



“Ensemble Squeeze” enabled Multiscale Localisation for the Local EAKF/EnSRF/EnKF - (MLEnKF)

Diego S. Carrió¹, Craig H. Bishop² and Xuguang Wang³

University of the Balearic Islands, km 7.5 Cra. Valldemossa km 7.5, Palma, Spain

Centre of Excellence for Climate Extremes and the School of Geography, Earth and Atmospheric Sciences, The University of Melbourne, Parkville, Victoria, Australia

Mesoscale Data Assimilation and Predictability (MAP) Laboratory, University of Oklahoma, Norman, USA



Main Motivation

Main Motivation of the “Ensemble Squeeze” Localization

➤ MAIN MOTIVATION:

- Implementation of the Local version of the *Gamma Inverse-Gamma (GIG; Bishop, 2016)* in a **LETKF framework, widely used in Operational Centers.**
- LETKF uses the “Ob-Error-Inflation localization” technique.

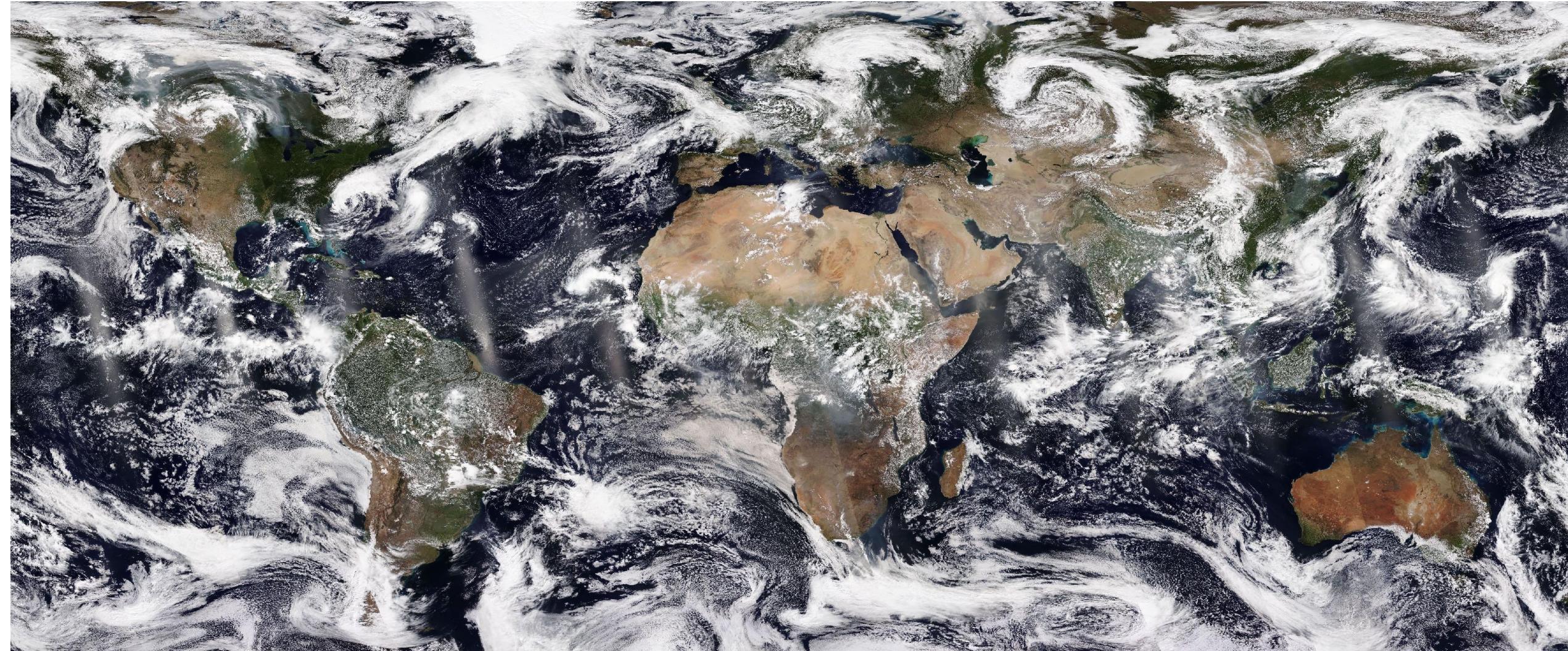
➤ MAIN PROBLEM:

- “**Ob-Error-Inflation**” localization method does not work for **GIG**.

➤ SOLUTIONS:

- Development of a **new localization technique “Ensemble Squeeze”** suitable for **GIG** which is **equivalent** to the “**Ob-Error-Inflation**” localization.
- The “**Ensemble Squeeze**” localization technique **can be applied to any other ensemble-based DA system.**

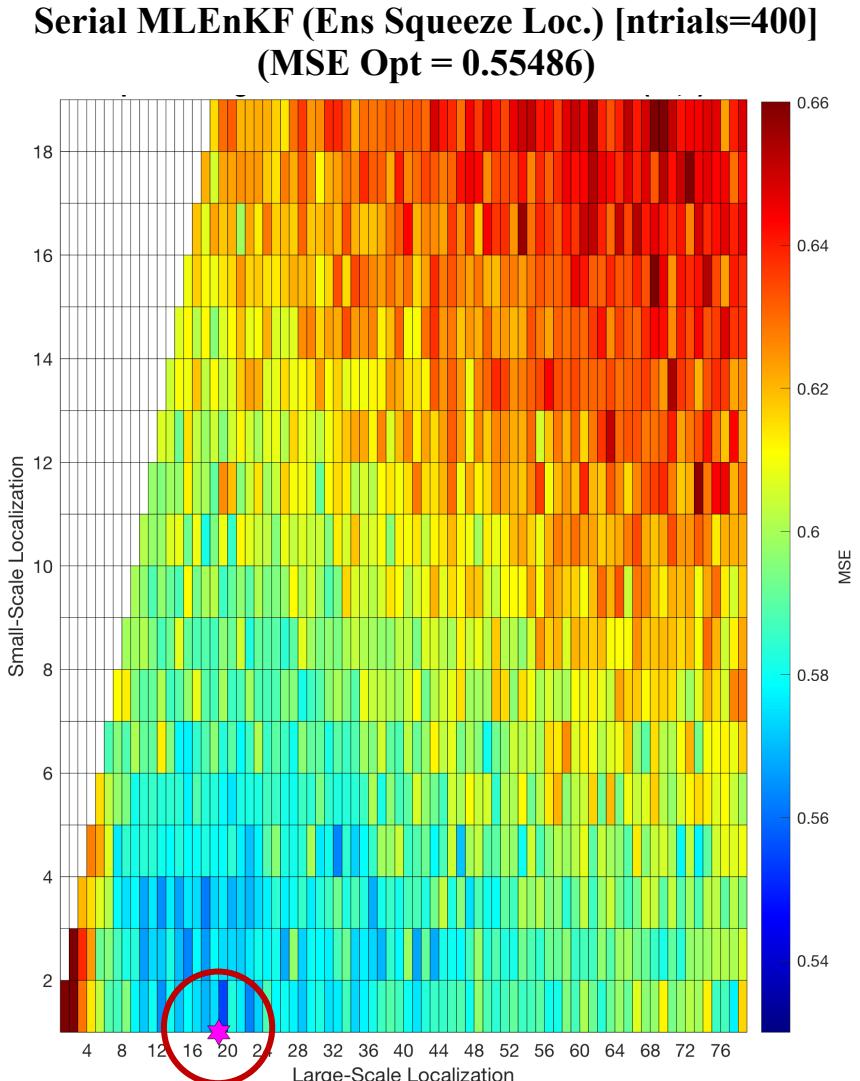
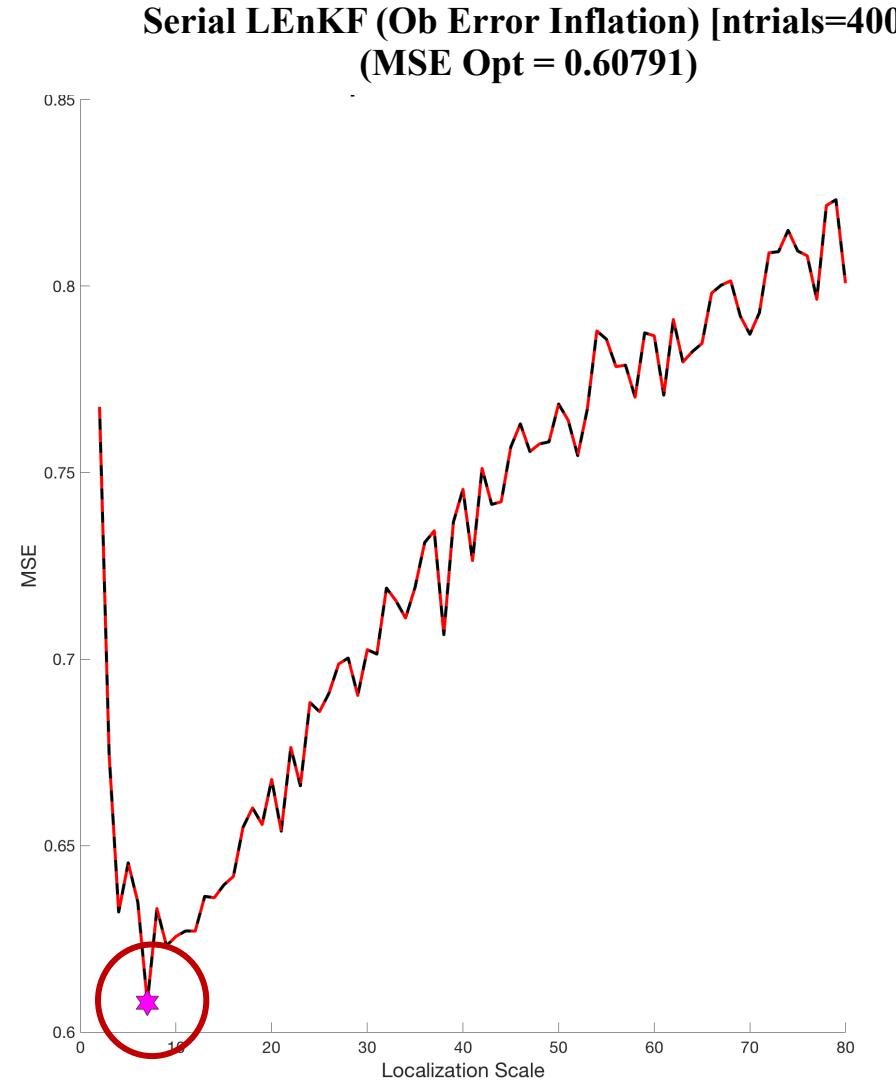
Main Motivation of this presentation



Brute Force Optimal Localization Length Scale

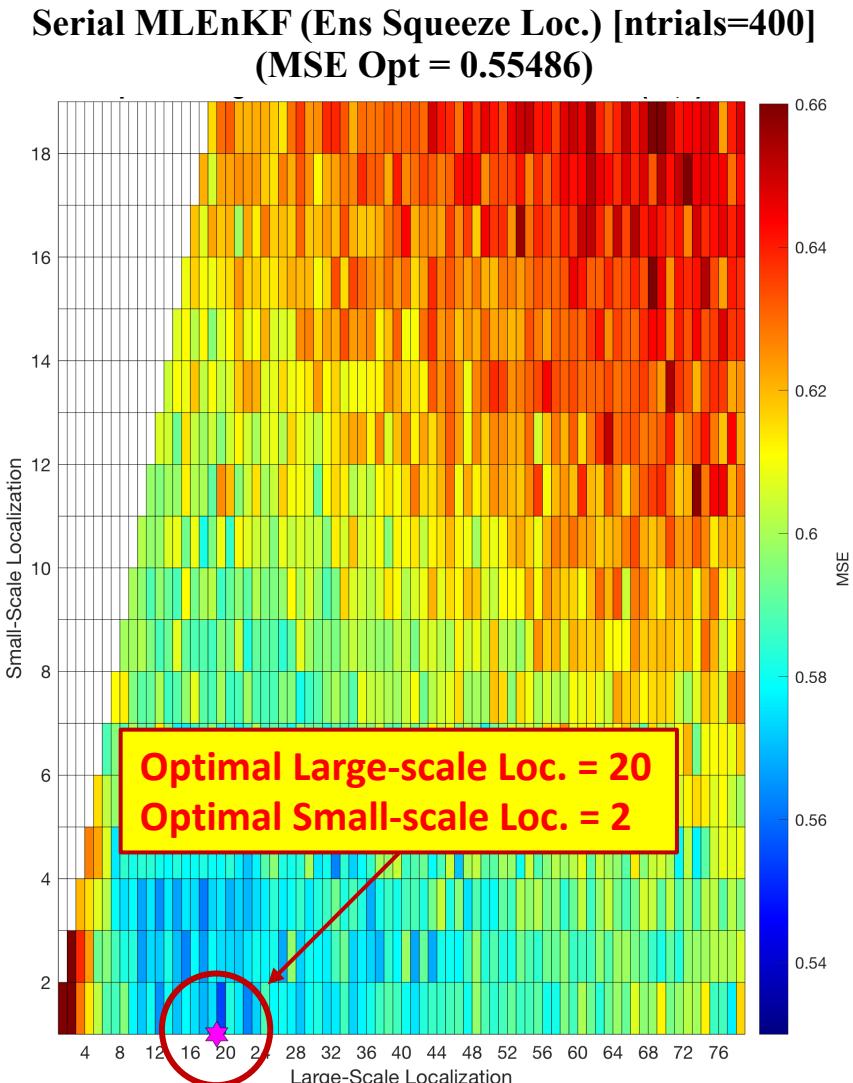
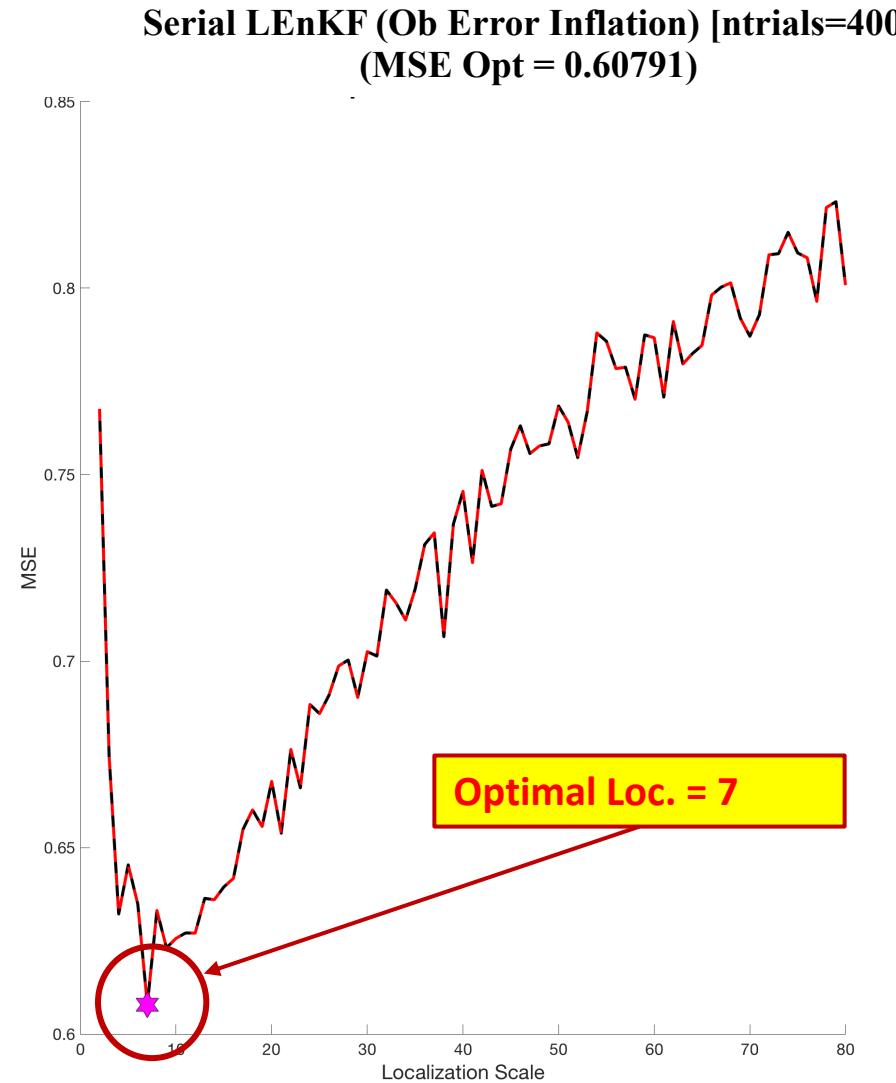
Main Results

n=120
K=10
p=30
ntrials=400
Ob Error var = 1.0

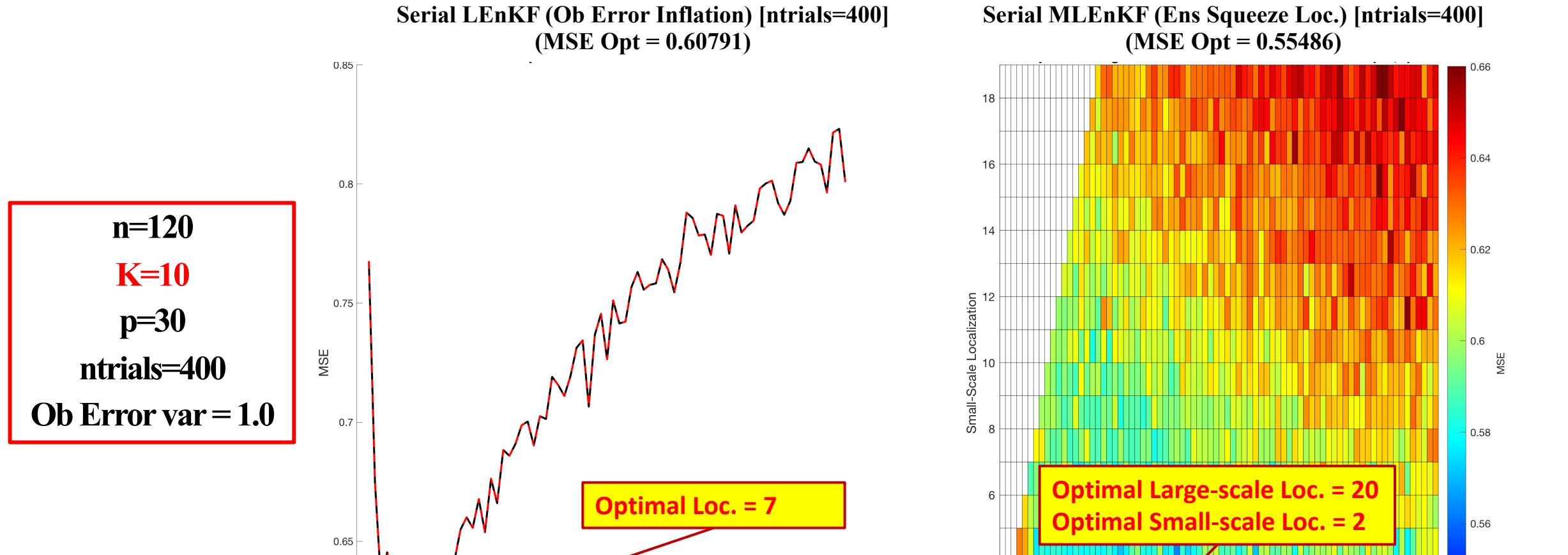


Main Results

n=120
K=10
p=30
ntrials=400
Ob Error var = 1.0



Main Results



Multi-Scale MLEnKF (MSE=0.55486) outperforms Single-Scale LEnKF (MSE=0.60791)

Approach to the “Optimal” Multi-Scale Localization Length Scale

1/K Rule: Getting “Optimal” Localization Length Scale from the Multi-Scale True Correlation Function.

Intersection 1/K with large-scale correlation function at a 30 grid model point distance.

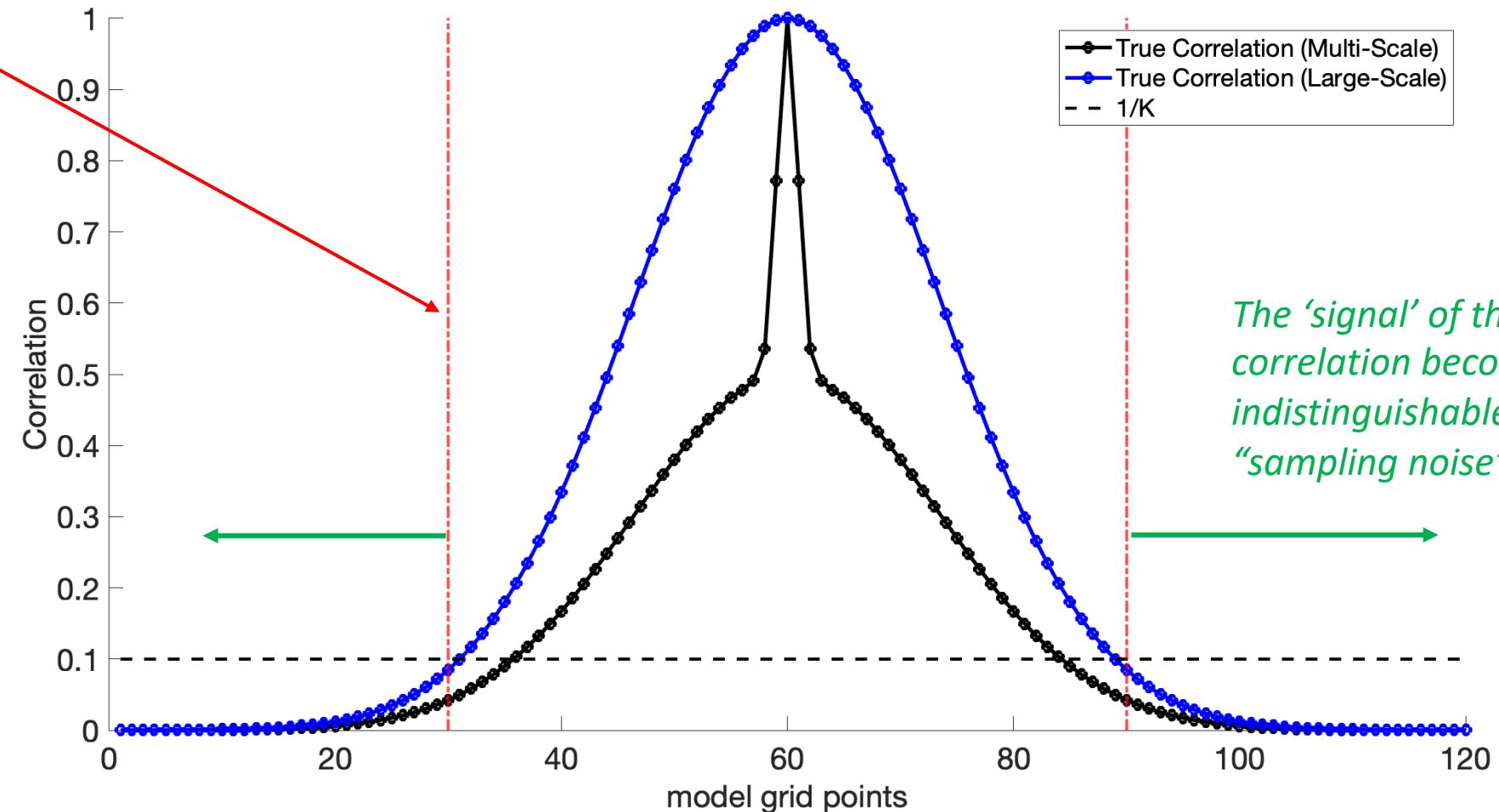


“optimal” localization is 30.
Brute force localization was 20

$K=10, n=120, p=30$

True Correlation (Multi-Scale)
True Correlation (Large-Scale)
- - 1/K

The ‘signal’ of the true correlation becomes indistinguishable from the “sampling noise”



1/K Rule: Getting “Optimal” Localization Length Scale from the Multi-Scale True Correlation Function.

Intersection 1/K with large-scale correlation function at a 30 grid model point distance.



“optimal” localization is 30.
Brute force localization was 20

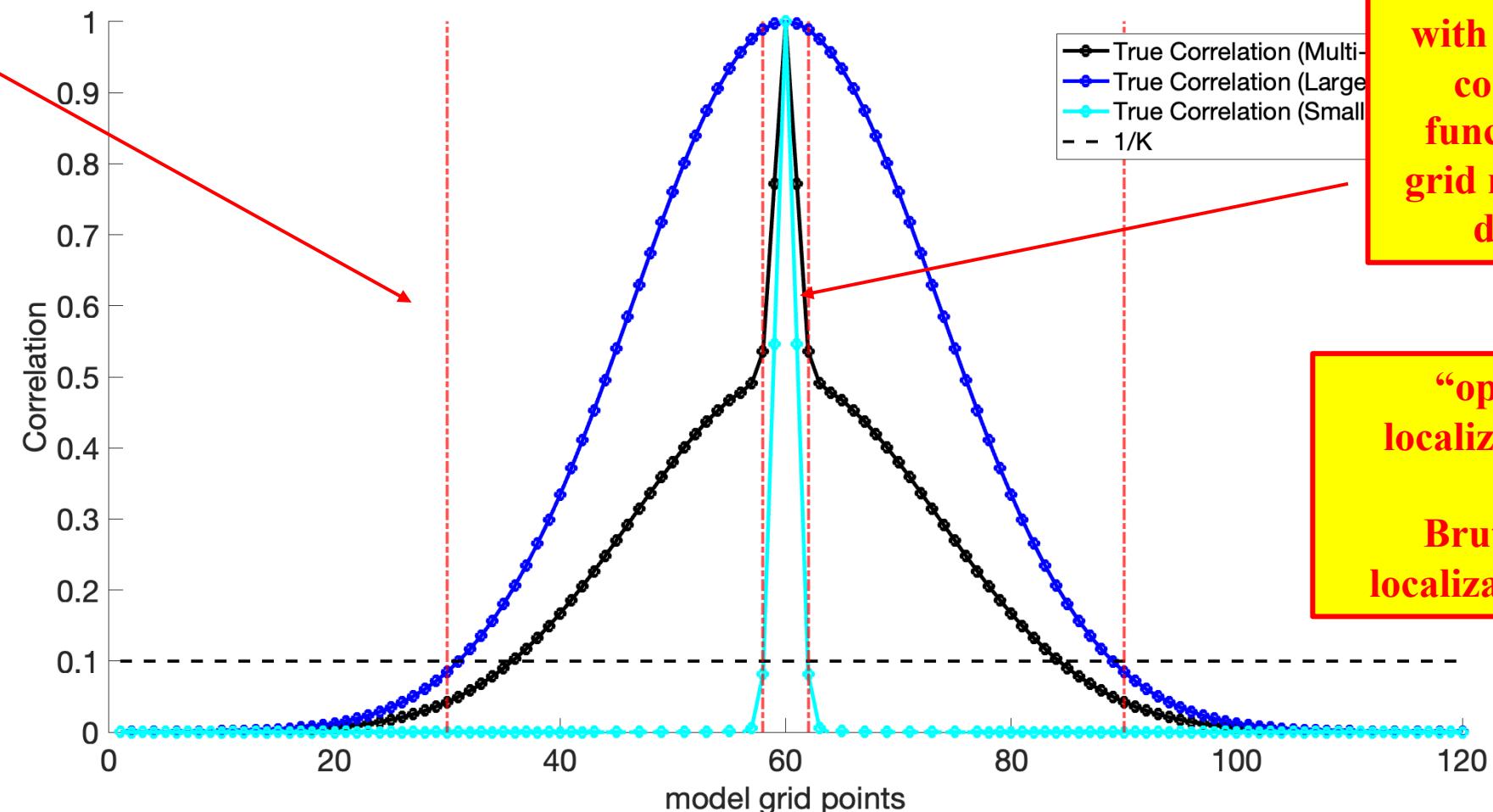
$K=10, n=120, p=30$

True Correlation (Multi-Scale)
True Correlation (Large Scale)
True Correlation (Small Scale)
1/K

Intersection 1/K with small-scale correlation function at a 2 grid model point distance.



“optimal” localization is 2.
Brute force localization was 2



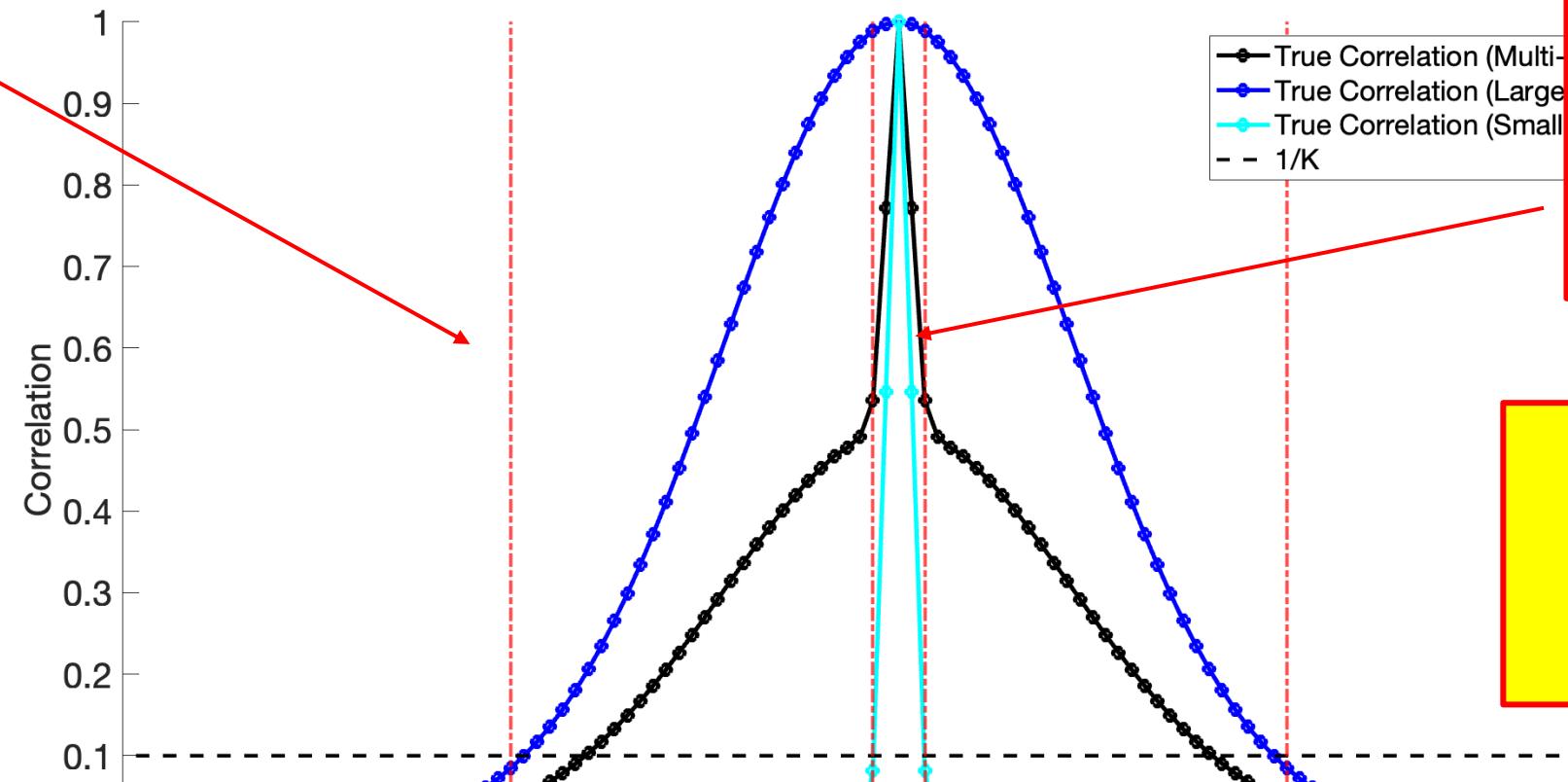
1/K Rule: Getting “Optimal” Localization Length Scale from the Multi-Scale True Correlation Function.

Intersection 1/K with large-scale correlation function at a 30 grid model point distance.



“optimal” localization is 30.
Brute force localization was 20

$K=10, n=120, p=30$



Intersection 1/K with small-scale correlation function at a 2 grid model point distance.



“optimal” localization is 2.
Brute force localization was 2

Results so far suggest that the intersection of 1/K with the climatological covariance function might help speed the tuning of the localization.



Theoretical aspects behind the “*Ensemble Squeeze*” Localization Method

The idea of “*Ob-Error-Inflation*” (Local Volume LETKF Perspective)

Serial EAKF (NCAR-DART)

for $j = 1:p$; % where p is the number of observations

Step 1: Obtain the observed value y_j^o , and the associated observation error variance, R_{jj}

Step 2: Update the ensemble estimate of the observed variable using

$$\begin{aligned} \langle y_j^a \rangle &= \langle y_j^f \rangle + \frac{H_j \mathbf{Z} (H_j \mathbf{Z})^T}{H_j \mathbf{Z} (H_j \mathbf{Z})^T + R_{jj}} (y_j^o - \langle y_j^f \rangle) = \langle y_j^f \rangle + \frac{H_j \mathbf{P}^f H_j^T}{H_j \mathbf{P}^f H_j^T + R_{jj}} (y_j^o - \langle y_j^f \rangle), \text{ and} \\ y_{ji}^a &= \langle y_j^a \rangle + \sqrt{1 - P^f (P^f + R_{jj})^{-1}} (y_{ji}^f - \langle y_j^f \rangle), \quad \text{for } i = 1, 2, \dots, K \quad (P^f = H_j \mathbf{P}^f H_j^T) \end{aligned}$$

Step 3: Find corresponding analysis ensemble for observations and model variables using

$$y_{ki}^a = y_{ki}^f + \frac{H_k \mathbf{Z} (H_j \mathbf{Z})^T}{(H_j \mathbf{Z}) (H_j \mathbf{Z})^T} (y_{ji}^a - y_{ji}^f) \text{ for } k = 1, 2, \dots, p \text{ and } i = 1, 2, \dots, K, \text{ and}$$

$$x_{\mu i}^a = x_{\mu i}^f + \frac{\mathbf{Z}(\mu, :) (H_j \mathbf{Z})^T}{(H_j \mathbf{Z}) (H_j \mathbf{Z})^T} (y_{ji}^a - y_{ji}^f) \text{ for } \mu = 1, 2, \dots, n, \text{ and } i = 1, 2, \dots, K$$

Step 4: Let the analysis ensemble be the prior ensemble for the next observation

$$y_{ki}^f = y_{ki}^a, \text{ for } k = 1, 2, \dots, p; i = 1, 2, \dots, K$$

$$x_{\mu i}^f = x_{\mu i}^a, \text{ for } \mu = 1, 2, \dots, n; i = 1, 2, \dots, K$$

The idea of “Ob-Error-Inflation” (Local Volume LETKF Perspective)

Local Serial EAKF “Ob-Error-Inflation Localization”

for $j = 1 : p$; % where p is the number of observations

Step 1: Obtain the observed value y_j^o , and the associated observation error variance, R_{jj}

Step 2: Update the ensemble estimate of the observed variable using ob error inflation factor $a_{\mu j}$

$$\begin{aligned} \langle y_j^a \rangle &= \langle y_j^f \rangle + \frac{H_j \mathbf{Z} (H_j \mathbf{Z})^T}{H_j \mathbf{Z} (H_j \mathbf{Z})^T + a_{\mu j} R_{jj}} (y_j^o - \langle y_j^f \rangle) = \langle y_j^f \rangle + \frac{H_j \mathbf{P}^f H_j^T}{H_j \mathbf{P}^f H_j^T + a_{\mu j} R_{jj}} (y_j^o - \langle y_j^f \rangle), \text{ and} \\ y_{ji}^a &= \langle y_j^a \rangle + \sqrt{1 - P^f (P^f + a_{\mu j} R_{jj})^{-1}} (y_{ji}^f - \langle y_j^f \rangle) \text{ for } i = 1, 2, \dots, K \quad (P^f = H_j \mathbf{P}^f H_j^T) \end{aligned}$$

Step 3: Find corresponding analysis ensemble for observations and model variables using

$$y_{ki}^a = y_{ki}^f + \frac{H_k \mathbf{Z} (H_j \mathbf{Z})^T}{(H_j \mathbf{Z}) (H_j \mathbf{Z})^T} (y_{ji}^a - y_{ji}^f) \text{ for } k = 1, 2, \dots, p \text{ and } i = 1, 2, \dots, K, \text{ and}$$

$$x_{\mu i}^a = x_{\mu i}^f + \frac{\mathbf{Z} (\mu, :) (H_j \mathbf{Z})^T}{(H_j \mathbf{Z}) (H_j \mathbf{Z})^T} (y_{ji}^a - y_{ji}^f) \text{ for } \mu = 1, 2, \dots, n, \text{ and } i = 1, 2, \dots, K$$

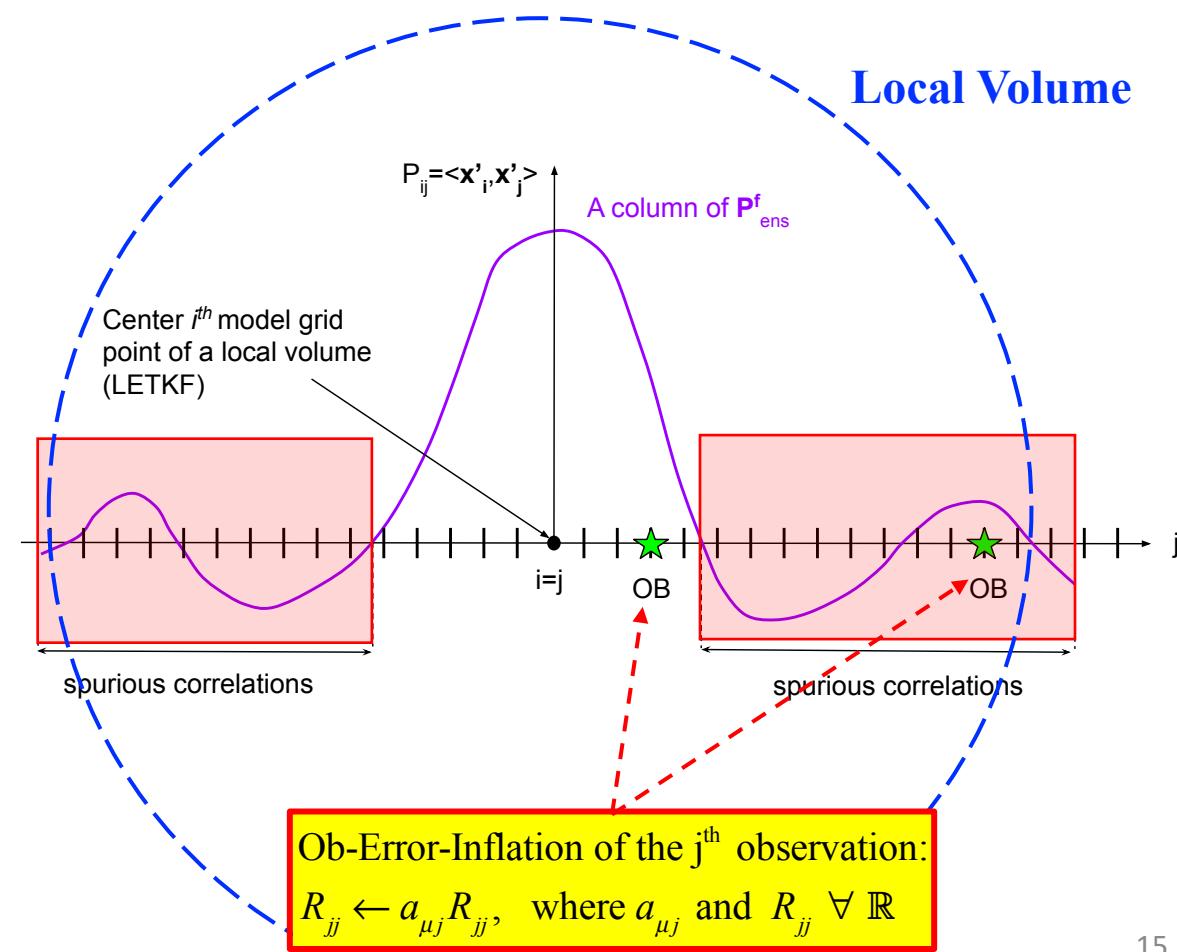
Step 4: Let the analysis ensemble be the prior ensemble for the next observation

$$y_{ki}^f = y_{ki}^a, \text{ for } k = 1, 2, \dots, p; i = 1, 2, \dots, K$$

$$x_{\mu i}^f = x_{\mu i}^a, \text{ for } \mu = 1, 2, \dots, n; i = 1, 2, \dots, K$$

end

1-dimensional “Ob-Error-Inflation” localization example



The Local Serial Observation EAKF ("Ensemble Squeeze" localization)

Local Serial EAKF "Ensemble Squeeze Localization"

for $j = 1 : p$; % where p is the number of observations

Step 1 : Obtain the observed value y_j^o , and the associated observation error variance, R_{jj}

Step 2 : Update the ensemble estimate of the observed variable using the ensemble squeeze factor $b_{\mu j} = a_{\mu j}^{-1/2}$

$$\begin{aligned}\langle y_j^a \rangle &= \langle y_j^f \rangle + \frac{b_{\mu j} H_j \mathbf{Z} (H_j \mathbf{Z})^T}{b_{\mu j} H_j \mathbf{Z} (b_{\mu j} H_j \mathbf{Z})^T + R_{jj}} (y_j^o - \langle y_j^f \rangle) = \langle y_j^f \rangle + \frac{(b_{\mu j}^2 H_j \mathbf{P}^f H_j^T)}{(b_{\mu j}^2 H_j \mathbf{P}^f H_j^T + R_{jj})} (y_j^o - \langle y_j^f \rangle), \text{ and} \\ y_{ji}^a &= \langle y_j^a \rangle + \sqrt{1 - \frac{b_{\mu j}^2 P^f}{(b_{\mu j}^2 P^f + R_{jj})^{-1}}} (y_{ji}^f - \langle y_j^f \rangle) \text{ for } i = 1, 2, \dots, K\end{aligned}$$

where $P^f = H_j \mathbf{P}^f H_j^T$.

Step 3 : Find corresponding analysis ensemble for observations and model variables using

$$y_{ki}^a = y_{ki}^f + \frac{H_k \mathbf{Z} (H_j \mathbf{Z})^T}{(H_j \mathbf{Z}) (H_j \mathbf{Z})^T} (y_{ji}^a - y_{ji}^f) \text{ for } k = 1, 2, \dots, p \text{ and } i = 1, 2, \dots, K, \text{ and}$$

$$x_{\mu i}^a = x_{\mu i}^f + \frac{\mathbf{Z} (\mu, :) (H_j \mathbf{Z})^T}{(H_j \mathbf{Z}) (H_j \mathbf{Z})^T} (y_{ji}^a - y_{ji}^f) \text{ for } \mu = 1, 2, \dots, n, \text{ and } i = 1, 2, \dots, K$$

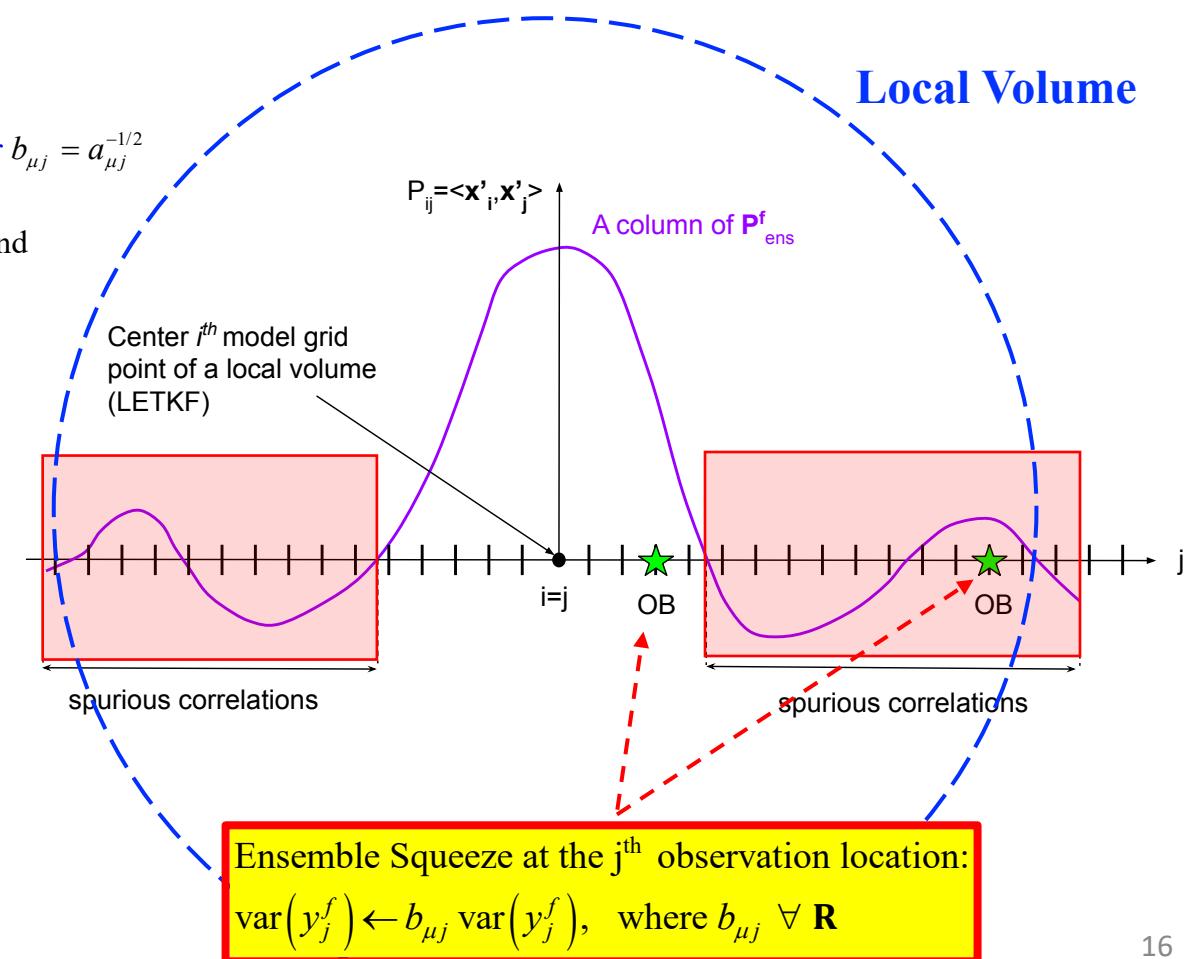
Step 4 : Let the analysis ensemble be the prior ensemble for the next observation

$$y_{ki}^f = y_{ki}^a, \text{ for } k = 1, 2, \dots, p; i = 1, 2, \dots, K$$

$$x_{\mu i}^f = x_{\mu i}^a, \text{ for } \mu = 1, 2, \dots, n; i = 1, 2, \dots, K$$

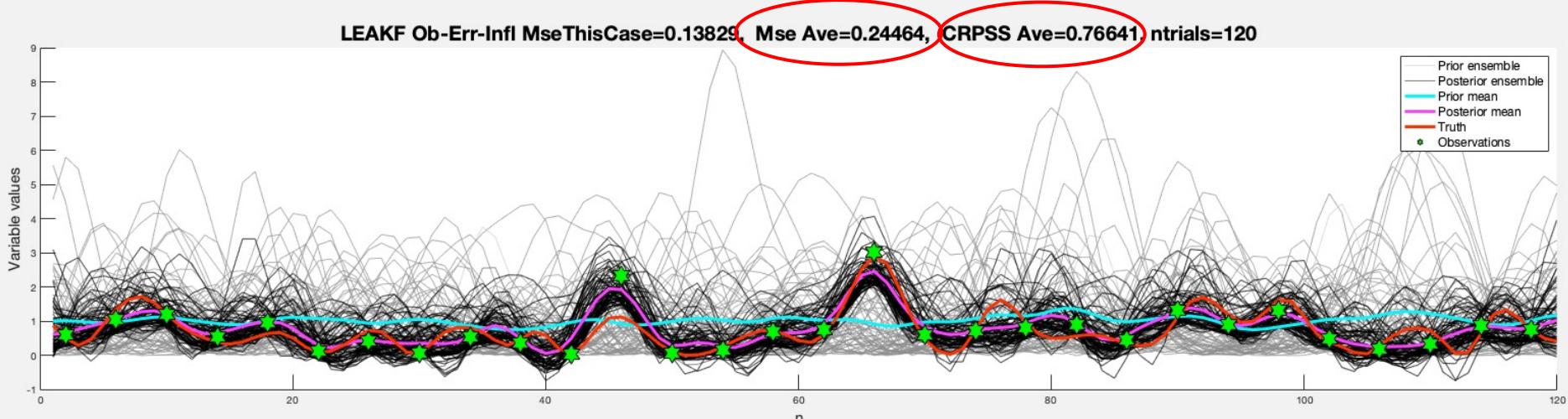
end

1-dimensional "Ensemble Squeeze" localization example



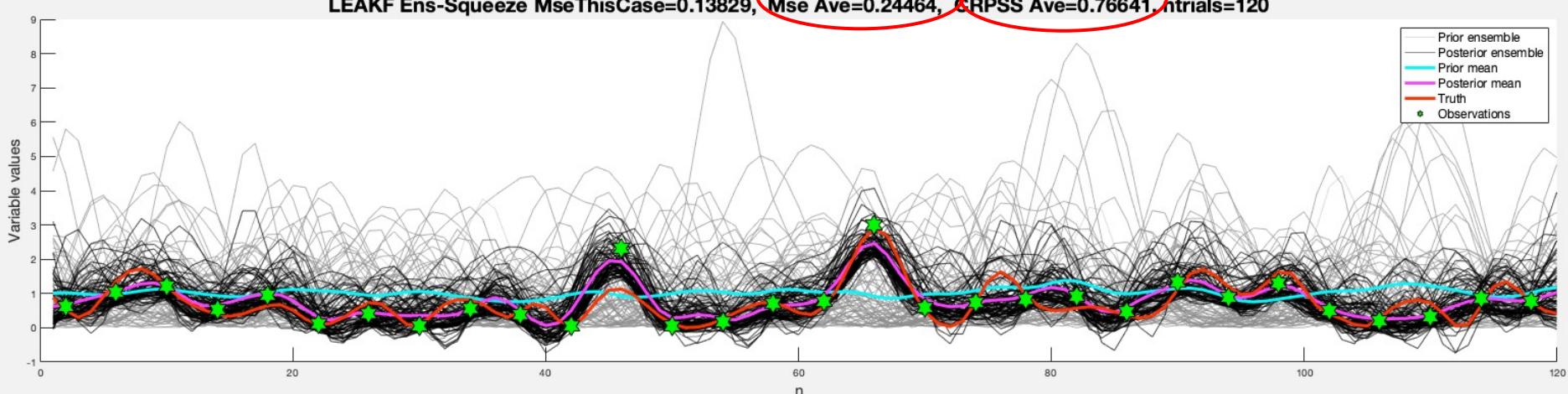
Equivalence Local Serial EAKF (*Ob-Error-Inflation*) vs Local Serial EAKF (*Ens Squeeze*)

LEAKF
(*Ob-Err-Infl*)



Both localization
methods produce
precisely the same
results!!

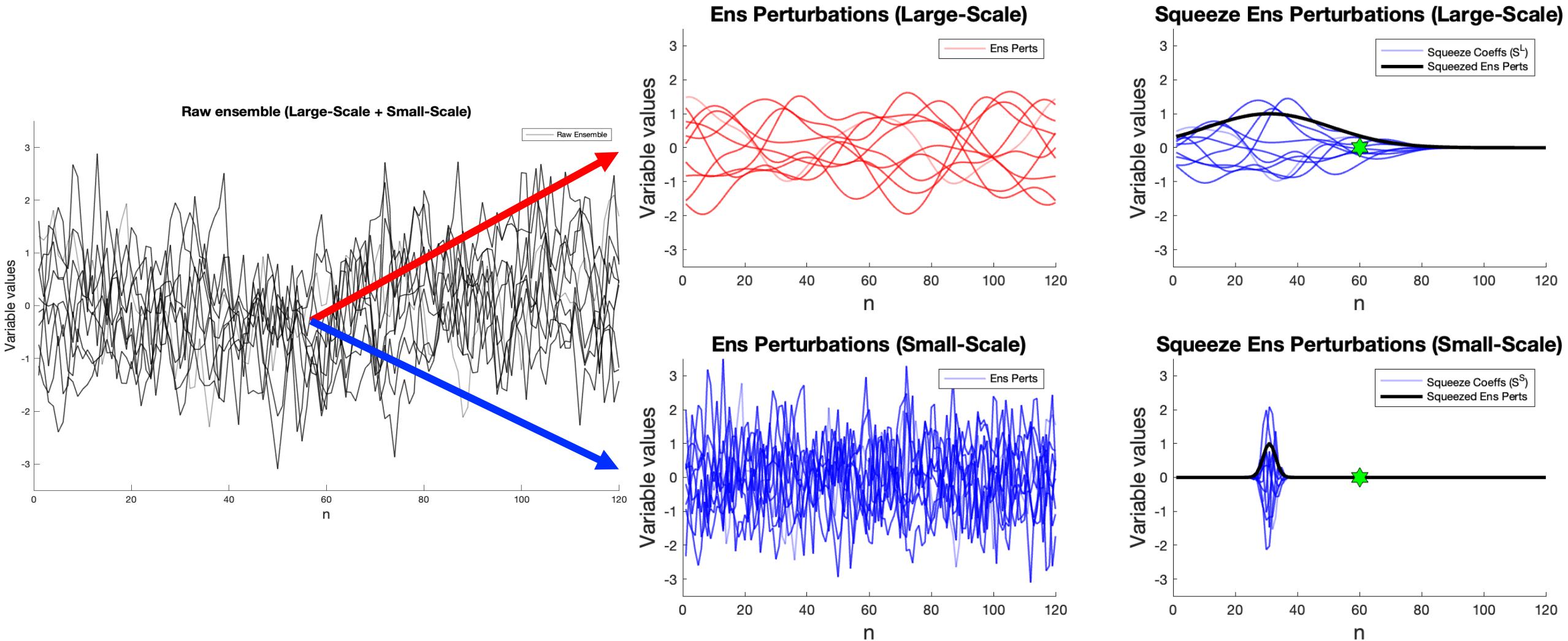
LEAKF
(*Ens-Squeeze*)





The “*Ensemble Squeeze*” Localization applied to
the Multi-Scale DA problem

The Multi-Scale Localization (Main Idea)



Multi-Scale Local MLEnKF using “Ensemble Squeeze Localization” (Pseudo-Code)

for $j = 1 : p$; % where p is the number of observations in the local volume

Step 0a : Find $\text{omk} = \mathbf{o}_{ji} = [\mathbf{o}_{ji}^L, \mathbf{o}_{ji}^S]$, for $i = 1, 2, \dots, K$ such that $y_{ji}^f = h_j(\mathbf{o}_{ji}) \cong h_j(\overline{\mathbf{o}_{ji}}) + \mathbf{h}_j^T(\mathbf{o}_{ji} - \overline{\mathbf{o}_{ji}})$

Step 0b : $\text{omk_squeeze} = \overline{\mathbf{o}_{ji}} + \mathbf{s} \odot (\mathbf{o}_{ji} - \overline{\mathbf{o}_{ji}}) = [\overline{\mathbf{o}_{ji}^L} + \mathbf{s}^L \odot (\mathbf{o}_{ji}^L - \overline{\mathbf{o}_{ji}^L}), \overline{\mathbf{o}_{ji}^S} + \mathbf{s}^S \odot (\mathbf{o}_{ji}^S - \overline{\mathbf{o}_{ji}^S})]$,

where elements of \mathbf{s} squeeze the perturbations based on how far they are from the grid point being updated.

We define $H_j \mathbf{Z}_{\text{omk}}^{\text{squeeze}}$ as:

$$H_j \mathbf{Z}_{\text{omk}}^{\text{squeeze}} = \left[\mathbf{h}_j^T \left[\mathbf{s}^L \odot (\mathbf{o}_{ji}^L - \overline{\mathbf{o}_{ji}^L}) \right], \mathbf{h}_j^T \left[\mathbf{s}^S \odot (\mathbf{o}_{ji}^S - \overline{\mathbf{o}_{ji}^S}) \right] \right] = \left[\mathbf{s}^L \mathbf{Z}^L[\text{omk}, :], \mathbf{s}^S \mathbf{Z}^S[\text{omk}, :] \right] (\sqrt{K-1})$$

Also note that:

$$H_k \mathbf{Z}_{\text{omk}}^{\text{squeeze}} = \left[\mathbf{h}_k^T \left[\mathbf{s}^L \odot (\mathbf{o}_{ki}^L - \overline{\mathbf{o}_{ki}^L}) \right], \mathbf{h}_k^T \left[\mathbf{s}^S \odot (\mathbf{o}_{ki}^S - \overline{\mathbf{o}_{ki}^S}) \right] \right] = \left[\mathbf{s}^L \mathbf{H}_k \mathbf{Z}^L[:, :], \mathbf{s}^S \mathbf{H}_k \mathbf{Z}^S[:, :] \right] (\sqrt{K-1})$$

Step 0c : $\text{squeeze_average} = [\mathbf{s}_L, \mathbf{s}_S] = \text{mean}([\mathbf{s}^L, \mathbf{s}^S], 1)$, then compute

$$\mathbf{w}_L = \frac{s_L}{\sqrt{s_L^2 + s_S^2}} \text{ and } \mathbf{w}_S = \frac{s_S}{\sqrt{s_L^2 + s_S^2}}, \text{ so that } w_L^2 + w_S^2 = 1$$

Step 1 : Do univariate Gaussian assimilation of y_j using squeezing factor to obtain y_{ji}^a , $i = 1, 2, \dots, K$

$$\begin{aligned} \langle y_j^a \rangle &= \langle y_j^f \rangle + \frac{\left[\mathbf{s}^L \mathbf{Z}^L[\text{omk}, :], \mathbf{s}^S \mathbf{Z}^S[\text{omk}, :] \right] \left[\mathbf{s}^L \mathbf{Z}^L[\text{omk}, :], \mathbf{s}^S \mathbf{Z}^S[\text{omk}, :] \right]^T}{\left[\mathbf{s}^L \mathbf{Z}^L[\text{omk}, :], \mathbf{s}^S \mathbf{Z}^S[\text{omk}, :] \right] \left[\mathbf{s}^L \mathbf{Z}^L[\text{omk}, :], \mathbf{s}^S \mathbf{Z}^S[\text{omk}, :] \right]^T + R_{jj}} \left(y_j^o - \langle y_j^f \rangle \right) \\ &= \langle y_j^f \rangle + \frac{H_j \mathbf{Z}_{\text{omk}}^{\text{squeeze}} \left(H_j \mathbf{Z}_{\text{omk}}^{\text{squeeze}} \right)^T}{H_j \mathbf{Z}_{\text{omk}}^{\text{squeeze}} \left(H_j \mathbf{Z}_{\text{omk}}^{\text{squeeze}} \right)^T + R_{jj}} \left(y_j^o - \langle y_j^f \rangle \right) \end{aligned}$$

$$y_{ji}^a = \langle y_j^a \rangle \left(1 - \frac{H_j \mathbf{Z}_{\text{omk}}^{\text{squeeze}} \left(H_j \mathbf{Z}_{\text{omk}}^{\text{squeeze}} \right)^T}{H_j \mathbf{Z}_{\text{omk}}^{\text{squeeze}} \left(H_j \mathbf{Z}_{\text{omk}}^{\text{squeeze}} \right)^T + R_{jj}} \right)^{1/2} \left[y_{ji}^f - \langle y_j^f \rangle \right], \quad \text{for } i = 1, 2, \dots, 2K$$



OMK (Observation Model Kernel) variables \mathbf{o}_j refers to the minimum list of model variables required to predict the y_j^f observation.

**Multi-Scale LEAKF
(Ens Squeeze Localization)**

Multi-Scale Local MLEnKF using “Ensemble Squeeze Localization” (Pseudo-Code)

Step 2 : Find corresponding analysis ensemble for observations and model variables using

$$y_{ki}^a = y_{ki}^f + \frac{\left\langle \mathbf{h}_k^T \left[\mathbf{w}_L \left(\mathbf{o}_{ki}^L - \overline{\mathbf{o}}_{ki}^L \right), \mathbf{w}_S \left(\mathbf{o}_{ki}^S - \overline{\mathbf{o}}_{ki}^S \right) \right] \left[\mathbf{h}_j^T \left[\mathbf{w}_L \left(\mathbf{o}_{ji}^L - \overline{\mathbf{o}}_{ji}^L \right), \mathbf{w}_S \left(\mathbf{o}_{ji}^S - \overline{\mathbf{o}}_{ji}^S \right) \right] \right]^T \right\rangle}{\left\langle \left[\mathbf{h}_j^T \left[\mathbf{w}_L \left(\mathbf{o}_{ji}^L - \overline{\mathbf{o}}_{ji}^L \right), \mathbf{w}_S \left(\mathbf{o}_{ji}^S - \overline{\mathbf{o}}_{ji}^S \right) \right] \right] \mathbf{h}_j^T \left[\mathbf{w}_L \left(\mathbf{o}_{ji}^L - \overline{\mathbf{o}}_{ji}^L \right), \mathbf{w}_L \left(\mathbf{o}_{ji}^S - \overline{\mathbf{o}}_{ji}^S \right) \right]^T \right\rangle} (y_{ji}^a - y_{ji}^f) \text{ for } k = 1, 2, \dots, p \text{ and } i = 1, 2, \dots, 2K,$$

$$x_{\mu i}^a = x_{\mu i}^f + \frac{\left\langle \left[\mathbf{w}_L \left(x_{\mu i}^{Lf} - \overline{x}_{\mu i}^{Lf} \right), \mathbf{w}_S \left(x_{\mu i}^{Sf} - \overline{x}_{\mu i}^{Sf} \right) \right], \left[\mathbf{w}_L \mathbf{h}_j^T \left(\mathbf{o}_{ji}^L - \overline{\mathbf{o}}_{ji}^L \right), \mathbf{w}_S \mathbf{h}_j^T \left(\mathbf{o}_{ji}^S - \overline{\mathbf{o}}_{ji}^S \right) \right]^T \right\rangle}{\left\langle \left[\mathbf{h}_j^T \left[\mathbf{w}_L \left(\mathbf{o}_{ji}^L - \overline{\mathbf{o}}_{ji}^L \right), \mathbf{w}_S \left(\mathbf{o}_{ji}^S - \overline{\mathbf{o}}_{ji}^S \right) \right] \right] \mathbf{h}_j^T \left[\mathbf{w}_L \left(\mathbf{o}_{ji}^L - \overline{\mathbf{o}}_{ji}^L \right), \mathbf{w}_S \left(\mathbf{o}_{ji}^S - \overline{\mathbf{o}}_{ji}^S \right) \right]^T \right\rangle} (y_{ji}^a - y_{ji}^f) \text{ for } \mu = 1, 2, \dots, n, \text{ and } i = 1, 2, \dots, K$$

Step 3 : Let the analysis ensemble perturbation be the prior ensemble perturbations for the next observation

$$\mathbf{Z} = \frac{1}{\sqrt{2K-1}} (\mathbf{x}^a - \mathbf{x}^f)$$

Step 4 : Let the analysis ensemble be the prior ensemble for the next observation

$$y_{ki}^f = y_{ki}^a, \text{ for } k = 1, 2, \dots, p; i = 1, 2, \dots, 2K$$

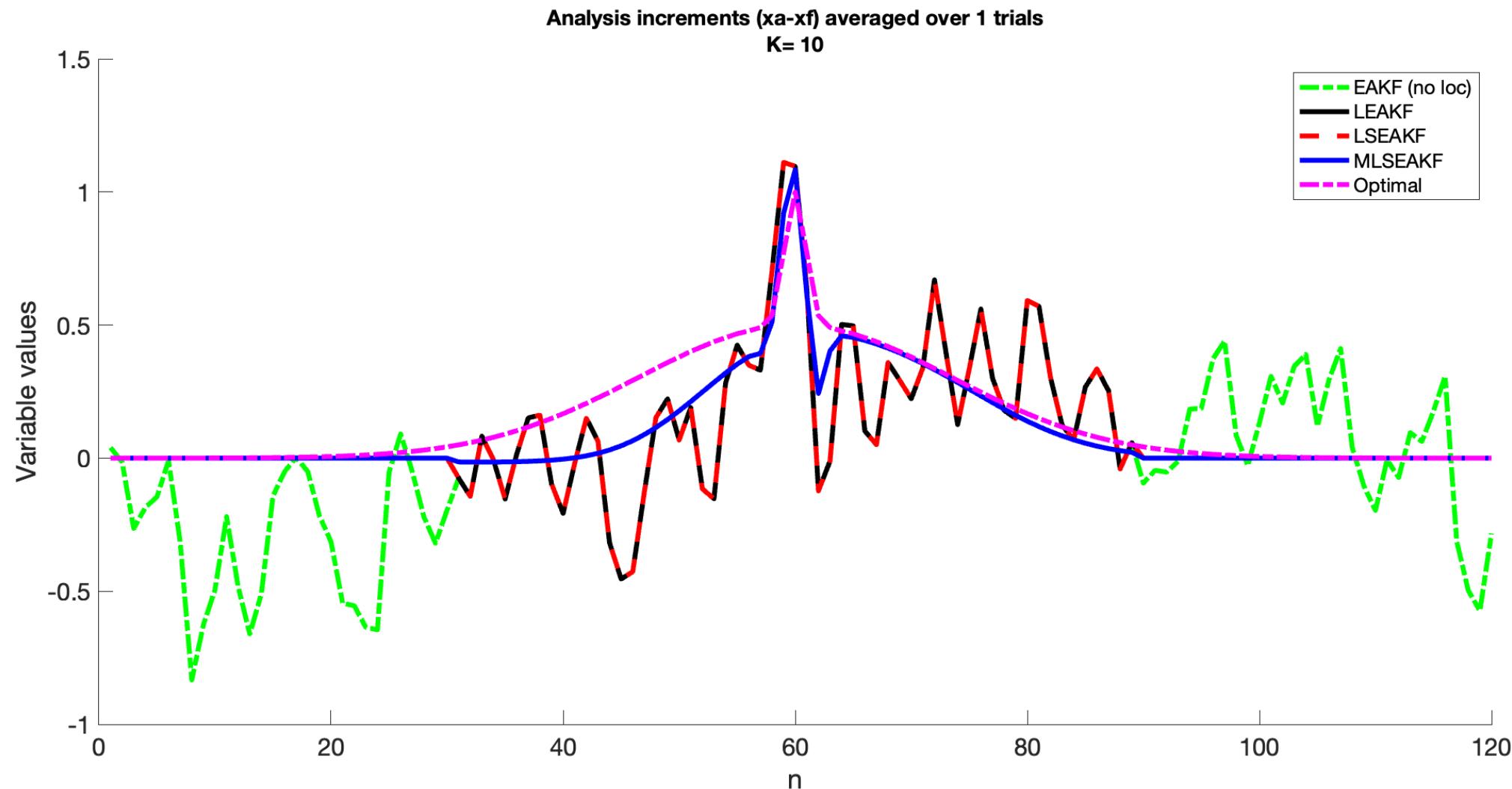
$$x_{\mu i}^f = x_{\mu i}^a, \text{ for } \mu = 1, 2, \dots, n; i = 1, 2, \dots, 2K$$

end

$$\mathbf{X}^a = \mathbf{x}^a + \sqrt{K-1} ((\mathbf{E} \Lambda_L \mathbf{E}^T) \mathbf{Z}^L + (\mathbf{E} \Lambda_S \mathbf{E}^T) \mathbf{Z}^S)$$

Results

Single Ob DA Correction (Optimal)



Comparison Single-Scale LEnKF and Multi-Scale MLEnKF with very accurate observations

1-dim Multi-Scale statistical model:

K = 10

n = 120

p = 30

ntrials = 200

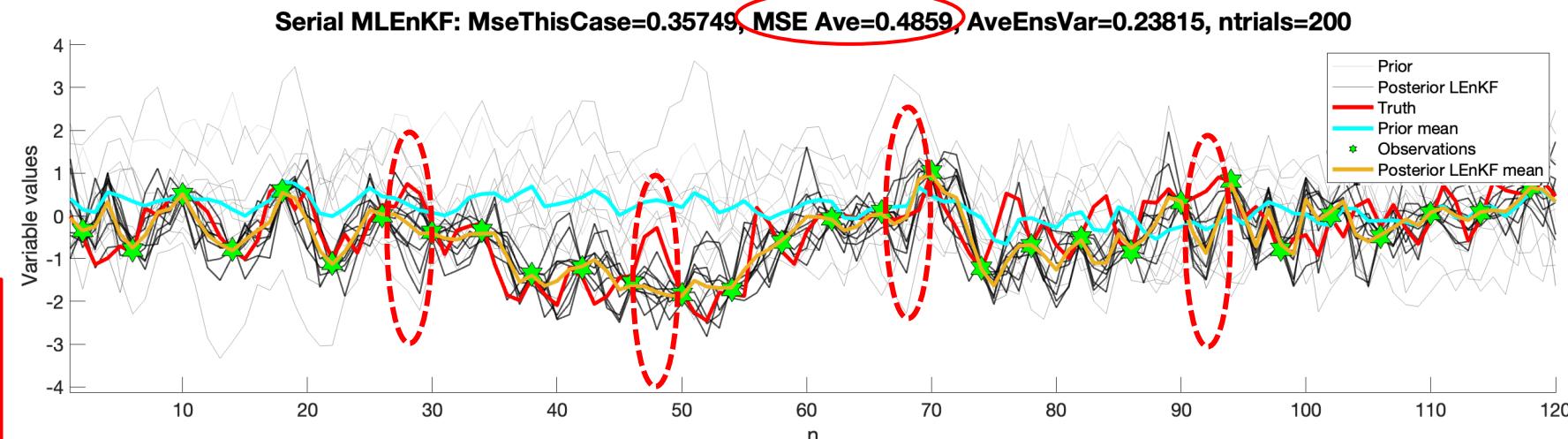
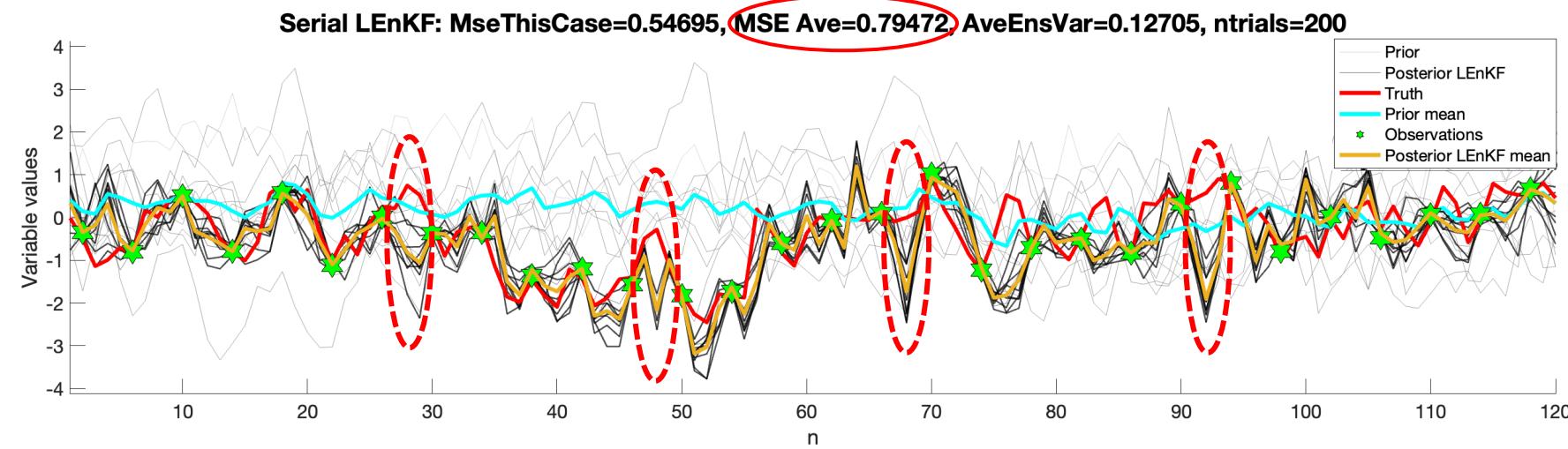
Ob error var = 0.01

Loc. single-scale = 7

Loc. Large multi-scale = 20

Loc. Small multi-scale = 2

**Multi-Scale LEAKF (MSE=0.4859)
beats the single-scale LEAKF
(MSE=0.79472)**



Comparison Single-Scale LEnKF and Multi-Scale MLEnKF with very accurate observations

1-dim Multi-Scale statistical model:

$K = 10$

$n = 120$

$p = 60$

$n_{trials} = 200$

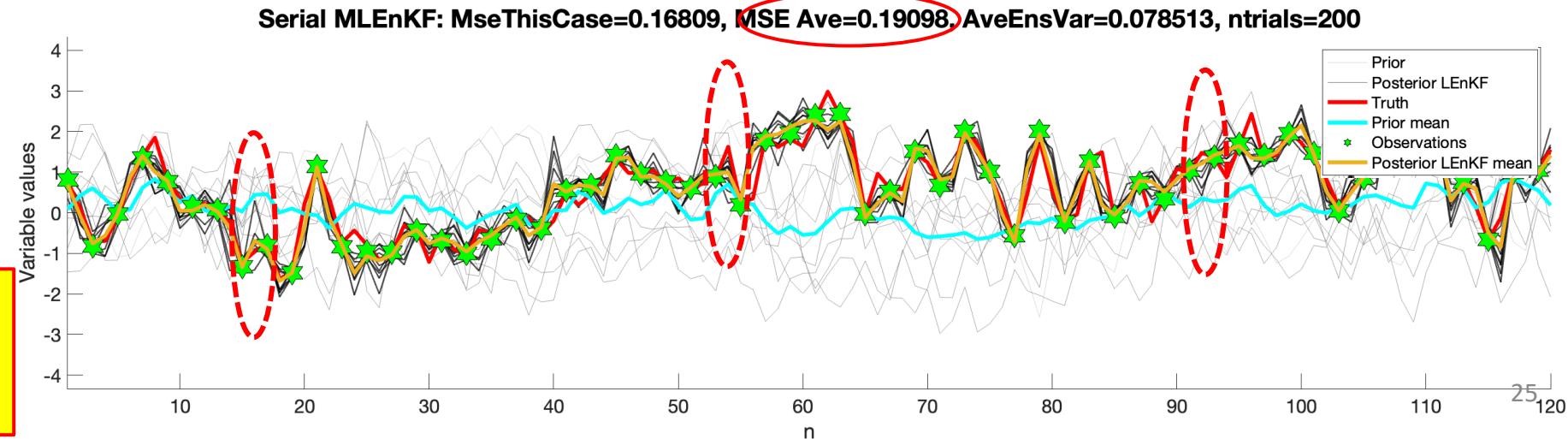
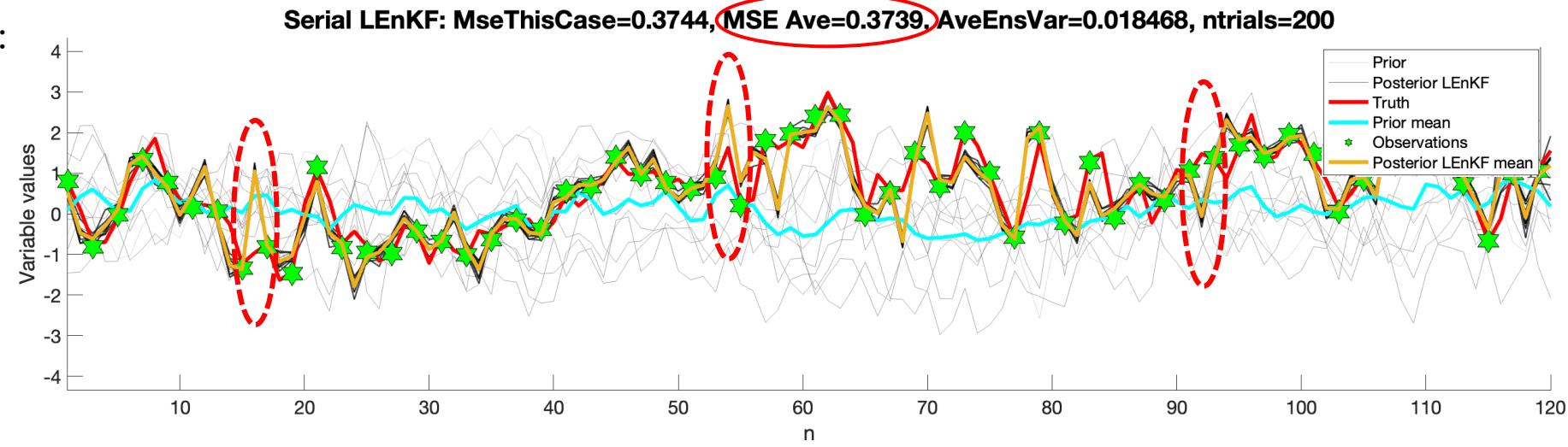
Ob error var = 0.01

Loc. single-scale = 7

Loc. Large multi-scale = 20

Loc. Small multi-scale = 2

Multi-Scale LEAKF (MSE=0.19098)
beats the single-scale LEAKF
(MSE=0.3739)



Comparison Single-Scale LEnKF and Multi-Scale MLEnKF with less accurate observations

1-dim Multi-Scale statistical model:

K = 10

n = 120

p = 30

ntrials = 200

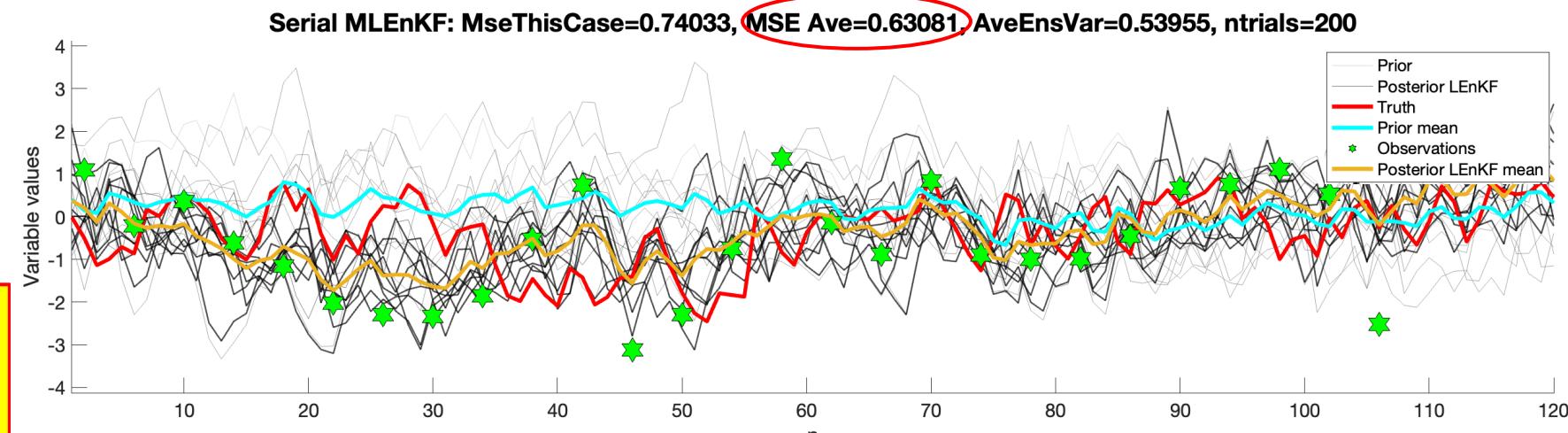
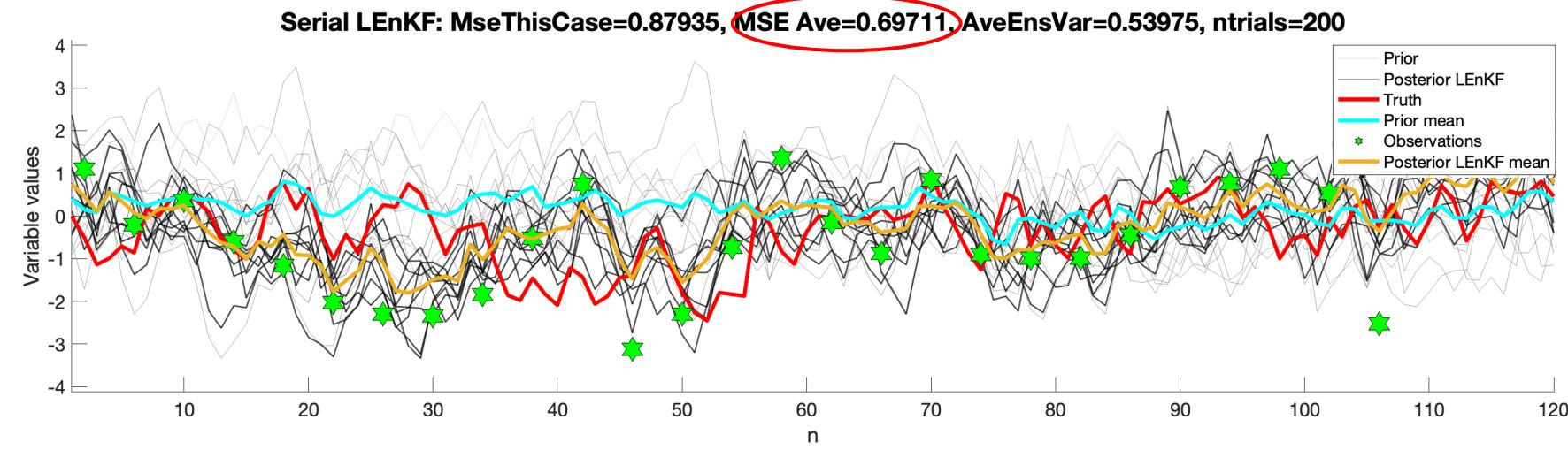
Ob error var = 1.0

Loc. single-scale = 7

Loc. Large multi-scale = 20

Loc. Small multi-scale = 2

**Multi-Scale LEAKF (MSE=0.63081)
beats the single-scale LEAKF
(MSE=0.69711)**



Comparison Single-Scale LEnKF and Multi-Scale MLEnKF with less accurate observations

1-dim Multi-Scale statistical model:

K = 10

n = 120

p = 60

ntrials = 200

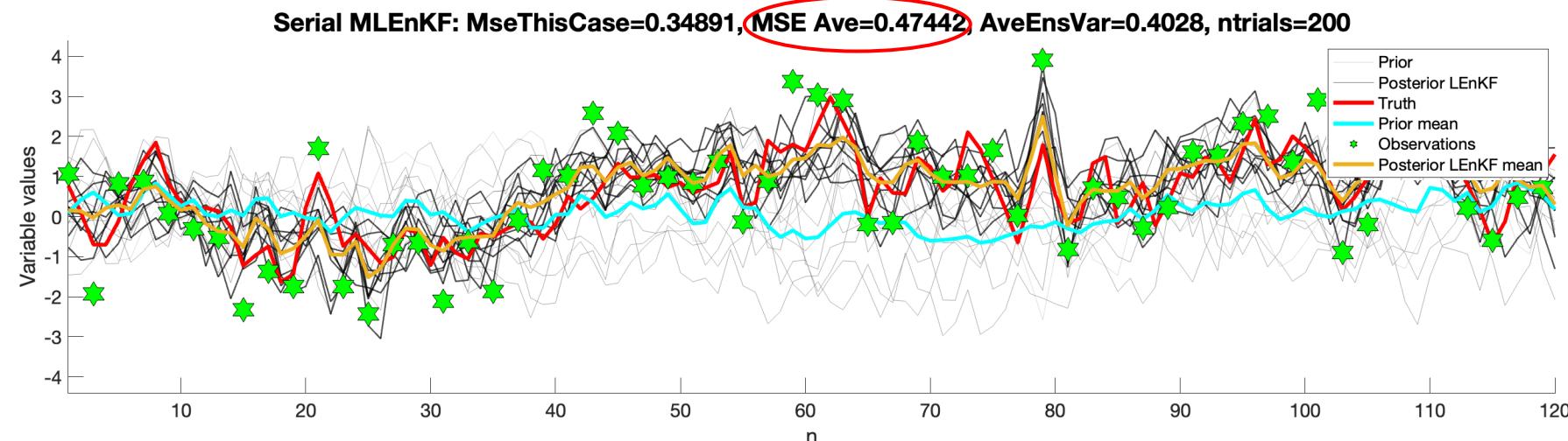
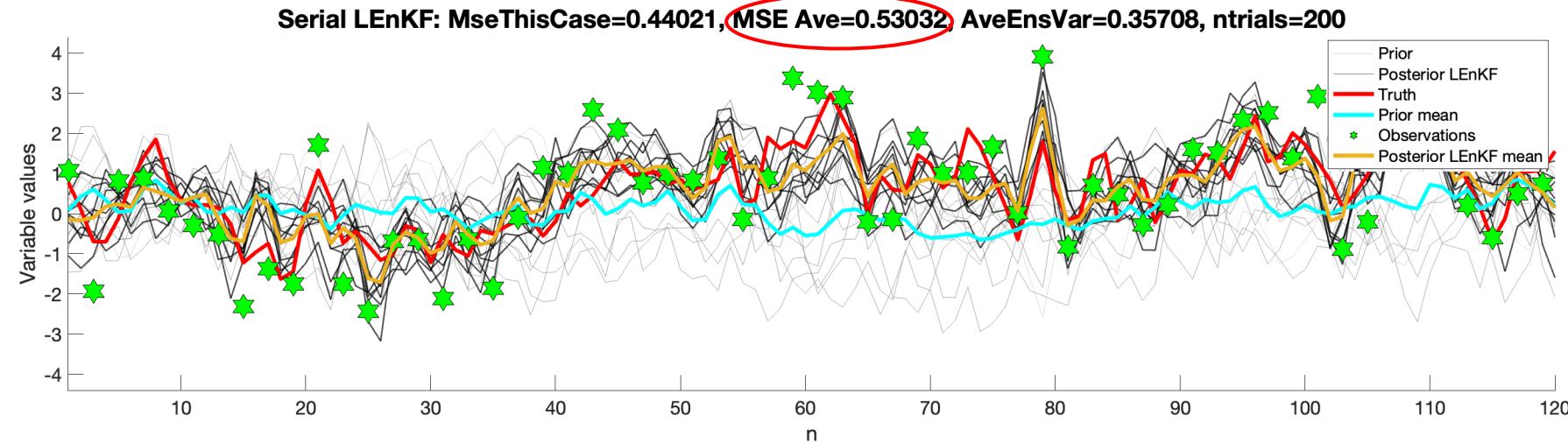
Ob error var = 1.0

Loc. single-scale = 7

Loc. Large multi-scale = 20

Loc. Small multi-scale = 2

**Multi-Scale LEAKF (MSE=0.47442)
beats the single-scale LEAKF
(MSE=0.53032)**



Concluding Remarks

- “*Ensemble Squeeze*” is equivalent to the “*Ob-Error-Inflation*” localization for same computational cost.
- “*Ensemble Squeeze*” localization works in a **serial framework** where we assimilate one observation after the other and it is **SIMPLER** to implement than other model space localization methods (e.g., Shunji et al, 2021).
- Preliminary results suggests that the **multi-scale DA** using the “*Ensemble Squeeze*” seems to produce **similar percentage error reduction** to the ones observed in Xuguang et al., 2020.

Potential use of the “*Ensemble Squeeze*” localization:

- “*Ensemble Squeeze*” localization could be useful for observations that are integral of the model state (**e.g., Satellite DA**)

FOR MORE DETAILS, PLEASE TALK WITH US LATER!!

Acknowledgements

The first author acknowledges the *Ministerio de Universidades (Plan de Recuperación, Transformación y Resiliencia)* funded by the European Union (**NextGenerationEU**), with the participation of the *University of the Balearic Islands*, which supports his current postdoctoral position as a *Maria Zambrano Fellowship*.



Financiado por
la Unión Europea
NextGenerationEU



Plan de Recuperación,
Transformación y Resiliencia



MINISTERIO
DE UNIVERSIDADES