Sampling error in the ensemble Kalman filter for small ensembles and high-dimensional states

- Chris Snyder, NSF NCAR, Boulder CO

This material is based on work supported by NSF NCAR, which is a major facility sponsored by the NSF under Cooperative Agreement 1852977.

A Map for This Talk

- notation, the Kalman filter (KF), and the ensemble Kalman filter (EnKF)
- previous results for work on error in the EnKF
- two tools: an optimal linear transformation & results for ”tall, skinny” random matrices
- sampling error in the EnKF for small ensemble size

Preliminaries

We wish to estimate the state $\mathbf{x}$ given observations $\mathbf{y}$.

$$\mathbf{x} = \text{discretized representation of atmosphere or other system}$$
$$\mathbf{y} = \text{concatenation of available measurements of the system}$$

$$N_x = \dim \mathbf{x}, \qquad N_y = \dim \mathbf{y}$$

[We can also concatenate different times into $\mathbf{x}$ and $\mathbf{y}$. All results today will apply to that case too.]

The Kalman Filter

Given: $\mathbf{x} \sim N(\mathbf{x}^f, \mathbf{P}), \quad \mathbf{y} = \mathbf{H}\mathbf{x} + \epsilon, \quad \epsilon \sim N(0, \mathbf{R})$.

Then $\mathbf{x}|\mathbf{y} \sim N(\mathbf{x}^a, \mathbf{P}^a)$, where

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

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P},$$

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

The Ensemble Kalman Filter (EnKF)

Work with forecast, analysis ensembles instead of **P**, $\mathbf{P}^a$

- storage and computations are feasible for ensemble size $N_e = 100$

The Ensemble Kalman Filter (EnKF)

Work with forecast, analysis ensembles instead of $\mathbf{P}$, $\mathbf{P}^a$

- storage and computations are feasible for ensemble size $N_e = 100$

Approximate covariances in KF by sample covariances (Evensen 1994)

- $\mathbf{P}$, $\mathbf{PH}^T$, $\mathbf{HPH}^T$ estimated from ensemble of forecasts at each analysis time
- generate ensemble of analyses, consistent with KF update

Ensemble Notation

Begin from $\{\mathbf{x}^i,\ i = 1, \ldots, N_e\}$, an ensemble drawn from $p(\mathbf{x})$.

$$\hat{\mathbf{x}} = {N_e}^{-1} \sum_{i=1}^{N_e} \mathbf{x}^i, \qquad \delta\mathbf{x}^i = \mathbf{x}^i - \hat{\mathbf{x}}$$

$$\mathbf{X} = (N_e - 1)^{-1/2} \left[\delta\mathbf{x}^1, \ldots, \delta\mathbf{x}^{N_e}\right]$$

$$\hat{\mathbf{P}} = \mathbf{X}\mathbf{X}^T = (N_e - 1)^{-1} \sum_{i=1}^{N_e} \delta\mathbf{x}^i {\delta\mathbf{x}^i}^T$$

EnKF Update Equations

EnKF is the KF with $\mathbf{P}$ replaced by $\hat{\mathbf{P}}$:

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

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

The analysis ensemble satisfies

$$\hat{\mathbf{P}}^a = (\mathbf{I} - \hat{\mathbf{K}}\mathbf{H})\hat{\mathbf{P}}.$$

[For "stochastic" EnKFs, this form of $\hat{\mathbf{P}}^a$ holds for expectation over realizations of the algorithm]

Analysis Errors for the EnKF

For the KF, expected squared error of $\mathbf{x}^a$ is given by

$$\mathbf{P}^a = \text{cov}(\mathbf{x}) = E\left((\mathbf{x} - \mathbf{x}^a)(\mathbf{x} - \mathbf{x}^a)^T\right)$$

The EnKF estimate of the expected squared analysis errors is

$$\hat{\mathbf{P}}^a = (\mathbf{I}_x - \hat{\mathbf{K}}\mathbf{H})\hat{\mathbf{P}},$$

and its analysis mean $\hat{\mathbf{x}}^a$ has expected squared errors:

$$\mathbf{A} = (\mathbf{I}_x - \hat{\mathbf{K}}\mathbf{H})\mathbf{P}(\mathbf{I}_x - \hat{\mathbf{K}}\mathbf{H})^T + \hat{\mathbf{K}}\mathbf{R}\hat{\mathbf{K}}$$

Tool #1: Optimal Linear Transformation

Helpful to work in the transformed coordinates [Snyder and Hakim 2022]:

$$\mathbf{x}' = \mathbf{V}^T \mathbf{P}^{-1/2} \mathbf{x}, \quad \mathbf{y}' = \mathbf{U}^T \mathbf{R}^{-1/2} \mathbf{y},$$

where columns of **U** and **V** contain singular vectors of

$$\tilde{\mathbf{H}} = \mathbf{R}^{-1/2} \mathbf{H} \mathbf{P}^{1/2} = \mathbf{U} \Lambda \mathbf{V}^T$$

Simplifications from Optimal Coordinates

In the transformed variables, KF is very simple

Covariances are identity matrices: $\mathbf{x}' \sim N(\mathbf{x}'^f, \mathbf{I}_x), \quad \epsilon \sim N(0, \mathbf{I}_y)$.

Observation operator is diagonal: $\mathbf{y}' = \Lambda \mathbf{x}' + \epsilon$, i.e., $y'_i = \lambda_i x'_i + \epsilon_i$.
Call λ_i the ith *canonical observation operator* (COO).

Gain is (rectangular) diagonal:

$$\mathbf{K} = \Lambda^T(\Lambda\Lambda^T + \mathbf{I}_y)^{-1}, \quad \text{i.e., } K_i = \lambda_i/(\lambda_i^2 + 1).$$

Updated (posterior) covariance is diagonal:

$$\mathbf{P}^a = \mathbf{I}_x - \Lambda^T(\Lambda\Lambda^T + \mathbf{I}_y)^{-1}\Lambda, \quad \text{i.e., } \mathrm{var}(x'_i|\mathbf{y}') = 1 - \lambda_i^2/(\lambda_i^2 + 1).$$

Importance of the COOs

The update depends only on the COOs $\{\lambda_i,\, i = 1, \ldots, N\}$.

Properties of the update that are independent of linear transformations (such as those related to information) are completely characterized by the COOs:

- degrees of freedom for signal (Rodgers 2000)
- mutual information of state and observations (Rodgers 2000, Xu 2007)
- conditioning of minimization, for either "**B**" or "**R**" preconditioning (Courtier 1997)
- minimal ensemble size required for particle filters (Snyder et al 2008)
- optimal low-rank approximations to update (Spantini et al 2015, Aulігné et al 2016, Bousserez and Henze 2018, Zupanski 2021)

EnKF in Optimal Coordinates

$$\hat{\mathbf{x}}^a = \hat{\mathbf{x}} + \hat{\mathbf{K}}(\mathbf{y} - \Lambda\hat{\mathbf{x}})$$

$$\hat{\mathbf{P}}^a = (\mathbf{I}_x - \hat{\mathbf{K}}\Lambda)\hat{\mathbf{P}}$$

$$\hat{\mathbf{K}} = \hat{\mathbf{P}}\Lambda^T(\Lambda\hat{\mathbf{P}}\Lambda^T + \mathbf{I}_y)^{-1}.$$

EnKF in Optimal Coordinates

$$\hat{\mathbf{x}}^a = \hat{\mathbf{x}} + \hat{\mathbf{K}}(\mathbf{y} - \Lambda\hat{\mathbf{x}})$$

$$\hat{\mathbf{P}}^a = (\mathbf{I}_x - \hat{\mathbf{K}}\Lambda)\hat{\mathbf{P}}$$

$$\hat{\mathbf{K}} = \hat{\mathbf{P}}\Lambda^T(\Lambda\hat{\mathbf{P}}\Lambda^T + \mathbf{I}_y)^{-1}.$$

And,

$$\mathbf{A} = (\mathbf{I}_x - \hat{\mathbf{K}}\Lambda)(\mathbf{I}_x - \hat{\mathbf{K}}\Lambda)^T + \hat{\mathbf{K}}\hat{\mathbf{K}}^T.$$

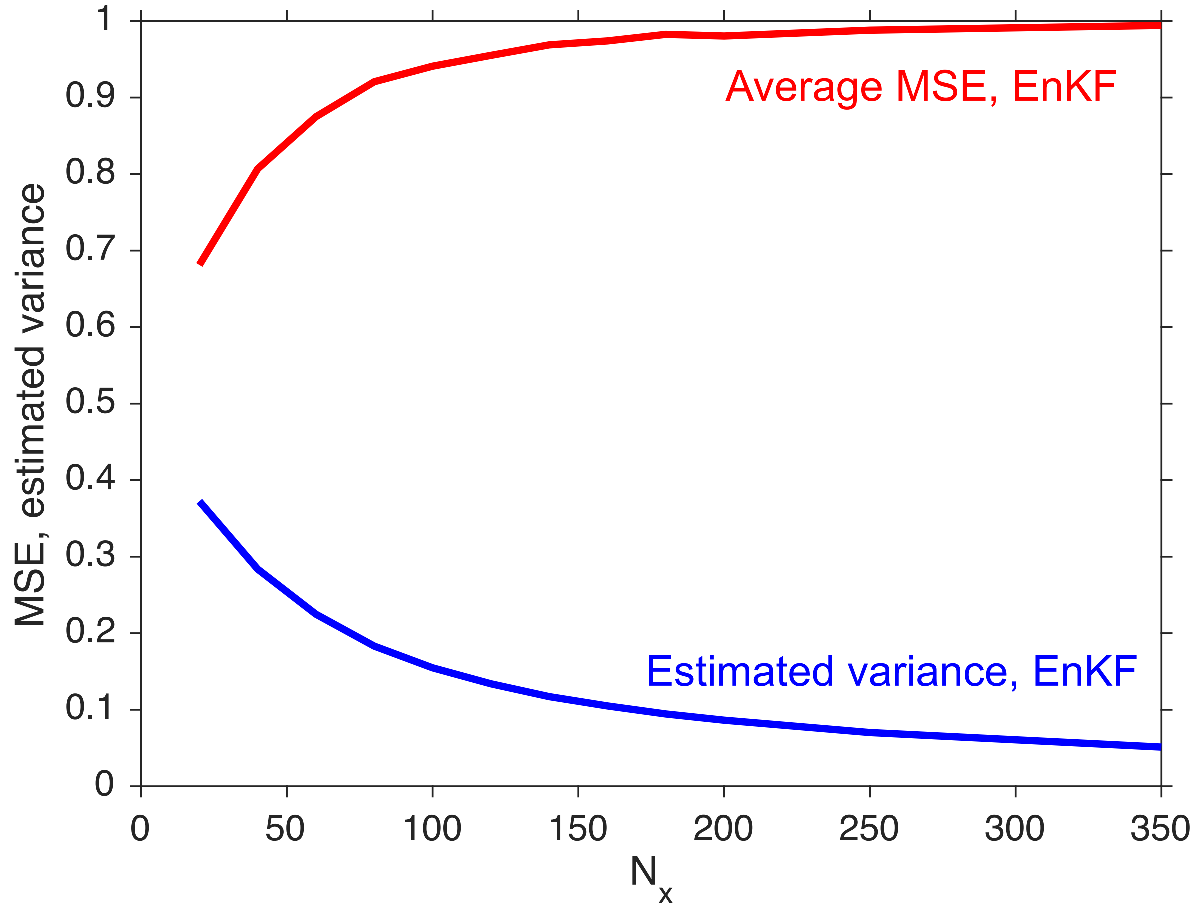
Sampling Error in the EnKF

Sample covariances have error of $O({N_e}^{-1/2})$

Sampling error is fundamental limitation on EnKF

- sample covariance matrices are rank deficient ("the rank problem")
- where correlations are small, covariances are swamped by noise
- How do such errors propagate through the algorithm?

An easy example: $\mathbf{P} = \mathbf{R} = \mathbf{H} = \mathbf{I}$



Sampling Error in the EnKF

Previous studies

- EnKF biased toward overconfidence: posterior covariance is too small
  (“in-breeding”, Houtekamer and Mitchell 1998; also van Leeuwen 1999)
- expansions for small sampling error, implicitly considering large ensembles
  (van Leeuwen 1999, Sacher and Bartello 2007)
- analysis or examples for scalar state (Whitaker and Hamill 2002, Sacher and Bartello 2007)
- analysis of pairwise update, i.e. single ob, single state variable
  (Anderson 2007 and sequels)
- ensemble size giving bounded error when $\mathbf{H} = \mathbf{I}_x$ (Furrer and Bengtsson 2007)
- explicit expression for $p(||\mathbf{x}^a - \hat{\mathbf{x}}^a||^2)$ (Kovalenko et al 2011)

Why Revisit EnKF Sampling Error?

- Seek explicit results for small ensembles, high-dimensional state and obs
- Clarify relative roles of state dimension, obs dimension, details of obs network

Tool #2: A High-Dimensional Approximation

Let $\mathbf{Z} = \Lambda \mathbf{X}$. The key approximation is

$$\mathbf{Z}^T\mathbf{Z} \approx \frac{b^2}{N_e - 1}\left(\mathbf{I}_e - {N_e}^{-1}\mathbf{1}\right), \quad \text{with } b^2 = \sum_{i=1}^{N} \lambda_i^2.$$

When valid, $\mathbf{Z}^T\mathbf{Z}$ acts approximately as a scalar multiple of the identity

(at least for "zero mean" vectors whose components sum to zero)

Tool #2: A High-Dimensional Approximation

Let $\mathbf{Z} = \Lambda\mathbf{X}$. The key approximation is

$$\mathbf{Z}^T\mathbf{Z} \approx \frac{b^2}{N_e - 1}\left(\mathbf{I}_e - {N_e}^{-1}\mathbf{1}\right), \quad \text{with } b^2 = \sum_{i=1}^{N} \lambda_i^2.$$

When valid, $\mathbf{Z}^T\mathbf{Z}$ acts approximately as a scalar multiple of the identity

(at least for "zero mean" vectors whose components sum to zero)

Heuristic from random-matrix lit: Tall matrices are isometries (e.g. Vershynin 2012)

Tool #2: A High-Dimensional Approximation

Intuition for approximation:

$(\mathbf{Z}^T\mathbf{Z})_{ij}$ is inner product of ith and jth perturbations (in obs space). Inner products are sums over many terms and become less variable when N is large.

Tool #2: A High-Dimensional Approximation

Intuition for approximation:

$(\mathbf{Z}^T\mathbf{Z})_{ij}$ is inner product of ith and jth perturbations (in obs space). Inner products are sums over many terms and become less variable when N is large.

More rigorously,

$$\text{var}(\mathbf{Z}^T\mathbf{Z})_{ij} \propto c^4/(N_e-1)^2, \quad \text{where } c^4 = \sum_{i=1}^{N} \lambda_i^4,$$

so elementwise approximation holds if c^2/b^2 is small.

Tool #2: A High-Dimensional Approximation

Intuition for approximation:

$(\mathbf{Z}^T\mathbf{Z})_{ij}$ is inner product of ith and jth perturbations (in obs space). Inner products are sums over many terms and become less variable when N is large.

More rigorously,

$$\text{var}(\mathbf{Z}^T\mathbf{Z})_{ij} \propto c^4/(N_e - 1)^2, \quad \text{where } c^4 = \sum_{i=1}^{N} \lambda_i^4,$$

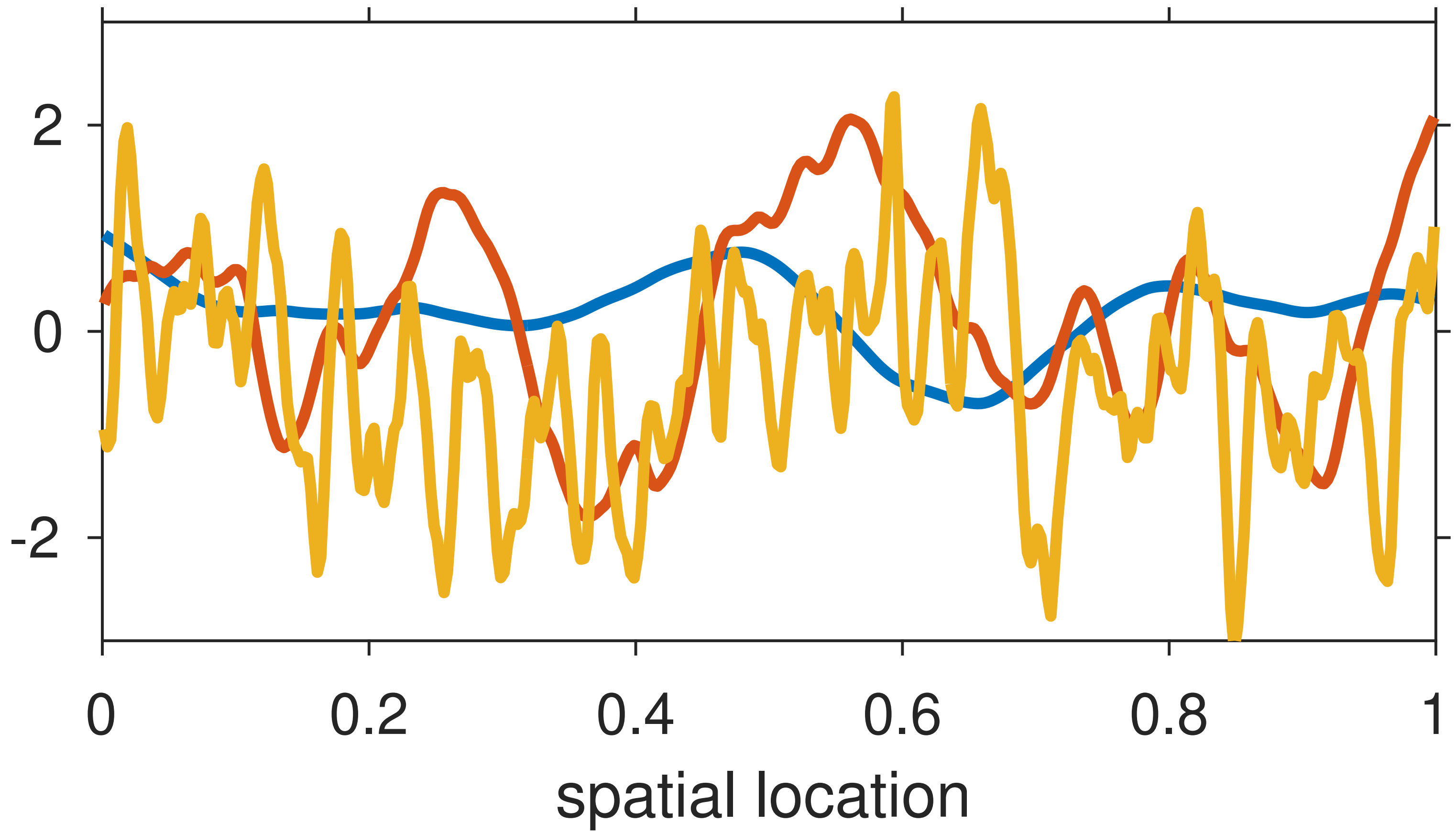
so elementwise approximation holds if c^2/b^2 is small.

Think of b^4/c^4 as an effective dimension: equals N if $\lambda_i = \text{const}$ and equals 1 in limit that λ_1^2 dominates sum.

An Idealized Example

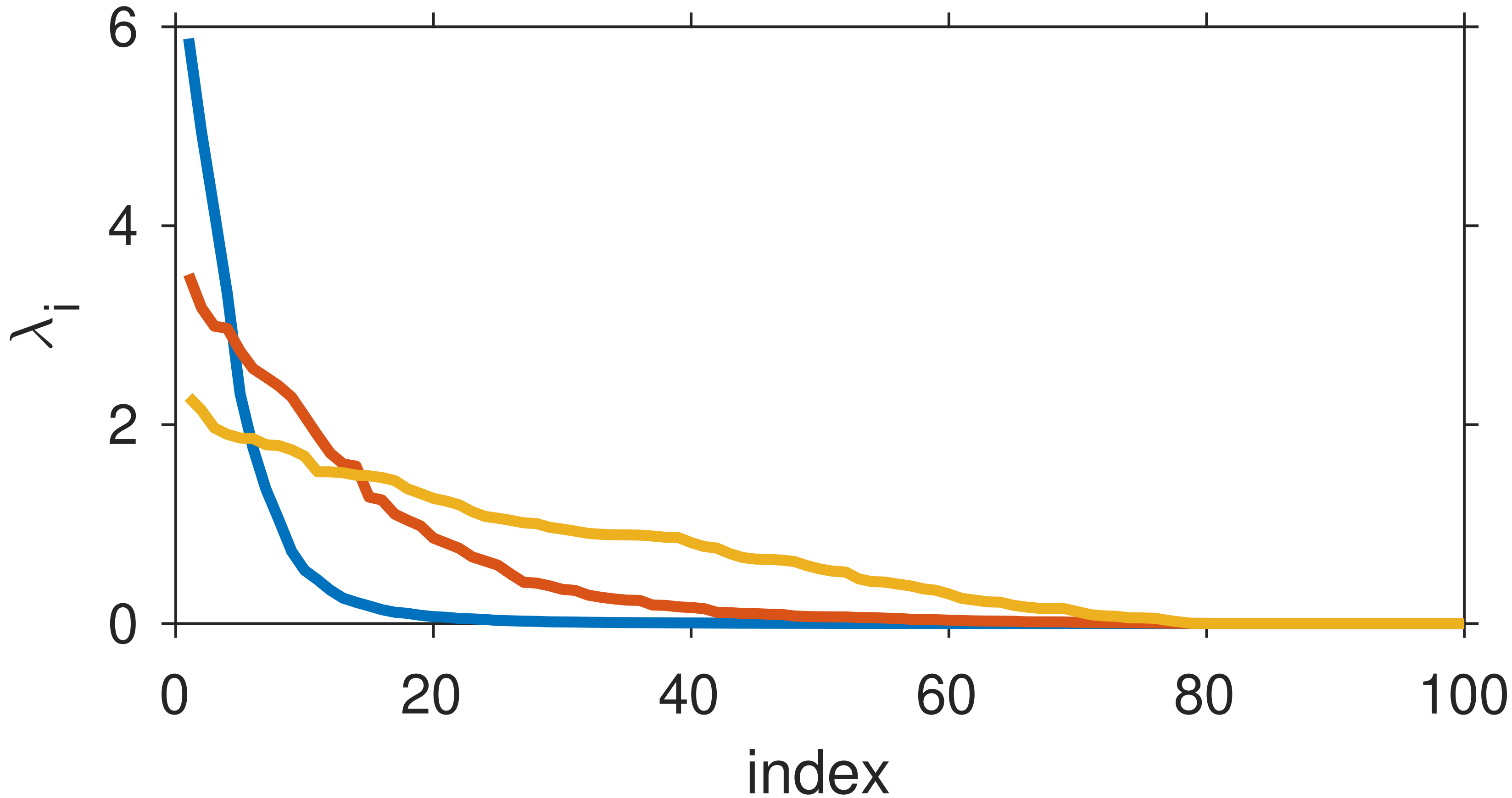
1D spatial example: homog. prior covariance + point obs at random locations

Consider 100 observations with iid errors $\sim N(0, 1)$



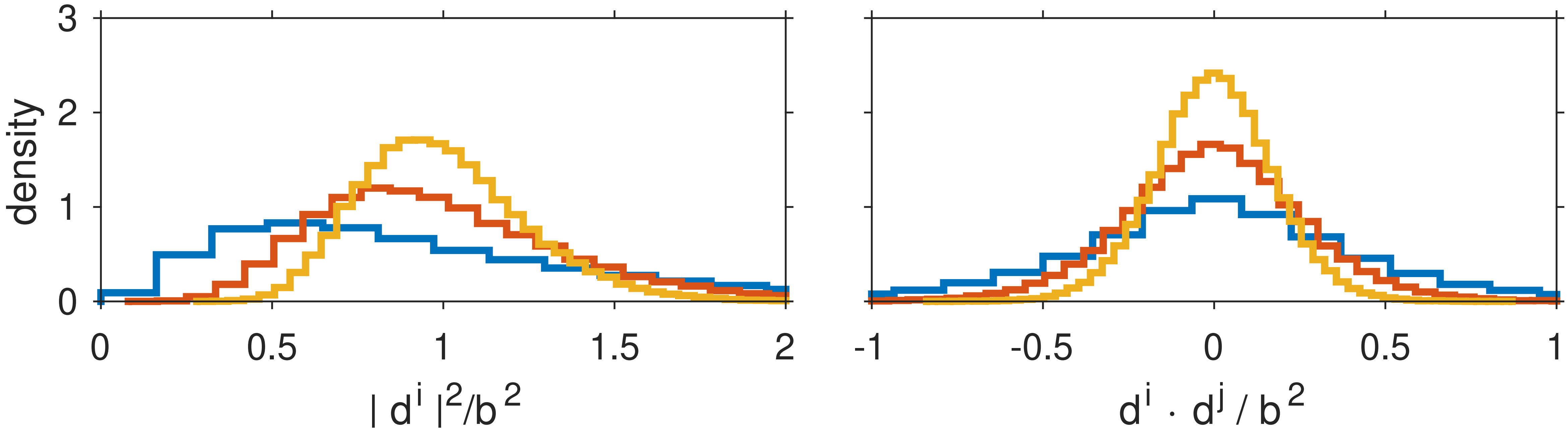
An Idealized Example

COOs when state has long, medium, or short spatial correlations (effective dimensions are 4.43, 14.4, and 33, respectively)



An Idealized Example

Histograms for diagonal (left) and off-diagonal (right) elements of $\mathbf{Z}^T\mathbf{Z}$



Tool #2 (cont.)

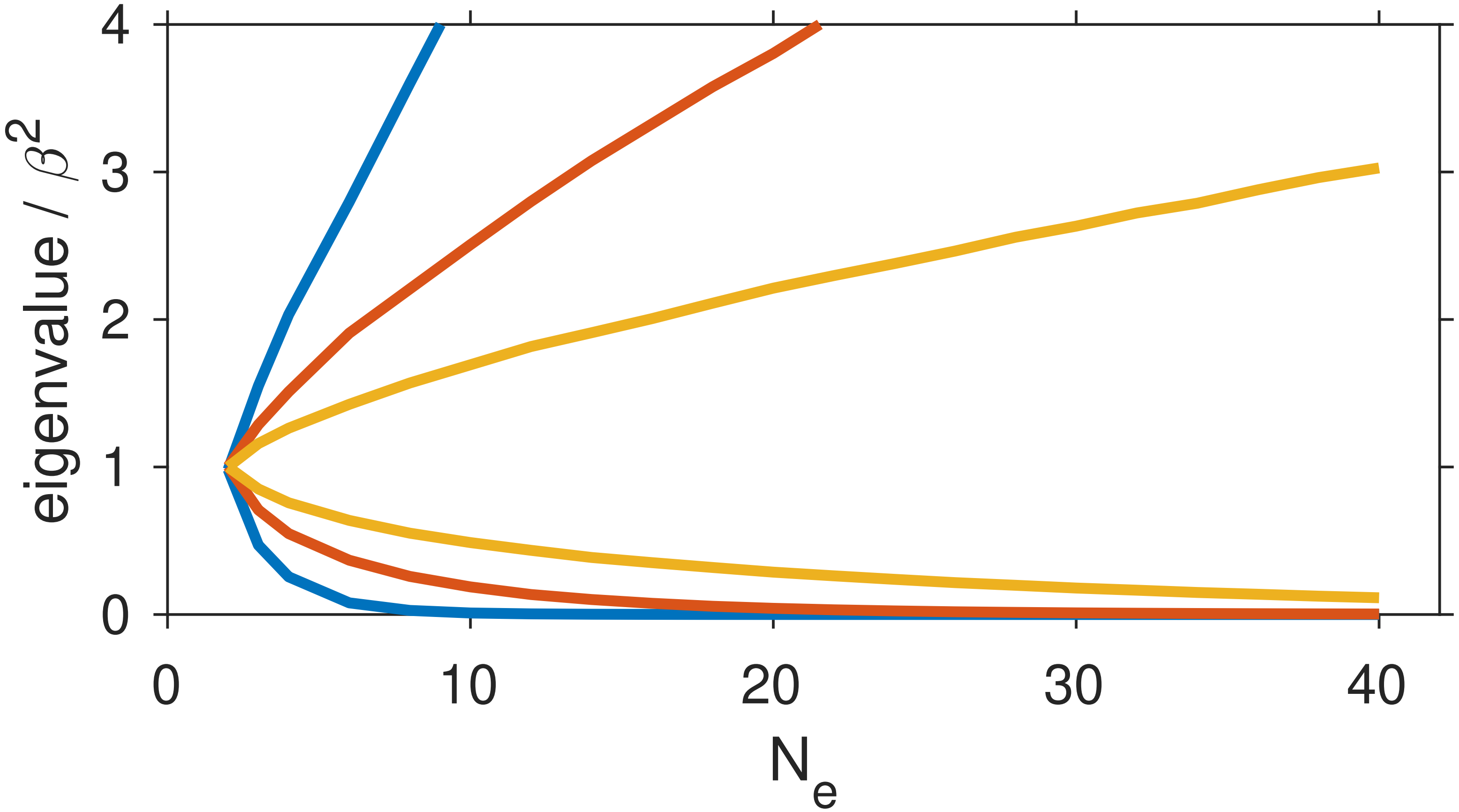
Also need eigenvalues of $\mathbf{Z}^T\mathbf{Z}$ to cluster around $\beta^2 = b^2/(N_e - 1)$.

(The approximation, when vaild, says that every direction in the ensemble subspace is equivalent and carries a variance of β^2.)

That clustering requires N_e small compared to the effective dimension, in addition to large effective dimension. (See Marchenko and Pastur 1967)

An Idealized Example

Maximum and minimum eigenvalues of $\mathbf{Z}^T\mathbf{Z}$ as a function of N_e.



The Effects of Sampling Error

Now ready to apply the tools, i.e. write EnKF in optimal coordinates and apply approximation of $\mathbf{Z}^T\mathbf{Z}$. Lots of good things happen.

Let $\beta^2 = b^2/(N_e - 1)$ and consider EnKF gain as an example:

$$\hat{\mathbf{K}} = \mathbf{X}\mathbf{Z}^T(\mathbf{Z}\mathbf{Z}^T + \mathbf{I}_y)^{-1}.$$

The Effects of Sampling Error

Now ready to apply the tools, i.e. write EnKF in optimal coordinates and apply approximation of $\mathbf{Z}^T\mathbf{Z}$. Lots of good things happen.

Let $\beta^2 = b^2/(N_e - 1)$ and consider EnKF gain as an example:

$$\hat{\mathbf{K}} = \mathbf{X}\mathbf{Z}^T(\mathbf{Z}\mathbf{Z}^T + \mathbf{I}_y)^{-1}.$$

$$(\mathbf{Z}\mathbf{Z}^T + \mathbf{I}_y)(a\mathbf{Z}\mathbf{Z}^T + \mathbf{I}_y) = \mathbf{I}_y \quad \Rightarrow \quad a = -(\beta^2 + 1)^{-1}$$

The Effects of Sampling Error

Now ready to apply the tools, i.e. write EnKF in optimal coordinates and apply approximation of $\mathbf{Z}^T\mathbf{Z}$. Lots of good things happen.

Let $\beta^2 = b^2/(N_e - 1)$ and consider EnKF gain as an example:

$$\hat{\mathbf{K}} = \mathbf{X}\mathbf{Z}^T(\mathbf{Z}\mathbf{Z}^T + \mathbf{I}_y)^{-1}.$$

$$(\mathbf{Z}\mathbf{Z}^T + \mathbf{I}_y)(a\mathbf{Z}\mathbf{Z}^T + \mathbf{I}_y) = \mathbf{I}_y \quad \Rightarrow \quad a = -(\beta^2 + 1)^{-1}$$

$$\hat{\mathbf{K}} = \mathbf{X}\mathbf{Z}^T(a\mathbf{Z}\mathbf{Z}^T + \mathbf{I}_y) = \mathbf{X}(a\beta^2\mathbf{Z}^T + \mathbf{Z}^T) = (\beta^2 + 1)^{-1}\mathbf{X}\mathbf{Z}^T$$

Effects on Mean Update

Obs-space analysis increment for mean is then

$$\begin{aligned}\Lambda(\hat{\mathbf{x}}^a - \hat{\mathbf{x}}) &= (\beta^2 + 1)\mathbf{Z}\mathbf{Z}^T(\mathbf{y} - \Lambda\hat{\mathbf{x}}) \\ &= (\beta^2 + 1)\mathbf{Z}\mathbf{Z}^T\mathbf{Z}\mathbf{a} \\ &\approx \beta^2(\beta^2 + 1)^{-1}\mathbf{Z}\mathbf{a} \qquad (1)\end{aligned}$$

Only projection of $\mathbf{y} - \Lambda\hat{\mathbf{x}}$ onto ensemble subspace matters to increment.

Gain in ensemble subspace is $\beta^2/(\beta^2 + 1)$, so analysis fits that projection of obs almost exactly

Effects on EnKF Analysis Ensemble

Continuing with similar manipulations leads to

$$\hat{\mathbf{P}}^a = (\beta^2 + 1)^{-1}\hat{\mathbf{P}}$$

EnKF retains little analysis variance—when N/N_e is large, β^2 will be large unless the obs are uniformative or redundant (COOs small)

Consider $\lambda_i = 1$ (as in case with $\mathbf{P} = \mathbf{H} = \mathbf{R} = \mathbf{I}$). Then EnKF analysis reduces variance by factor of approximately N_e/N, while KF reduces by factor of 1/2.

Effects on Analysis Errors

Finally, using the approximation for $\mathbf{Z}^T\mathbf{Z}$, the analysis-error covariance becomes

$$\mathbf{A} = \mathbf{I}_x - (\beta^2+1)^{-1}\left(\hat{\mathbf{P}}\Lambda^T\Lambda + \Lambda^T\Lambda\hat{\mathbf{P}}\right) + (\beta^2+1)^{-2}(\beta^2+\gamma^4)\hat{\mathbf{P}},$$

whose diagonal entries are

$$a_{ii} = \begin{cases} 1 + (\beta^2+1)^{-2}\left(\gamma^4+\beta^2-2(\beta^2+1)\lambda_i^2\right)\hat{p}_{ii}, & i \le N \\ 1 + (\beta^2+1)^{-2}\left(\gamma^4+\beta^2\right)\hat{p}_{ii} & i > N \end{cases}$$

Note $\gamma^4 = (N_e - 1)^{-1}\sum\lambda^4$.

Effects on Analysis Errors

Finally, using the approximation for $\mathbf{Z}^T\mathbf{Z}$, the analysis-error covariance becomes

$$\mathbf{A} = \mathbf{I}_x - (\beta^2+1)^{-1}\left(\hat{\mathbf{P}}\Lambda^T\Lambda + \Lambda^T\Lambda\hat{\mathbf{P}}\right) + (\beta^2+1)^{-2}(\beta^2+\gamma^4)\hat{\mathbf{P}},$$

whose diagonal entries are

$$a_{ii} = \begin{cases} 1 + (\beta^2+1)^{-2}\left(\gamma^4+\beta^2-2(\beta^2+1)\lambda_i^2\right)\hat{p}_{ii}, & i \leq N \\ 1 + (\beta^2+1)^{-2}\left(\gamma^4+\beta^2\right)\hat{p}_{ii} & i > N \end{cases}$$

Note $\gamma^4 = (N_e-1)^{-1}\sum\lambda^4$.

a_{ii} always $increases$, relative to the prior variance, in unobserved directions.

In observed directions, a_{ii} is smaller than prior variance if λ_i is big enough:

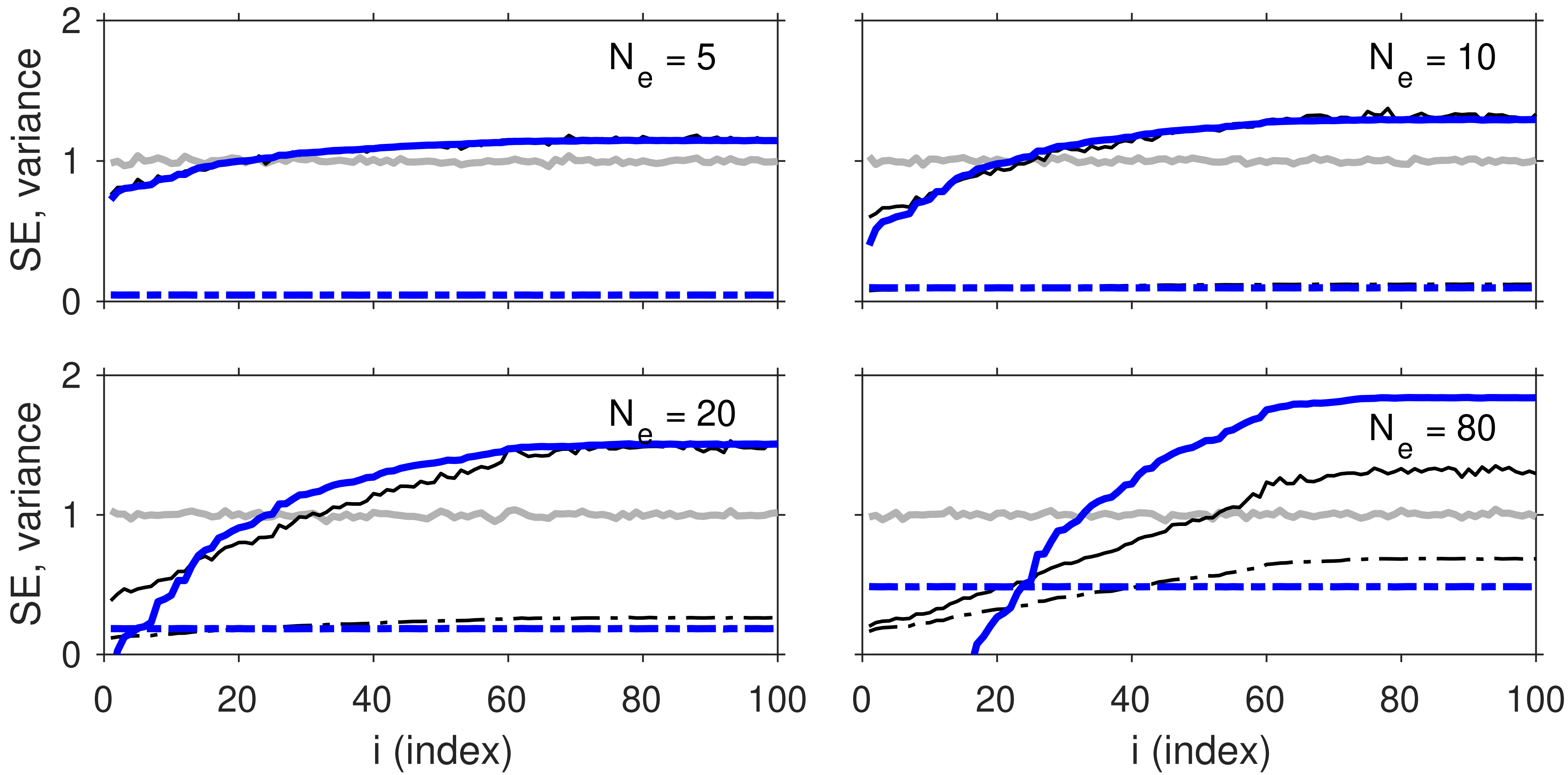
$$\lambda_i > \frac{\gamma^4+\beta^2}{2(\beta^2+1)}$$

Approximation Accuracy

Return to simple spatial example, with length scale giving effective dimension ≈33

Check approximation against actual EnKF results

Approximation Accuracy



black: EnKF results, blue: approximation, gray: prior; solid: squared analysis error, dashed: analysis variance

Summary & Discussion

Summary & Discussion

Sampling error in EnKF

- explicit results for $N_x, N_y, N \gg N_e$
- leverage: "optimal" coordinates + approx. for tall, thin random matrices
- sampling-error effects fully determined by COOs, $\{\lambda_i,\, i = 1, \ldots, N\}$

Summary & Discussion

Sampling error in EnKF

Covariance localization vs inflation

- overconfidence of EnKF often envisioned as problem that slowly accumulates
- in fact, underestimation of $\mathbf{A}$ by $\hat{\mathbf{P}}^a$ can be catastrophic in single update
- localization is essential for practical EnKF

Summary & Discussion

Sampling error in EnKF

Covariance localization vs inflation

Additional interesting directions

- quantify COOs for, say, global NWP
- estimate COOs, then use those estimates to modify algorithm
- role of cross validation (e.g., "double" EnKF)

Role of Gaussianity

$p(\mathbf{x})$, $p(\epsilon)$ are not always Gaussian, and **y** may depend nonlinearly on **x**.

. . . true of all practical applications of the EnKF

In that case,

- interpret KF eqns as best linear unbiased estimator (BLUE)
- only assumptions are existence of $E(\mathbf{x})$, $E(\mathbf{y})$, cov($\mathbf{x}$), cov($\mathbf{y}$), and cov($\mathbf{x}, \mathbf{y}$)
- all results on sampling error here still hold