

Sampling strategies and SQRT analysis schemes for the EnKF

Geir Evensen

Norsk Hydro Research Centre, Bergen, Norway

Based on Evensen 2004, submitted to Ocean Dynamics

Background

- EnKF/EnKS in native formulation are pure Monte Carlo methods.
- Uses Monte Carlo sampling for initial ensemble, model noise and measurement perturbations.
- Uses stochastic equation for ensemble integration.
- Computes analysis based on ensemble perturbations and measurement perturbations.
- Review by Evensen (2003), *Ocean Dynamics*, 53, 343–367.

Part one: Outline

Sampling errors can be reduced by:

- Wise sampling of initial ensemble and model noise as motivated by
 - Pham (2001), MWR.
 - Nerger et al (to appear 2004), MWR.
- Elimination of measurement perturbations in the analysis scheme is possible by use of “square root” algorithms as shown by
 - Andersen (2001), MWR.
 - Bishop et al (2001), MWR.
 - Whitaker and Hamill (2002) MWR.
 - Tippett et al (2003) MWR.

EnKF: Ensemble representation

- Define the ensemble matrix

$$\mathbf{A} = (\psi_1, \psi_2, \dots, \psi_N) \in \mathfrak{R}^{n \times N}.$$

- The ensemble mean is (defining $\mathbf{1}_N \in \mathfrak{R}^{N \times N} \equiv 1/N$)

$$\bar{\mathbf{A}} = \mathbf{A}\mathbf{1}_N.$$

- The ensemble perturbations becomes

$$\mathbf{A}' = \mathbf{A} - \bar{\mathbf{A}} = \mathbf{A}(\mathbf{I} - \mathbf{1}_N).$$

- The ensemble covariance matrix $\mathbf{P}_e \in \mathfrak{R}^{n \times n}$ becomes

$$\mathbf{P}_e = \frac{\mathbf{A}'(\mathbf{A}')^T}{N - 1}$$

EnKF: Measurement perturbations

- Given a vector of measurements $\mathbf{d} \in \mathbb{R}^m$, define

$$\mathbf{d}_j = \mathbf{d} + \boldsymbol{\epsilon}_j, \quad j = 1, \dots, N,$$

stored in

$$\mathbf{D} = (\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_N) \in \mathbb{R}^{m \times N},$$

- The ensemble perturbations are stored in

$$\mathbf{E} = (\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2, \dots, \boldsymbol{\epsilon}_N) \in \mathbb{R}^{m \times N},$$

thus, the measurement error covariance matrix becomes

$$\mathbf{R}_e = \frac{\mathbf{E}\mathbf{E}^T}{N-1}.$$

EnKF: Analysis equation

- The analysis equation can now be written

$$A^a = A + P_e H^T (H P_e H^T + R_e)^{-1} (D - H A).$$

- Defining the innovations $D' = D - H A$ and using previous definitions:

$$\begin{aligned} A^a &= A + A' A'^T H^T (H A' A'^T H^T + E E^T)^{-1} D' \\ &= A + A' S^T C^{-1} D', \\ &= A X \end{aligned}$$

where $S = H A'$ and $C = (S S^T + E E^T)$ and

$$X = I + S^T C^{-1} D'$$

EnKF with linear exact model

- Linear noise free model

$$\mathbf{A}_k = \mathbf{F}^k \mathbf{A}_0$$

- EnKF with linear noise free model

$$\mathbf{A}_k = \mathbf{F}^k \mathbf{A}_0 \prod_{i=1}^k \mathbf{X}_i$$

- With $\text{rank}(\mathbf{F}) = n$ and $\text{rank}(\mathbf{X}_i) = N$, the quality of the EnKF solution is dependent on the rank and conditioning of the initial ensemble \mathbf{A}_0 .

Improved sampling: Introduction

- Full covariance

$$P = Z\Lambda Z^T$$

- Ensemble covariance

$$\begin{aligned} P_e &= \frac{1}{N-1} A' (A')^T \\ &= \frac{1}{N-1} U \Sigma V^T V \Sigma^T U^T = \frac{1}{N-1} U \Sigma \Sigma^T U^T \end{aligned}$$

- When $N \rightarrow \infty$

$$P_e = \frac{1}{N-1} U \Sigma \Sigma^T U^T \rightarrow P = Z \Lambda Z$$

or $U \rightarrow Z$ and $\Sigma \Sigma^T / (N-1) \rightarrow \Lambda$.

Improved sampling: Approach

- Generate an ensemble which best possibly represents P .
- Should be constructed using the first N eigenvectors of P but these are too expensive to compute.
- Approximate eigenvectors Z by singular vectors U computed from a **large** ensemble of perturbations:

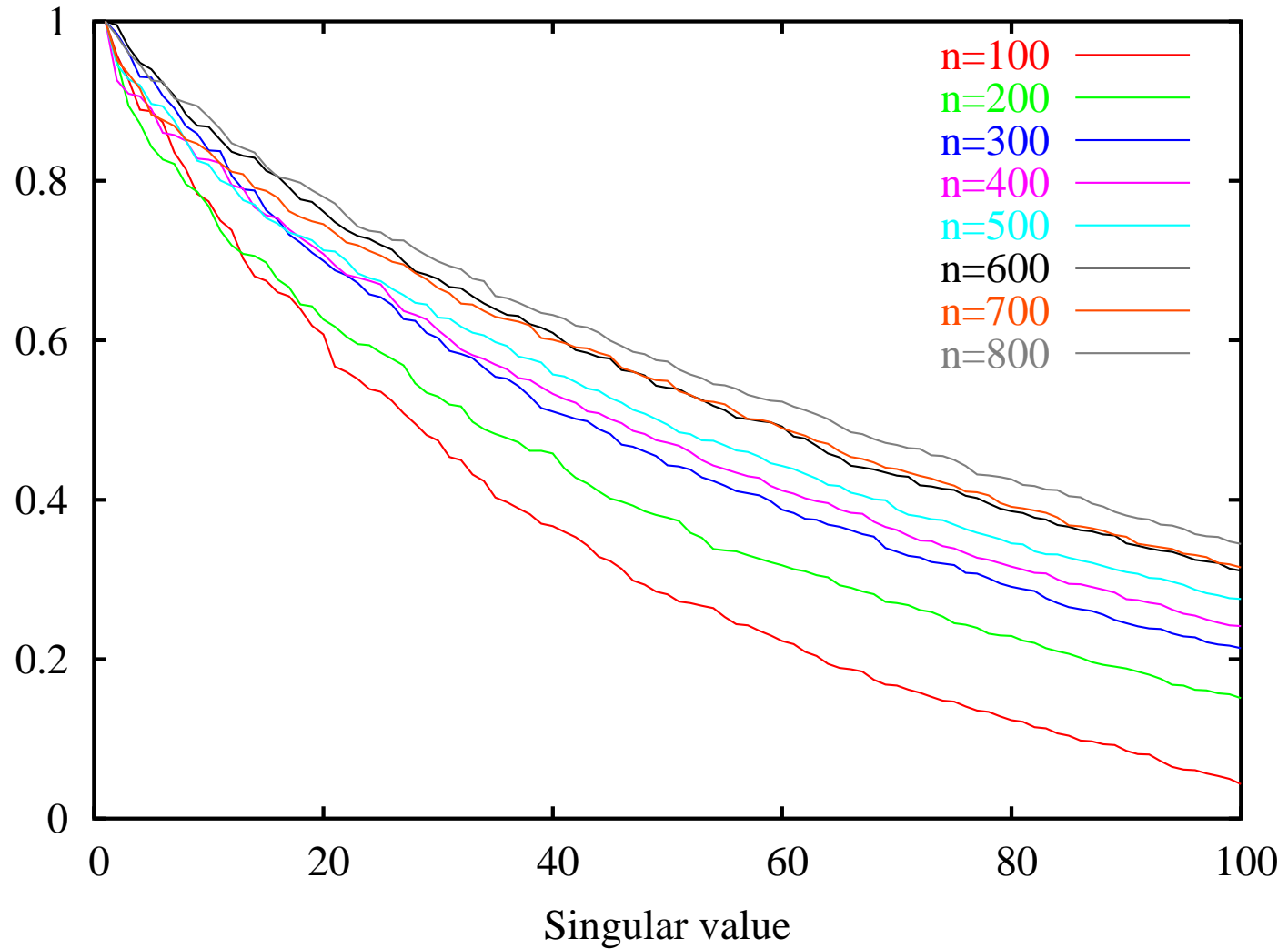
$$\hat{A}' = \hat{U} \hat{\Sigma} \hat{V}^T \in \mathfrak{R}^{n \times \alpha N}.$$

- Store first N singular vectors of \hat{U} in U_1 .
- Store N dominant singular values of $\hat{\Sigma}$ in Σ_1 .
- Generate a **random** orthogonal matrix $V_1^T \in \mathfrak{R}^{N \times N}$.
- Compute

$$A' = U_1 \alpha^{-1} \Sigma_1 V_1^T$$

Improved sampling: Conditioning

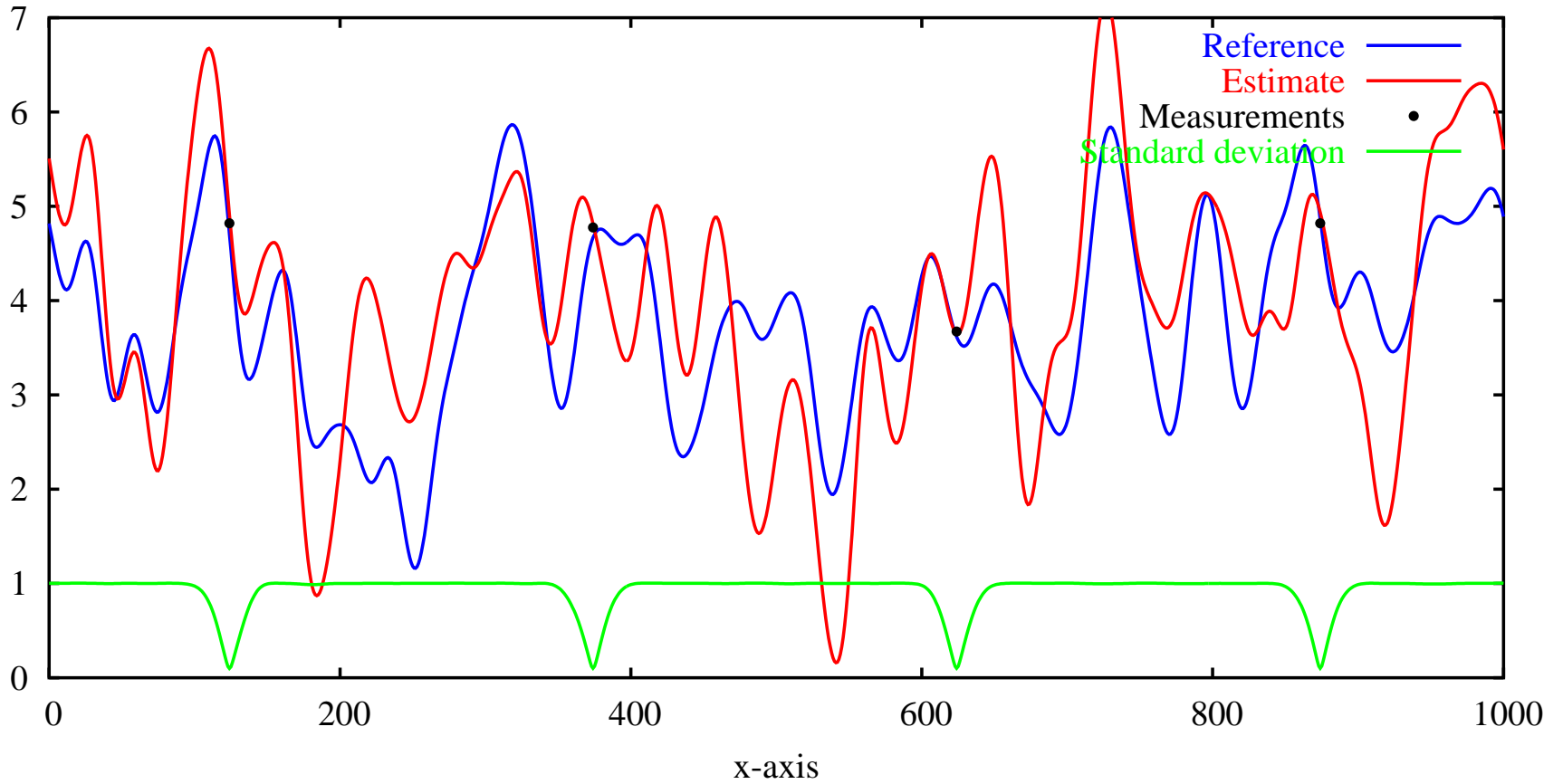
Singular value spectrums



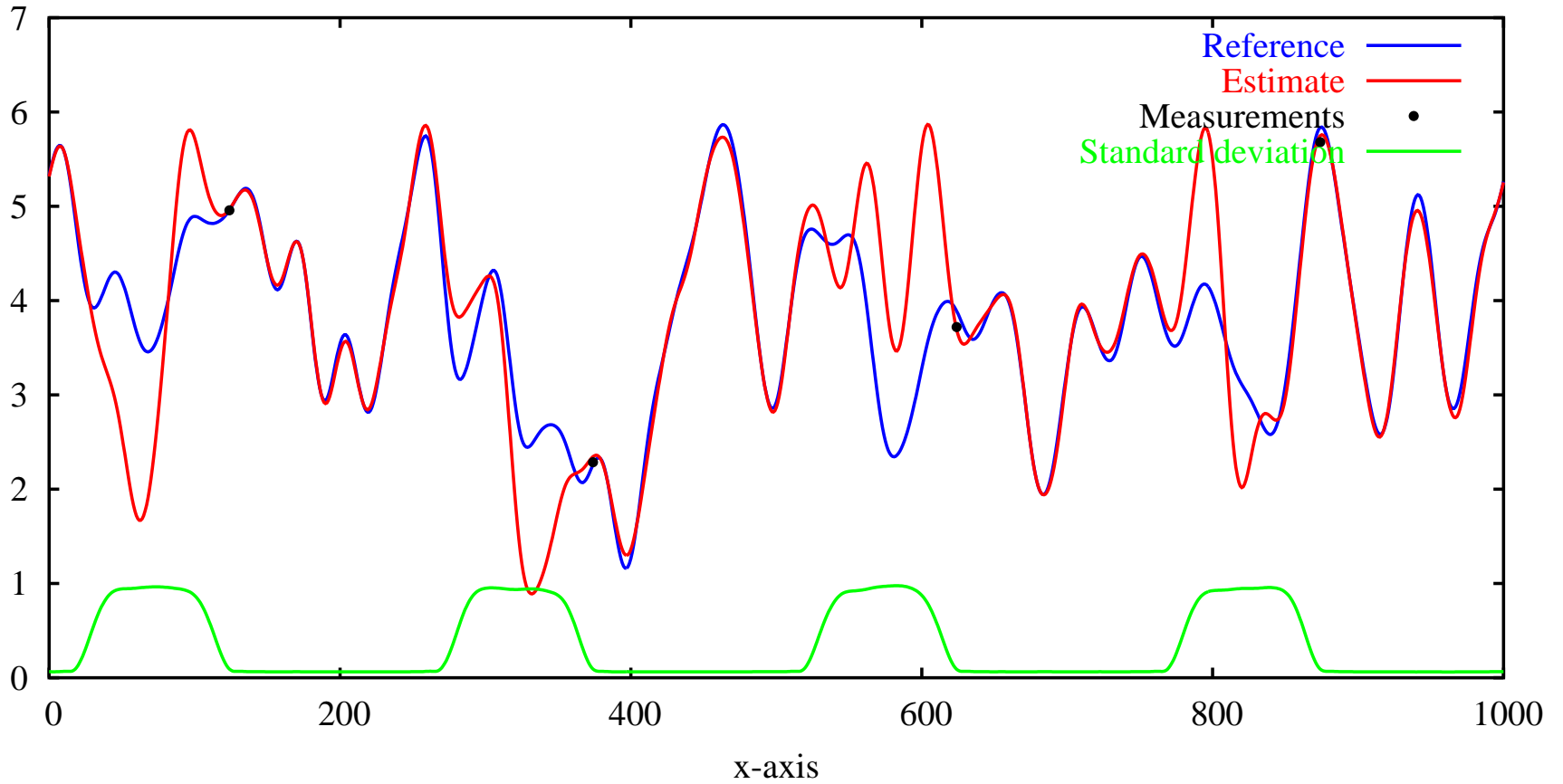
Example: Model

- Linear one-dimensional “exact” advection model on periodic domain.
- Advection speed is 1.0, $\Delta x = 1.0$ and time step is $\Delta t = 1.0$.
- True initial condition sampled from the distribution Φ which has mean equal to zero, variance equal to one and spatial decorrelation length 20.
- First guess is true state plus another sample drawn from Φ , thus initial variance is assumed to be one.
- Initial ensemble is generated by adding samples drawn from Φ , to the first guess.
- Four measurements every 5th time step with std. dev. 0.1.
- Integration length is 300 time units.

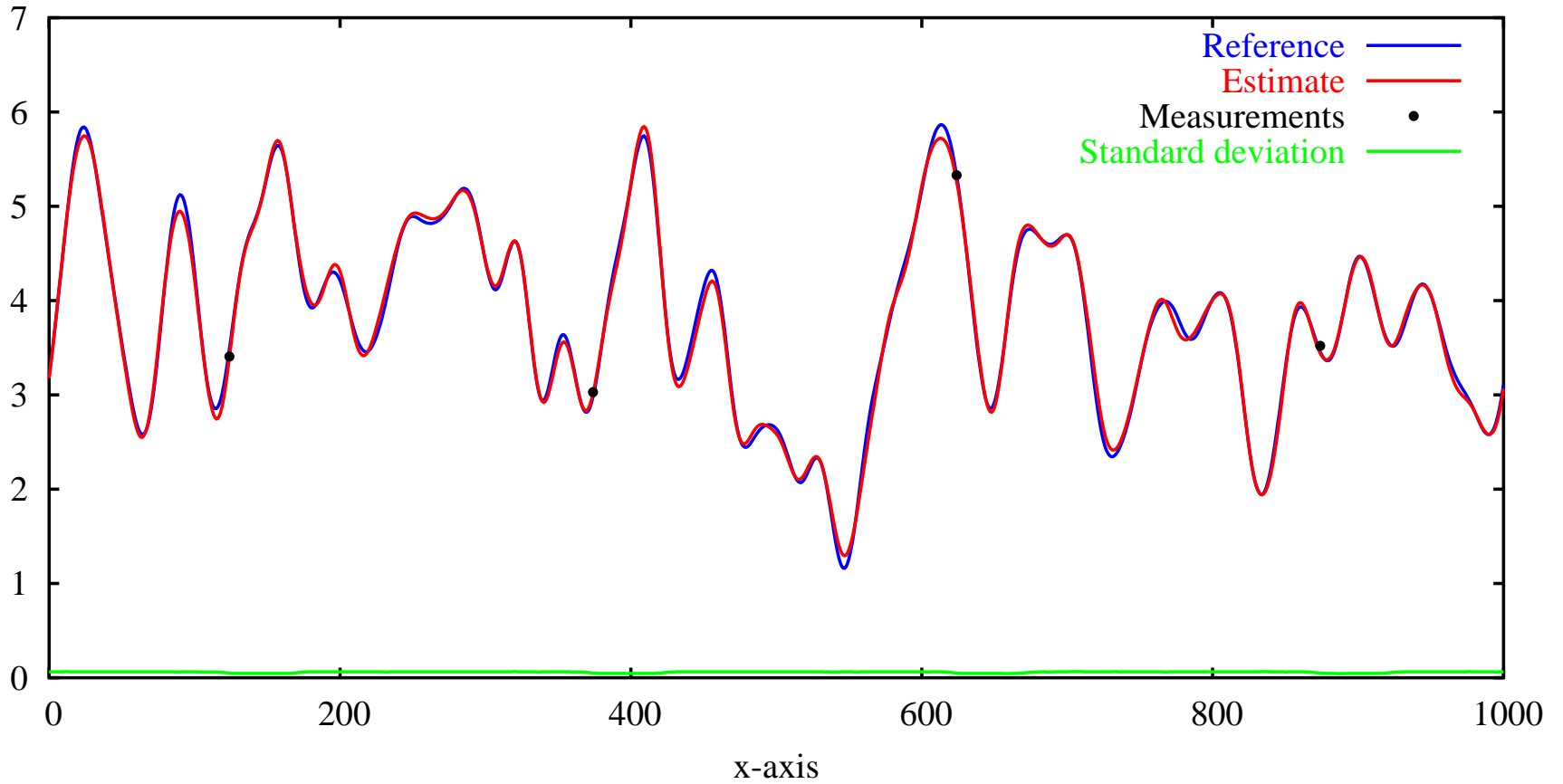
Example (Time t=3)



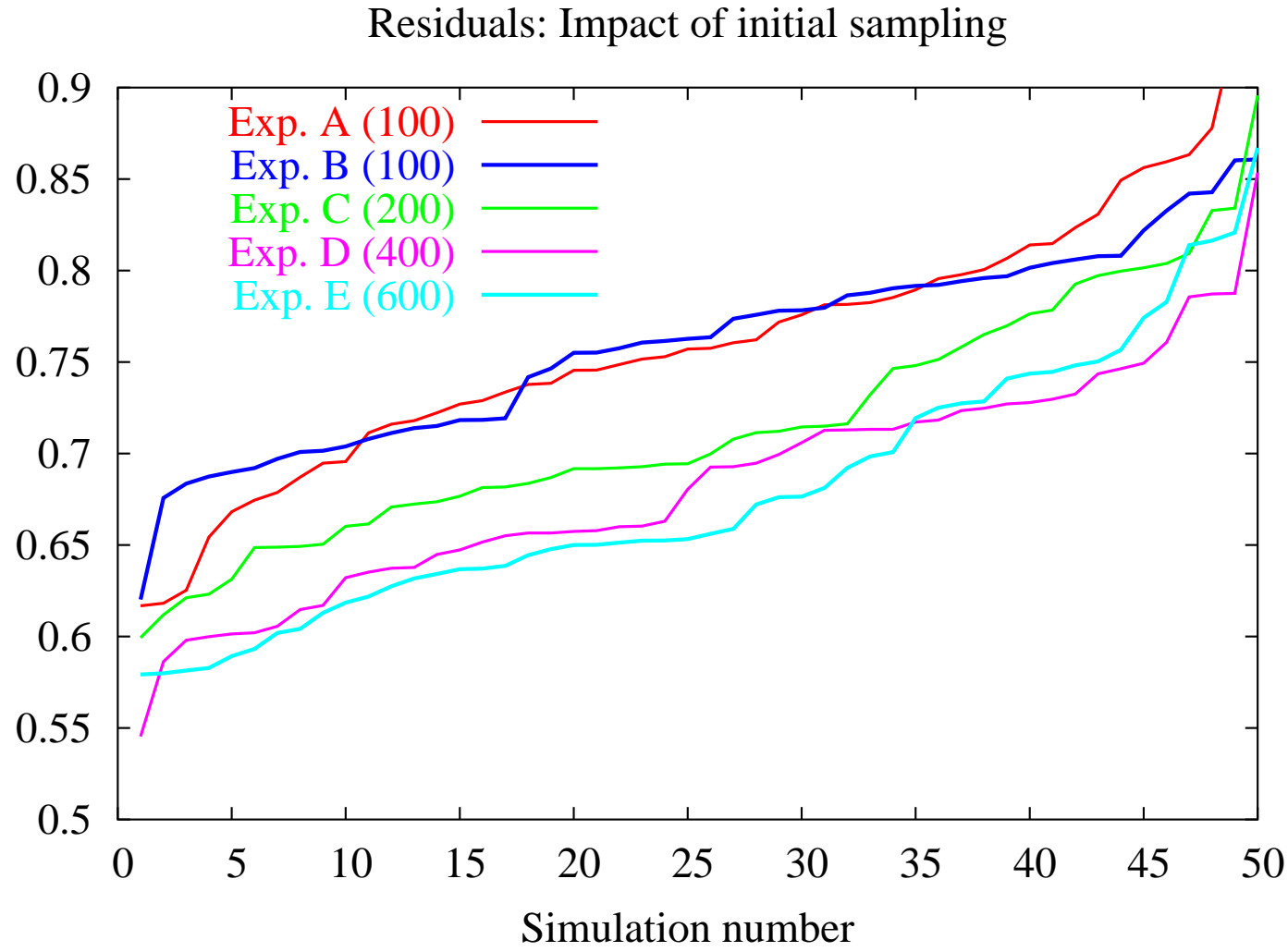
Example (Time t=121)



Example (Time t=241)

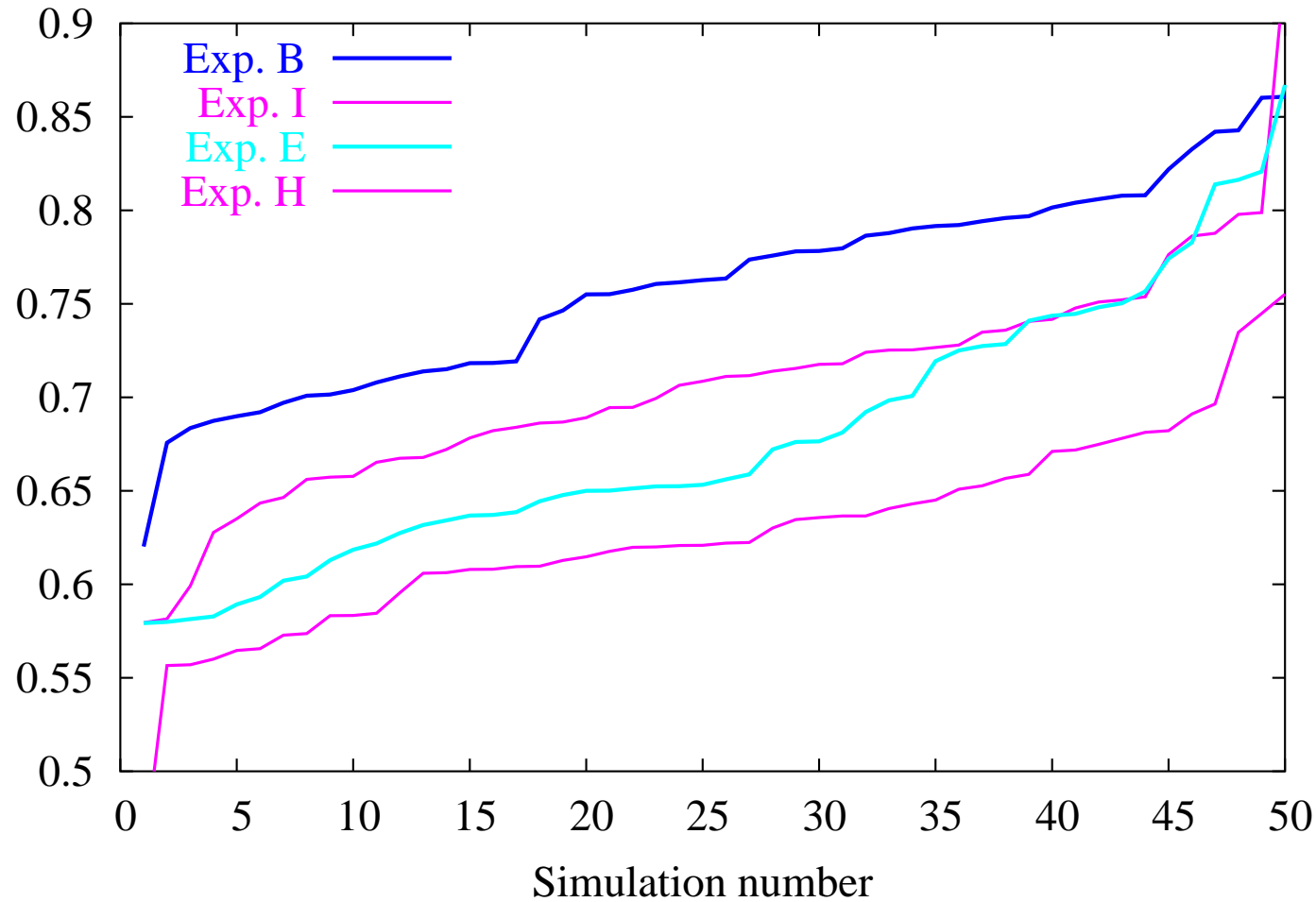


Improved sampling: Initial ensemble



Improved sampling: Measurements

Residuals: Impact of improved sampling of measurement perturbations



A square root analysis scheme (1)

- The ensemble mean can be updated from

$$\bar{\psi}^a = \bar{\psi}^f + A' S^T C^{-1} (d - H \bar{\psi}^f).$$

- The analysis covariance is defined as

$$P^a = P^f - P^f H^T (H P^f H^T + R)^{-1} H P^f,$$

or, in ensemble notation

$$A^{a'} A^{a'T} = A' (I - S^T C^{-1} S) A'^T.$$

- Inverse of C

$$C = Z \Lambda Z^T \Rightarrow C^{-1} = Z \Lambda^{-1} Z^T$$

A square root analysis scheme (2)

• We get

$$\begin{aligned} A^{a'} A^{a'T} &= A' (I - S^T Z \Lambda^{-1} Z^T S) A'^T \\ &= A' \left(I - \left(\sqrt{\Lambda^{-1}} Z^T S \right)^T \left(\sqrt{\Lambda^{-1}} Z^T S \right) \right) A'^T \\ &= A' (I - X_2^T X_2) A'^T, \end{aligned}$$

with the SVD of X_2 defined as

$$U_2 \Sigma_2 V_2^T = X_2 = \sqrt{\Lambda^{-1}} Z^T S.$$

A square root analysis scheme (3)

- We then get

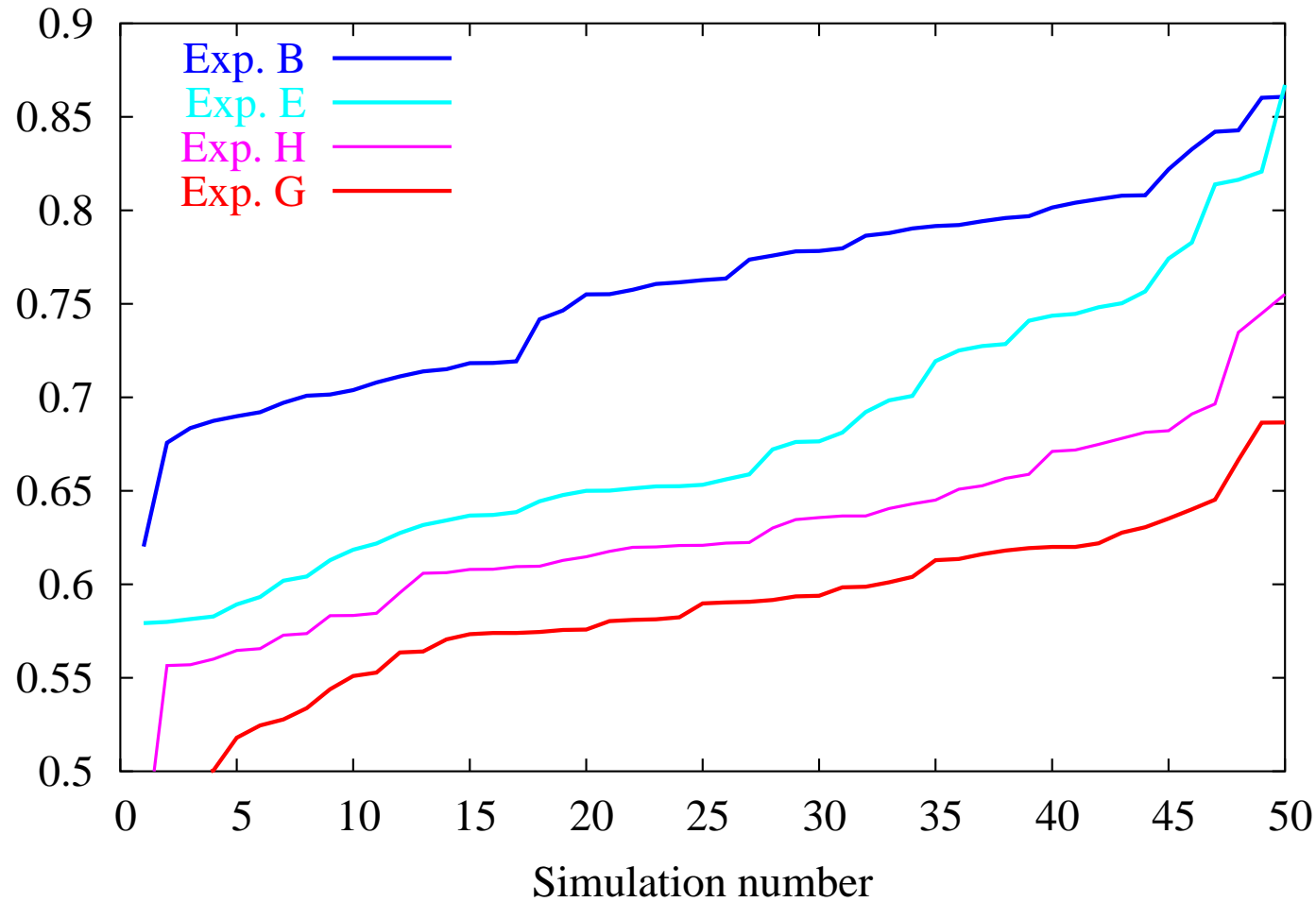
$$\begin{aligned} A^{a'} A^{a'T} &= A' (I - [U_2 \Sigma_2 V_2^T]^T [U_2 \Sigma_2 V_2^T]) A'^T \\ &= A' (I - V_2 \Sigma_2^T \Sigma_2 V_2^T) A'^T \\ &= A' V_2 (I - \Sigma_2^T \Sigma_2) V_2^T A'^T \\ &= \left(A' V_2 \sqrt{I - \Sigma_2^T \Sigma_2} \right) \left(A' V_2 \sqrt{I - \Sigma_2^T \Sigma_2} \right)^T. \end{aligned}$$

- The analysis equation becomes

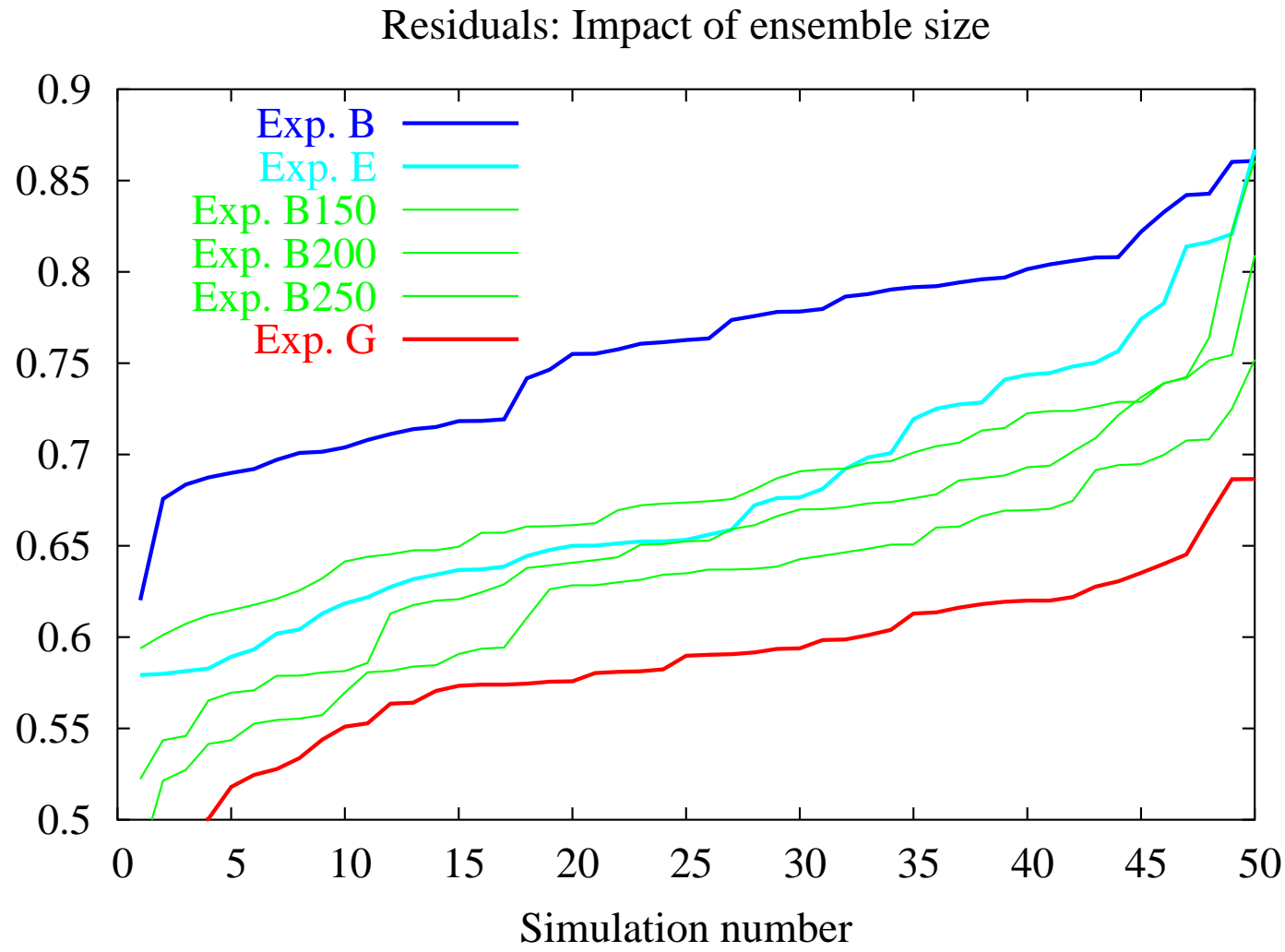
$$A^{a'} = A' V_2 \sqrt{I - \Sigma_2^T \Sigma_2}.$$

A square root analysis scheme (4)

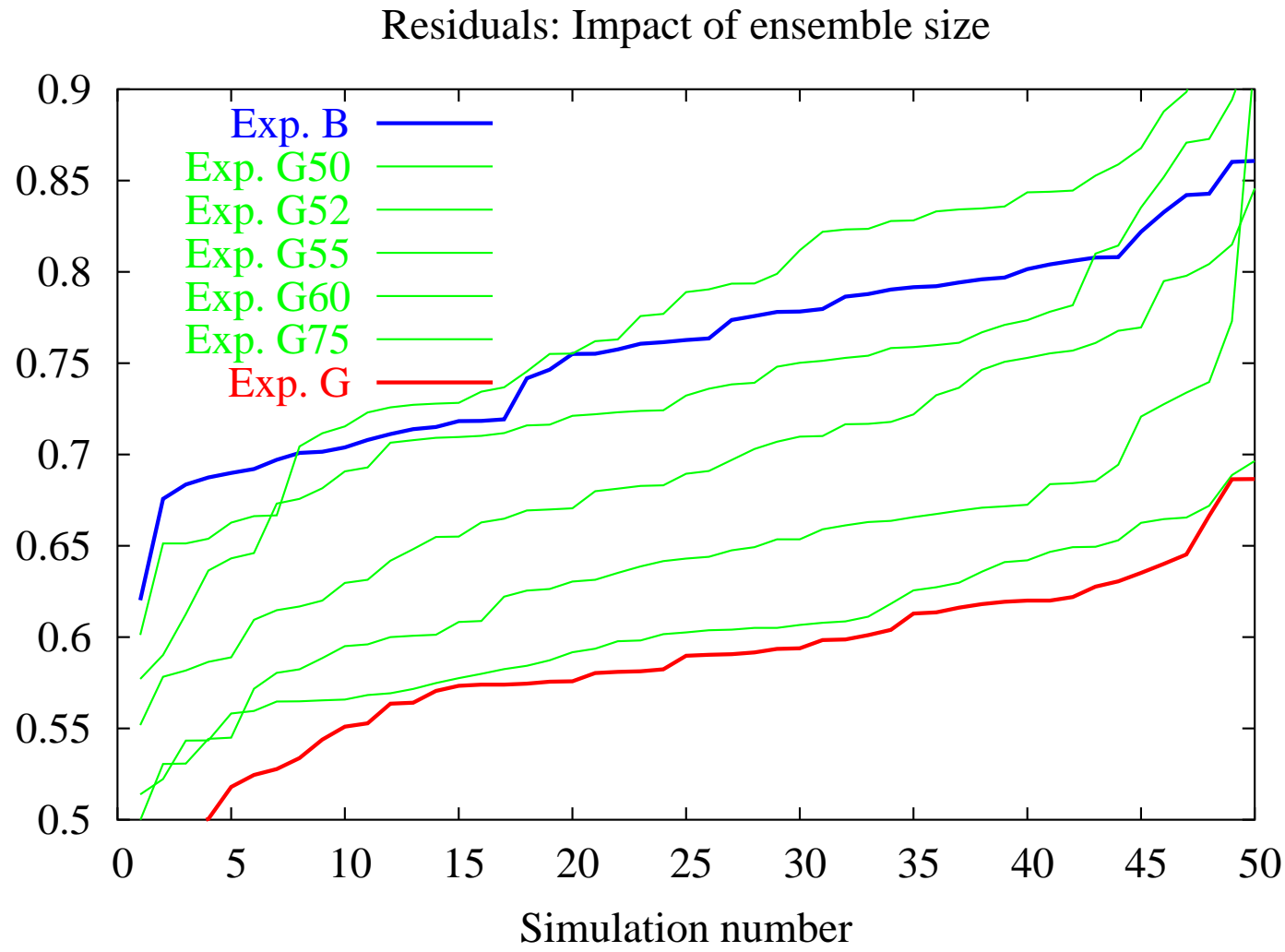
Residuals: Impact of Sqrt analysis



Impact of ensemble size (1)



Impact of ensemble size (2)



Summary: Part one

- Size matters!
- With the right technique size is not all!
 - Sampling of initial ensemble (and model noise).
 - Square root formulation for analysis scheme.
- Details in Evensen 2004, Ocean Dynamics.
- F90 code for new routines available from <http://www.nersc.no/~geir/EnKF>