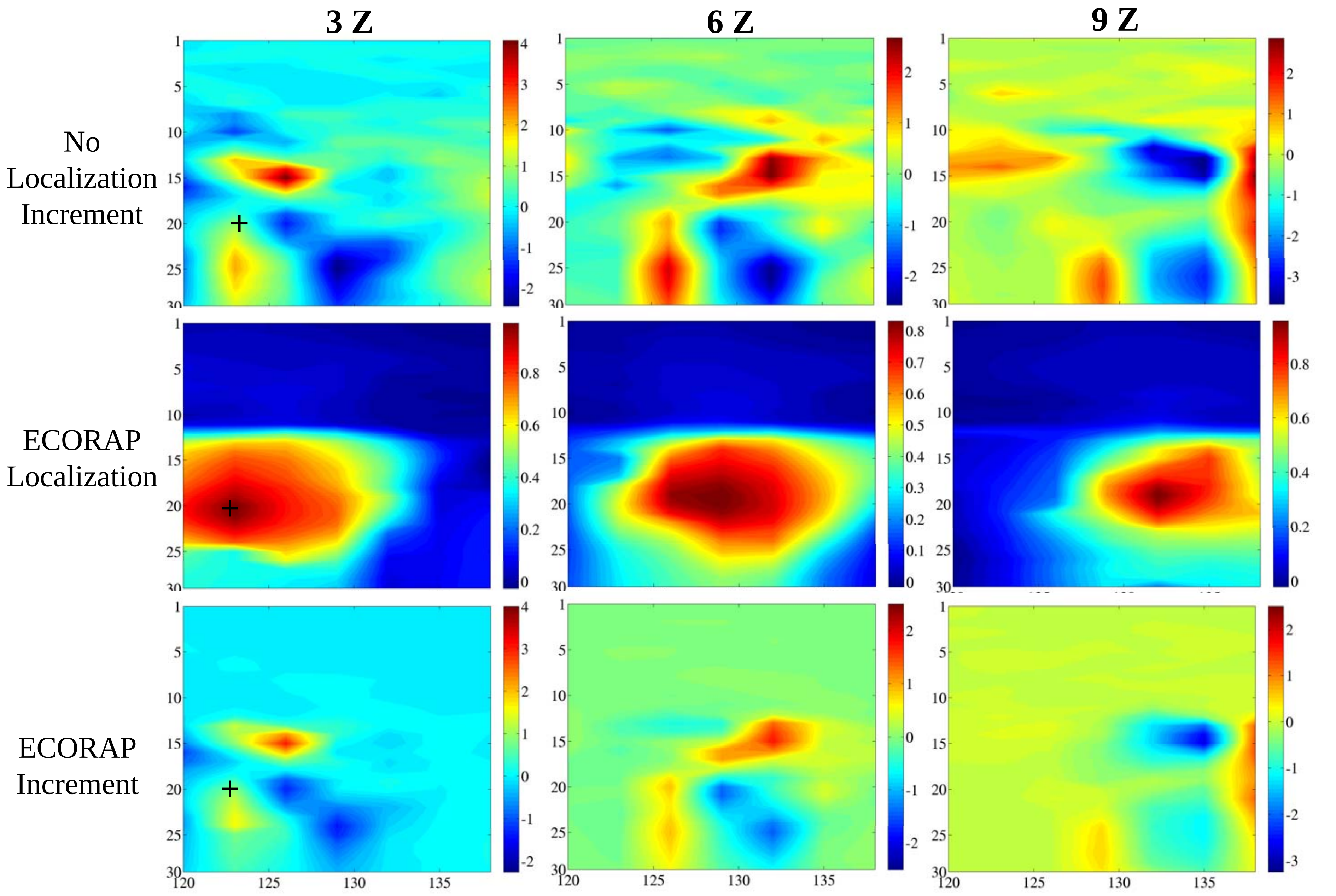
Since columns of $\left[\overline{\mathbf{C}_K^{\circ n} \tilde{\mathbf{E}} \tilde{\Lambda}^{1/2}}\right]$ are in spectral space, for broad localization functions can truncate leaving $\left[\overline{\mathbf{C}_K^{\circ n} \tilde{\mathbf{E}} \tilde{\Lambda}^{1/2}}\right]$ only has L columns ($L < N$).

Separability assumption for ECO-RAP

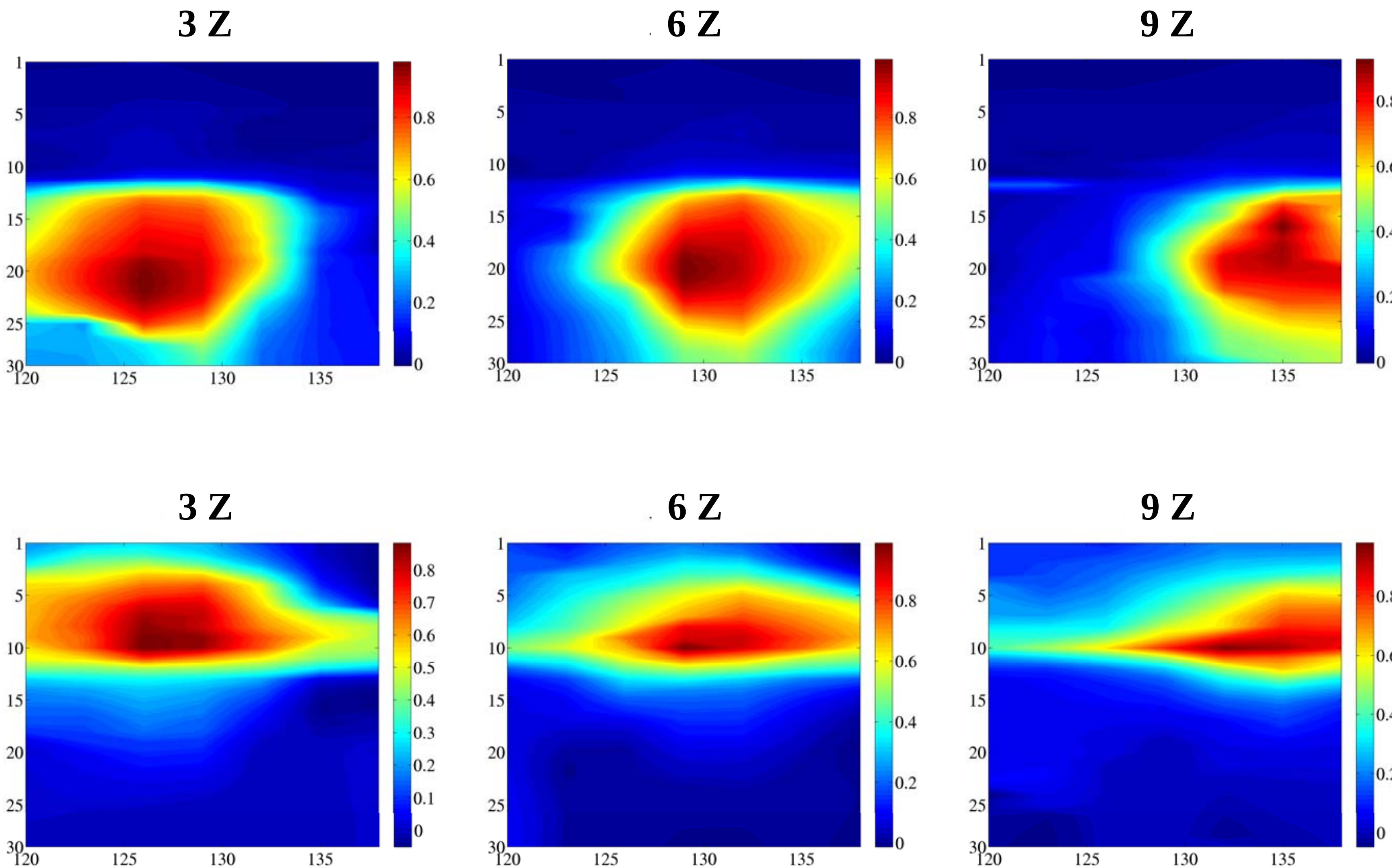
The cost of computing $\mathbf{C}_K^{\circ n}$ may be reduced from N^2 to $(nx+ny+nz)*nv*nt$ by approximating the rows of $\mathbf{C}_K^{\circ n}$ by separable functions of $(x,t),(y,t)$ and (z,t).

This separability assumption also reduces markedly reduces the computation and memory requirements for $\left[\overline{\mathbf{C}_K^{\circ n}\tilde{\mathbf{E}}\tilde{\Lambda}^{1/2}}\right]$, the square root of ECO-RAP.

v Increment From a Single T Ob.



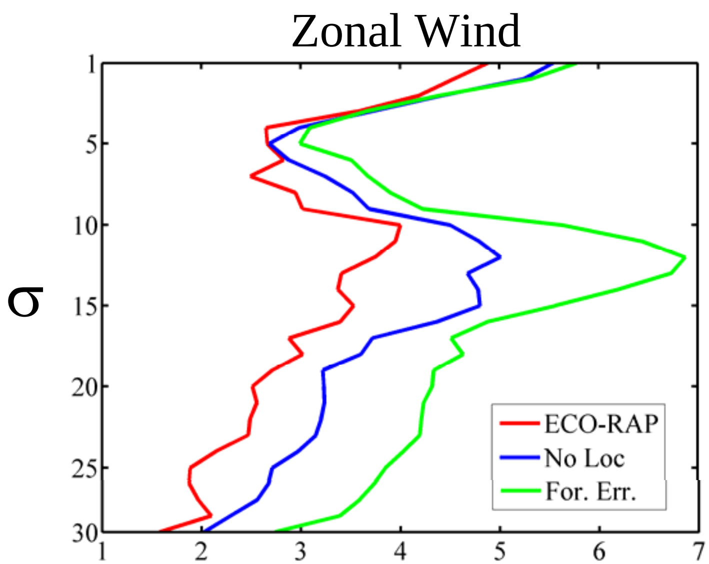
Example ECO-RAP Localization Functions

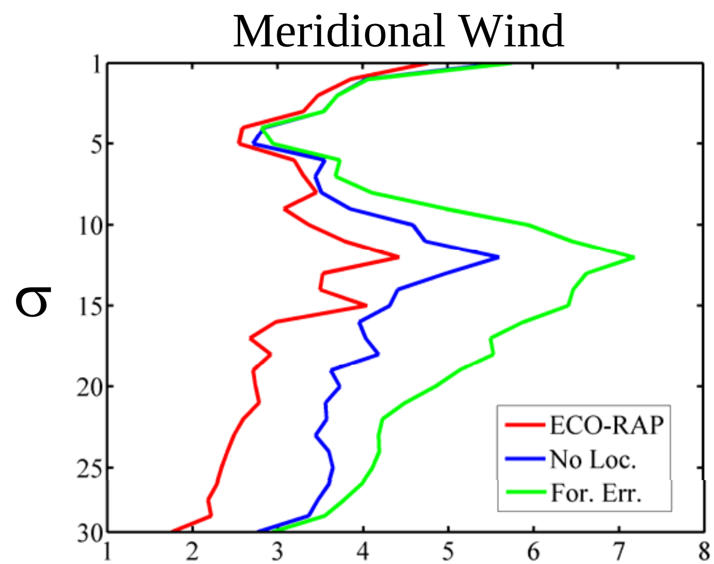


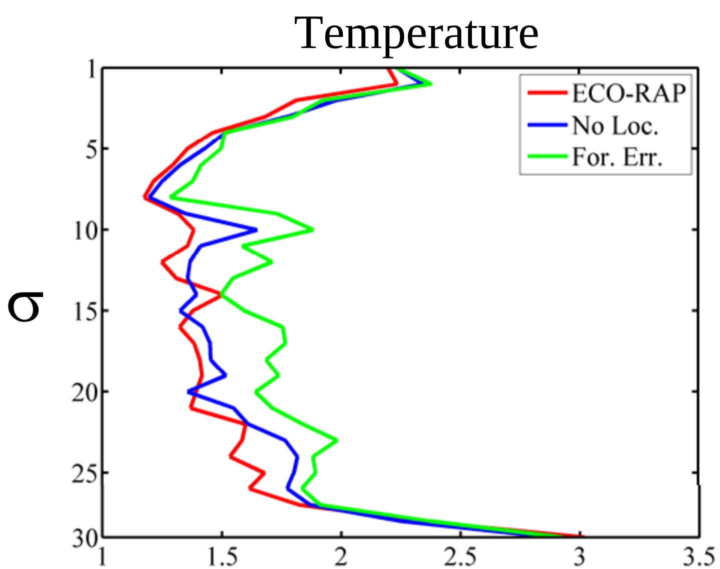
LETKF using CALECO: Preliminary Experiment

- *K* = 27 member ensemble, T119L30 NWP model (NOGAPS).
- 7x7x30 grid box size .
- 3^o grid resolution.
- We observe *u*,*v*,*T* at every point within the box at 3Z and 9Z, and attempt to estimate the state at 6Z.
- ‘Truth’ is assumed to be a 21-27 hour forecast.
- First guess/ensemble come from last 6 hrs of 9 hr forecast valid at the same time.
- Observations are the ‘truth’ plus random number
- Observation error variance is 1 m^2/s^2 and 1 K^2
- Number of obs = 8820, Number of variables=13230
- $K\_{Turbulence}$=1640
- Smoothed ensemble perturbations before applying ECO-RAP
- Correlations were raised to the 12^{th} power

Globally Averaged Results







Global RMS Error

	Forecast	ECORAP	No Loc.
u	4.6	3.1	3.8
v	4.8	3.1	3.9
T	1.8	1.6	1.7

Summary

- Ensemble localization is equivalent to running ensemble through a 1-step turbulent cascade where energy containing eddies are the raw ensemble and the columns of the square root of the localization covariance matrix.
- Turbulence analogy, separability, and spectral truncation enable computationally efficient DA algorithm – cost governed by error dimension.
- ECO-RAP allows larger observation volumes in LETKF - outperforms raw ensemble.

Review

- ECO-RAP is a new flow-adaptive localization method for ensemble DA.
- It raises ensemble correlations to powers (Hadamard products) to selectively reduce spurious correlations.
- Broad localization functions are obtained by sandwiching non-adaptive localization matrices between correlation matrices raised to a power.
- ECO-RAP adapts to changes in the propagation and scale characteristics of errors.
- ECO-RAP is as good as non-adaptive localization when error distribution is invariant.

Intermediate Summary

- ECO-RAP is a new flow-adaptive localization function for ensemble DA, which uses:
  - The largest ensemble correlations to predict the location and scale of evolving error correlation structures
  - Raises ensemble correlations to powers (Hadamard products) to selectively reduce spurious correlations
  - Uses matrix products and spectral smoothing to obtain broad, smooth localization functions
- ECO-RAP pays no penalty when the true errors are not varying
- ECO-RAP adapts to error correlation structures undergoing both propagation and scale changes

Local Ensemble Transform Kalman Filter

- Local Ensemble Transform Kalman Filter (LETKF) [Hunt et al (2007; Physica D)]
  - Each grid point is updated only with the observations lying within grid point's observation volume.
  - Each grid point can be updated independently so algorithm is scalable.
  - Finite observation volume is needed to limit the effect of spurious long-distance correlations.
  - Problematic for observations of vertical integrals of model variables such as satellite obs.
  - Problematic for 4D assimilation when errors propagate further than the localization width over the time window of interest.
  - Redundancy in observation processing since there is a high degree of overlap between volumes.

LETKF using CALECO turbulence

- When CALECO is used in LETKF size of observation volumes is unconstrained because localization is implicit in CALECO.
- Larger observation volumes are appropriate for satellite DA and 4D-DA
- Larger observation volumes enable entire grid columns (or indeed the entire globe) to be updated simultaneously and hence avoids redundancy in observation processing.
- Note that ensemble size is now given by the size of the turbulence ensemble.
- The size of the turbulence ensemble is an upper bound on the dimension of the error and is usually < number of obs.

Multi-variate or Uni-variate ECO-RAP

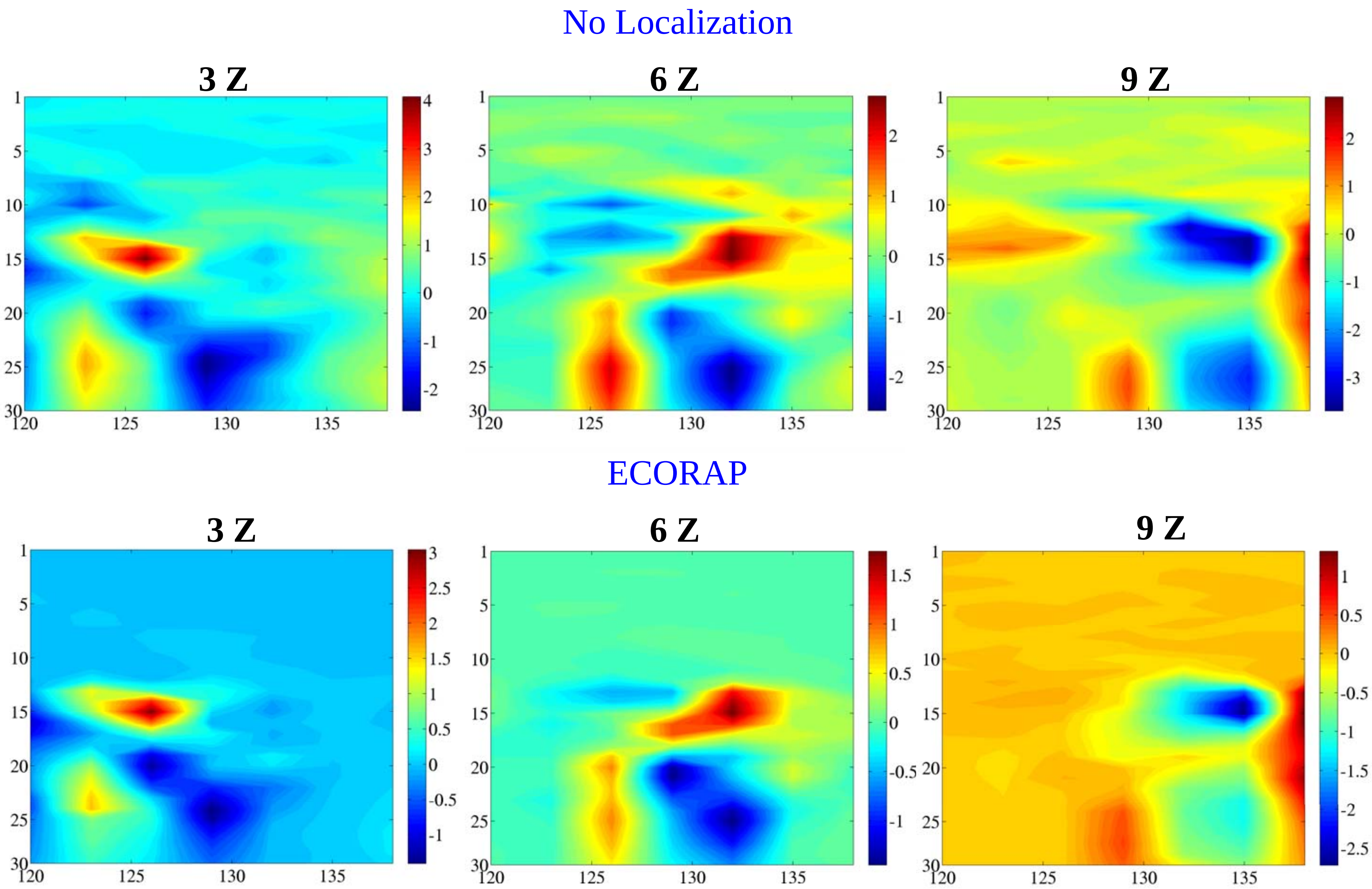
- ECO-RAP can provide multi-variate “localization”.
- However, in this experiment, to further increase the computational efficiency of ECO-RAP, we chose a single variable θ_e to localize u,v,T
- Future work will consider fully multivariate ECO-RAP together with alternative univariate formulations (e.g. using ϕ)

ECO-RAP, Part 2:
Inexpensive Huge Ensembles

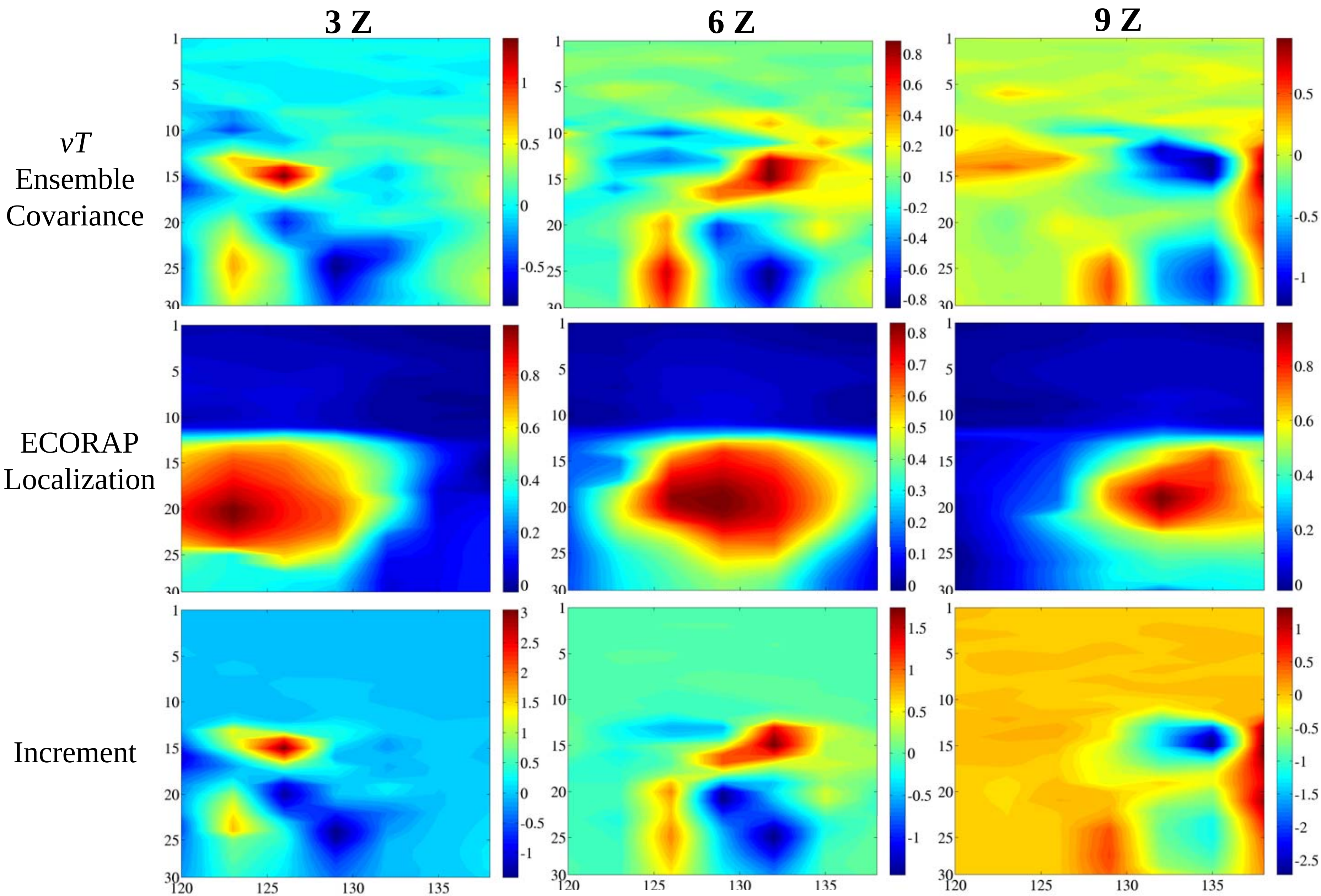
Craig H. Bishop, Daniel Hodyss,
William. F. Campbell, and Justin G. Mclay

Naval Research Laboratory, Monterey, California

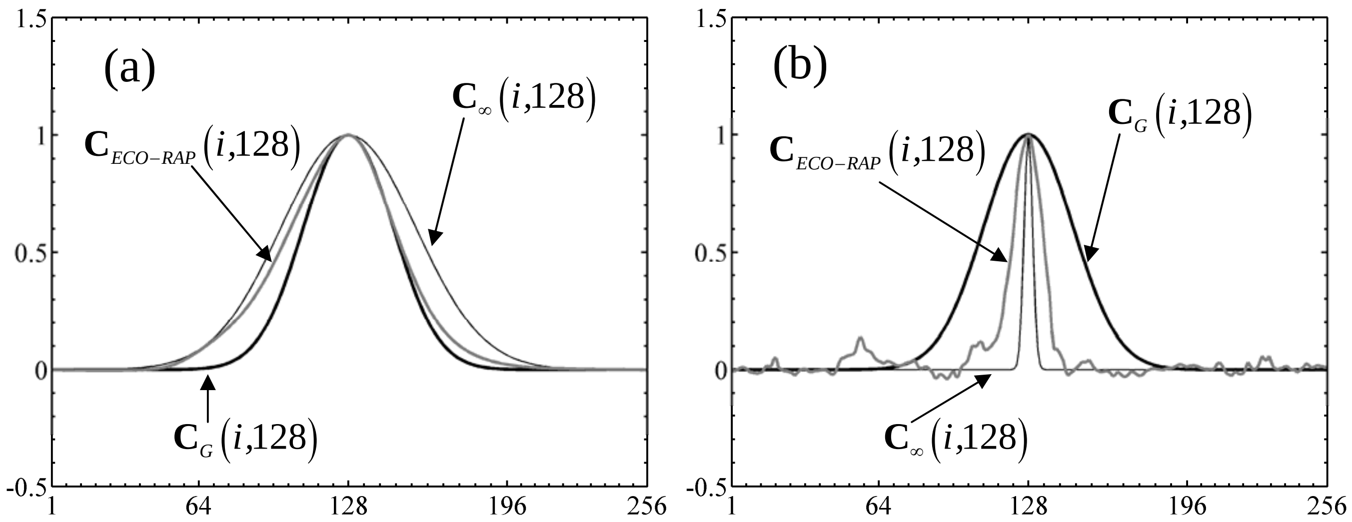
ECORAP versus No Localization



ECORAP in the LETKF

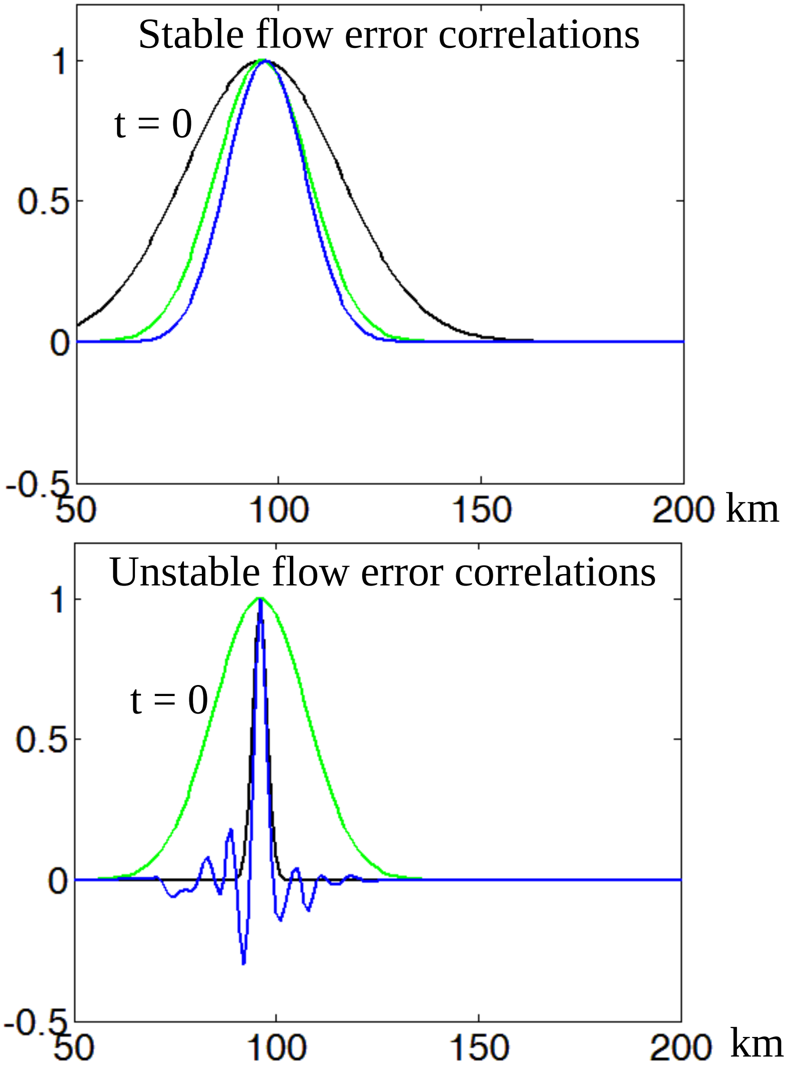


Length Scale Variability Experiments



Small Ensembles and Spurious Correlations

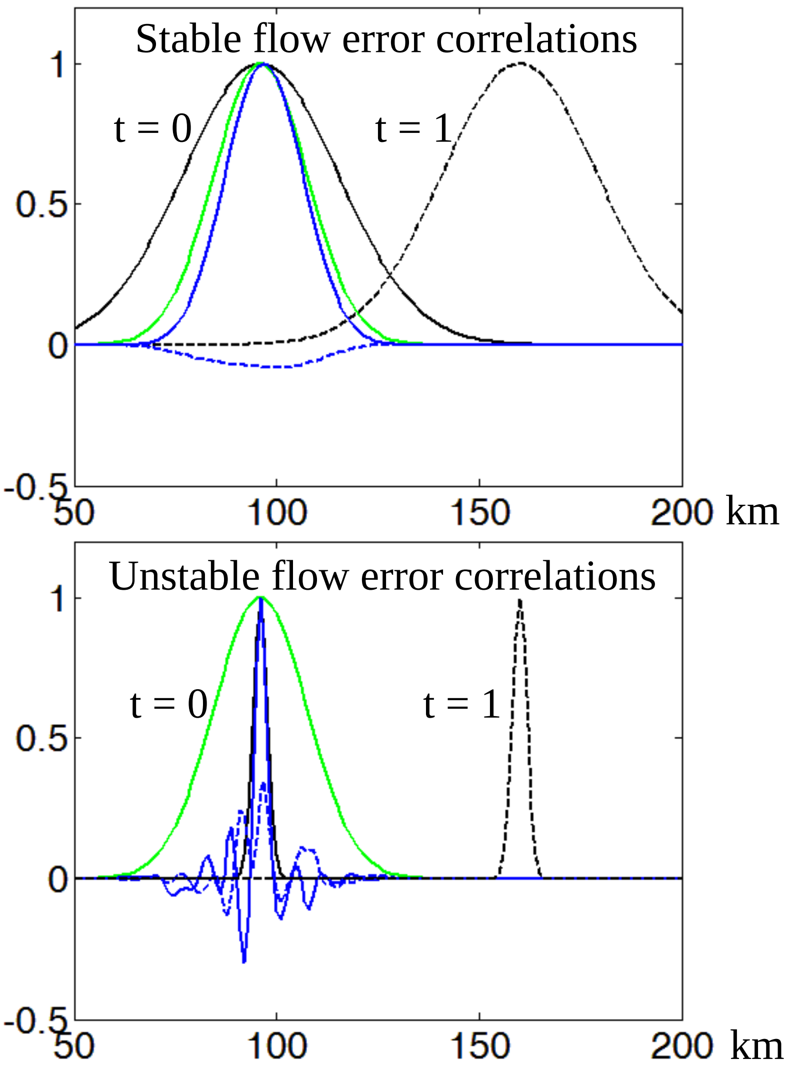
- Current ensemble localization functions poorly represent propagating error correlations.



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