Towards the estimation of optimal covariance matrices in multi-ensemble data assimilation

David Titley-Peloquin
McGill University, Montreal, Canada

Joint work with S. Gratton and E. Simon
INPT, IRIT, Toulouse, France

AVENUE Project Workshop 2017

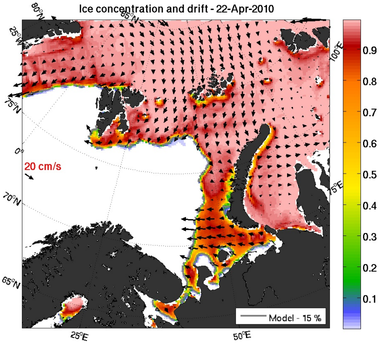
Outline

- Motivation
- A framework for maximum likelihood estimation
- Maximum likelihood estimation with multi-ensembles
- Application to ensemble-based Kalman filters
- Numerical experiments

Outline

- Motivation
- A framework for maximum likelihood estimation
- Maximum likelihood estimation with multi-ensembles
- Application to ensemble-based Kalman filters
- Numerical experiments

Data assimilation combines in an optimal way the heterogeneous and uncertain information provided by model and observations in order to estimate the state of a system.



source: http://topaz.nersc.no

Linear Gaussian framework

$$\begin{array}{lll} \text{Advance in time:} & \mathbf{x}_k = \mathbf{M}_k \mathbf{x}_{k-1} + \mathbf{v}_k, & \mathbf{v}_k \sim \mathcal{N}(0, \mathbf{Q}_k) \\ \text{Observation:} & \mathbf{z}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{w}_k, & \mathbf{w}_k \sim \mathcal{N}(0, \mathbf{R}_k) \end{array}$$

Linear Gaussian framework

$$
\begin{array}{lll}
\text{Advance in time:} & \mathbf{x}_k = \mathbf{M}_k \mathbf{x}_{k-1} + \mathbf{v}_k, & \mathbf{v}_k \sim \mathcal{N}(0, \mathbf{Q}_k) \\
\text{Observation:} & \mathbf{z}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{w}_k, & \mathbf{w}_k \sim \mathcal{N}(0, \mathbf{R}_k)
\end{array}
$$

Kalman filter

Forecast

$$
\begin{aligned}
\mathbf{x}_k^f &= \mathbf{M}_k \mathbf{x}_{k-1}^a \\
\mathbf{P}_k^f &= \mathbf{M}_k \mathbf{P}_{k-1}^a \mathbf{M}_k^T + \mathbf{Q}_k
\end{aligned}
$$

Analysis

$$
\begin{aligned}
\mathbf{K}_k &= \mathbf{P}_k^f \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k^T + \mathbf{R}_k)^{-1} \\
\mathbf{x}_k^a &= \mathbf{x}_k^f + \mathbf{K}_k (\mathbf{z}_k - \mathbf{H}_k \mathbf{x}_k^f) \\
\mathbf{P}_k^a &= (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k^f
\end{aligned}
$$

Linear Gaussian framework

$$\begin{array}{lll} \text{Advance in time:} & \mathbf{x}_k = \mathbf{M}_k \mathbf{x}_{k-1} + \mathbf{v}_k, & \mathbf{v}_k \sim \mathcal{N}(0, \mathbf{Q}_k) \\ \text{Observation:} & \mathbf{z}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{w}_k, & \mathbf{w}_k \sim \mathcal{N}(0, \mathbf{R}_k) \end{array}$$

Kalman filter

Forecast

$$\mathbf{x}_k^f = \mathbf{M}_k \mathbf{x}_{k-1}^a$$
$$\mathbf{P}_k^f = \mathbf{M}_k \mathbf{P}_{k-1}^a \mathbf{M}_k^T + \mathbf{Q}_k$$

Analysis

$$\mathbf{K}_k = \mathbf{P}_k^f \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k^T + \mathbf{R}_k)^{-1}$$
$$\mathbf{x}_k^a = \mathbf{x}_k^f + \mathbf{K}_k (\mathbf{z}_k - \mathbf{H}_k \mathbf{x}_k^f)$$
$$\mathbf{P}_k^a = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k^f$$

Challenges

- storage of covariance matrices
- non-linearity
- non-Gaussian noise

Ensemble Kalman filter

- To avoid updating the error covariance matrix explicitly, the Kalman equations are applied to an ensemble of state vectors (with $i = 1, \dots, n$).

$$\begin{aligned} \text{Forecast:} \quad & \mathbf{x}_k^{f,i} = \mathcal{M}_k(\mathbf{x}_{k-1}^{a,i}) + \mathbf{v}_k, & \mathbf{v}_k \sim \mathcal{N}(0, \mathbf{Q}_k) \\ \text{Analysis:} \quad & \mathbf{x}_k^{a,i} = \mathbf{x}_k^{f,i} + \mathbf{K}_k(\mathbf{z}_k - \mathbf{H}_k \mathbf{x}_k^{f,i} + \mathbf{w}_k), & \mathbf{w}_k \sim \mathcal{N}(0, \mathbf{R}_k) \end{aligned}$$

with $\mathbf{K}_k = \mathbf{P}_k^f \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k^T + \mathbf{R}_k)^{-1}$ the Kalman gain matrix.

Ensemble Kalman filter

- To avoid updating the error covariance matrix explicitly, the Kalman equations are applied to an ensemble of state vectors (with $i = 1, \ldots, n$).

$$\begin{aligned} &\text{Forecast:} \quad \mathbf{x}_k^{f,i} = \mathcal{M}_k(\mathbf{x}_{k-1}^{a,i}) + \mathbf{v}_k, && \mathbf{v}_k \sim \mathcal{N}(0, \mathbf{Q}_k) \\ &\text{Analysis:} \quad \mathbf{x}_k^{a,i} = \mathbf{x}_k^{f,i} + \mathbf{K}_k(\mathbf{z}_k - \mathbf{H}_k \mathbf{x}_k^{f,i} + \mathbf{w}_k), && \mathbf{w}_k \sim \mathcal{N}(0, \mathbf{R}_k) \end{aligned}$$

  with $\mathbf{K}_k = \mathbf{P}_k^f \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k^T + \mathbf{R}_k)^{-1}$ the Kalman gain matrix.

- The mean and error covariance are approximated by the ensemble mean and covariance:

$$\bar{\mathbf{x}}_k^f = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_k^{f,i}, \quad \mathbf{P}_k^f = \frac{1}{n-1} \mathbf{A}^f (\mathbf{A}^f)^T \;\; \text{with} \;\; \mathbf{A}^f = \left[\mathbf{x}_k^{f,1} - \bar{\mathbf{x}}_k^f, \ldots, \mathbf{x}_k^{f,n} - \bar{\mathbf{x}}_k^f \right]$$

Ensemble Kalman filter

- To avoid updating the error covariance matrix explicitly, the Kalman equations are applied to an ensemble of state vectors (with $i = 1, \dots, n$).

$$\begin{aligned} \text{Forecast:} \quad & \mathbf{x}_k^{f,i} = \mathcal{M}_k(\mathbf{x}_{k-1}^{a,i}) + \mathbf{v}_k, & \mathbf{v}_k \sim \mathcal{N}(0, \mathbf{Q}_k) \\ \text{Analysis:} \quad & \mathbf{x}_k^{a,i} = \mathbf{x}_k^{f,i} + \mathbf{K}_k(\mathbf{z}_k - \mathbf{H}_k \mathbf{x}_k^{f,i} + \mathbf{w}_k), & \mathbf{w}_k \sim \mathcal{N}(0, \mathbf{R}_k) \end{aligned}$$

with $\mathbf{K}_k = \mathbf{P}_k^f \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k^T + \mathbf{R}_k)^{-1}$ the Kalman gain matrix.

- The mean and error covariance are approximated by the ensemble mean and covariance:

$$\bar{\mathbf{x}}_k^f = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_k^{f,i}, \quad \mathbf{P}_k^f = \frac{1}{n-1} \mathbf{A}^f (\mathbf{A}^f)^T \ \text{ with } \ \mathbf{A}^f = \left[\mathbf{x}_k^{f,1} - \bar{\mathbf{x}}_k^f, \dots, \mathbf{x}_k^{f,n} - \bar{\mathbf{x}}_k^f \right]$$

- Feasible for large applications with ~ 100 ensemble members with moderation techniques (inflation, localization).

Motivation for multi-ensembles

- Large-scale geophysical applications may involve
  - ▷ several numerical models
  - ▷ different grid sizes
  - ▷ significant i/o costs for storage of model output

$$\begin{aligned}
\text{Forecast:}\quad & \mathbf{x}_k^{f,i} = \mathcal{M}_k(\mathbf{x}_{k-1}^{a,i}) + \mathbf{v}_k, && \mathbf{v}_k \sim \mathcal{N}(0, \mathbf{Q}_k) \\
\text{Analysis:}\quad & \mathbf{x}_k^{a,i} = \mathbf{x}_k^{f,i} + \mathbf{K}_k(\mathbf{z}_k - \mathbf{H}_k\mathbf{x}_k^{f,i} + \mathbf{w}_k), && \mathbf{w}_k \sim \mathcal{N}(0, \mathbf{R}_k)
\end{aligned}$$

Motivation for multi-ensembles

- Large-scale geophysical applications may involve
  - ▷ several numerical models
  - ▷ different grid sizes
  - ▷ significant i/o costs for storage of model output

$$\begin{array}{lll} \text{Forecast:} & \mathbf{x}_k^{f,i} = \mathcal{M}_k(\mathbf{x}_{k-1}^{a,i}) + \mathbf{v}_k, & \mathbf{v}_k \sim \mathcal{N}(0, \mathbf{Q}_k) \\ \text{Analysis:} & \mathbf{x}_k^{a,i} = \mathbf{x}_k^{f,i} + \mathbf{K}_k(\mathbf{z}_k - \mathbf{H}_k \mathbf{x}_k^{f,i} + \mathbf{w}_k), & \mathbf{w}_k \sim \mathcal{N}(0, \mathbf{R}_k) \end{array}$$

- Can we apply ensemble-based Kalman filters with ensembles of different accuracies?
  - ▷ How do we take into account the differences in accuracy between ensembles?
  - ▷ How can we estimate the error covariance matrix given ensembles of different accuracies?

Outline

- Motivation
- A framework for maximum likelihood estimation
- Maximum likelihood estimation with multi-ensembles
- Application to ensemble-based Kalman filters
- Numerical experiments

Setting (I)

- Random variable:

$$\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \text{ with } \boldsymbol{\mu} \in \mathbb{R}^m \text{ and } \boldsymbol{\Sigma} \in \mathbb{R}^{m \times m} \text{ unknown.}$$

- Samples available:

$$\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n] \in \mathbb{R}^{m \times n}$$

- Goal: Compute estimates $\hat{\mu} \approx \boldsymbol{\mu}$ and $\widehat{\boldsymbol{\Sigma}} \approx \boldsymbol{\Sigma}$

Maximum likelihood estimation

- Maximize the likelihood function of the normal distribution:

$$\mathcal{L}(\hat{\mu}, \widehat{\mathbf{\Sigma}}|\mathbf{X}) \propto \prod_{i=1}^{n} \det(\widehat{\mathbf{\Sigma}})^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{x_i} - \hat{\mu})^T \widehat{\mathbf{\Sigma}}^{-1}(\mathbf{x_i} - \hat{\mu})\right)$$

Maximum likelihood estimation

- Maximize the likelihood function of the normal distribution:

$$\mathcal{L}(\hat{\mu}, \widehat{\mathbf{\Sigma}} | \mathbf{X}) \propto \prod_{i=1}^{n} \det(\widehat{\mathbf{\Sigma}})^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{x_i} - \hat{\mu})^T \widehat{\mathbf{\Sigma}}^{-1} (\mathbf{x_i} - \hat{\mu})\right)$$

- Minimize the negative log-likelihood:

$$g(\hat{\mu}, \widehat{\mathbf{\Sigma}}) = \frac{n}{2} \log\det(\widehat{\mathbf{\Sigma}}) + \frac{1}{2}\mathrm{trace}(\widehat{\mathbf{\Sigma}}^{-1} \mathbf{A}\mathbf{A}^T)$$

where **A** is the matrix of anomalies:

$$\mathbf{A} = [\mathbf{x}_1 - \hat{\mu}, \ldots, \mathbf{x}_n - \hat{\mu}]$$

Maximum likelihood estimation

- Minimize the negative log-likelihood:

$$g(\hat{\mu}, \widehat{\mathbf{\Sigma}}) = \frac{n}{2} \log \det(\widehat{\mathbf{\Sigma}}) + \frac{1}{2} \text{trace}(\widehat{\mathbf{\Sigma}}^{-1} \mathbf{A}\mathbf{A}^T)$$

Maximum likelihood estimation

- Minimize the negative log-likelihood:

$$g(\hat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}}) = \frac{n}{2} \log \det(\widehat{\boldsymbol{\Sigma}}) + \frac{1}{2} \mathrm{trace}(\widehat{\boldsymbol{\Sigma}}^{-1} \mathbf{A}\mathbf{A}^T)$$

- Differentiate with respect to $\hat{\boldsymbol{\mu}}$:

$$\sum_{i=1}^{n} (\mathbf{x}_i - \hat{\boldsymbol{\mu}})^T \widehat{\boldsymbol{\Sigma}}^{-1} \{\mathrm{d}\hat{\boldsymbol{\mu}}\} = 0 \quad \Rightarrow \quad \hat{\boldsymbol{\mu}} = \bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i$$

Maximum likelihood estimation

- Minimize the negative log-likelihood:

$$g(\hat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}}) = \frac{n}{2} \log \det(\widehat{\boldsymbol{\Sigma}}) + \frac{1}{2} \mathrm{trace}(\widehat{\boldsymbol{\Sigma}}^{-1} \mathbf{A}\mathbf{A}^T)$$

- Differentiate with respect to $\hat{\boldsymbol{\mu}}$:

$$\sum_{i=1}^{n} (\mathbf{x}_i - \hat{\boldsymbol{\mu}})^T \widehat{\boldsymbol{\Sigma}}^{-1} \{\mathrm{d}\hat{\boldsymbol{\mu}}\} = 0 \quad \Rightarrow \quad \hat{\boldsymbol{\mu}} = \bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i$$

- Differentiate with respect to $\widehat{\boldsymbol{\Sigma}}$:

$$\mathrm{trace}\left(\widehat{\boldsymbol{\Sigma}}^{-1} \{\mathrm{d}\widehat{\boldsymbol{\Sigma}}\} \left(n\mathbf{I} - \widehat{\boldsymbol{\Sigma}}^{-1} \mathbf{A}\mathbf{A}^T\right)\right) = 0 \quad \Rightarrow \quad \widehat{\boldsymbol{\Sigma}} = \frac{1}{n} \mathbf{A}\mathbf{A}^T$$

Maximum likelihood estimation

- Minimize the negative log-likelihood:

$$g(\hat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}}) = \frac{n}{2} \log \det(\widehat{\boldsymbol{\Sigma}}) + \frac{1}{2} \text{trace}(\widehat{\boldsymbol{\Sigma}}^{-1} \mathbf{A}\mathbf{A}^T)$$

- Differentiate with respect to $\hat{\boldsymbol{\mu}}$:

$$\sum_{i=1}^{n} (\mathbf{x}_i - \hat{\boldsymbol{\mu}})^T \widehat{\boldsymbol{\Sigma}}^{-1} \{d\hat{\boldsymbol{\mu}}\} = 0 \quad \Rightarrow \quad \hat{\boldsymbol{\mu}} = \bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i$$

- Differentiate with respect to $\widehat{\boldsymbol{\Sigma}}$:

$$\text{trace}\left(\widehat{\boldsymbol{\Sigma}}^{-1} \{d\widehat{\boldsymbol{\Sigma}}\} \left(n\mathbf{I} - \widehat{\boldsymbol{\Sigma}}^{-1} \mathbf{A}\mathbf{A}^T\right)\right) = 0 \quad \Rightarrow \quad \widehat{\boldsymbol{\Sigma}} = \frac{1}{n} \mathbf{A}\mathbf{A}^T$$

provided $n > m$

Setting (II)

- Random variable:

$$\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \text{ with } \boldsymbol{\mu} \in \mathbb{R}^m \text{ and } \boldsymbol{\Sigma} \in \mathbb{R}^{m \times m} \text{ unknown.}$$

- Very few samples available:

$$\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n] \in \mathbb{R}^{m \times n} \quad \text{with } n < m$$

- Goal: Compute estimates $\hat{\boldsymbol{\mu}} \approx \boldsymbol{\mu}$ and $\widehat{\boldsymbol{\Sigma}} \approx \boldsymbol{\Sigma}$

Maximum likelihood estimation when $n < m$

- Pick a subspace of dimension $p \ll m$:

$$\mathbf{Q} \in \mathbb{R}^{m \times p}, \quad \mathbf{Q}^T\mathbf{Q} = \mathbf{I}$$

- Search for an estimate of the covariance matrix of the form

$$\widehat{\boldsymbol{\Sigma}} = \mathbf{QMQ}^T = \mathbf{QLL}^T\mathbf{Q}^T$$

where $\mathbf{M} \in \mathbb{R}^{p \times p}$ is symmetric positive definite and $\mathbf{M} = \mathbf{LL}^T$ is either a Cholesky or symmetric square root factorization.

Maximum likelihood estimation when $n < m$

- Pick a subspace of dimension $p \ll m$:

$$\mathbf{Q} \in \mathbb{R}^{m \times p}, \quad \mathbf{Q}^T\mathbf{Q} = \mathbf{I}$$

- Search for an estimate of the covariance matrix of the form

$$\widehat{\mathbf{\Sigma}} = \mathbf{QMQ}^T = \mathbf{QLL}^T\mathbf{Q}^T$$

  where $\mathbf{M} \in \mathbb{R}^{p \times p}$ is symmetric positive definite and $\mathbf{M} = \mathbf{LL}^T$ is either a Cholesky or symmetric square root factorization.

- Similar to regularization/shrinkage techniques (Ledoit and Wolf 2004, Ueno et. al. 2009, 2010, 2014, Johns and Mandel 2010, etc.)

Maximum likelihood estimation when $n < m$

- Maximize the likelihood function with respect to the degenerate normal distribution:

$$\mathcal{L}(\widehat{\boldsymbol{\Sigma}}|\mathbf{X}) \propto \prod_{i=1}^{n} \det{}^*(\widehat{\boldsymbol{\Sigma}})^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{x_i} - \bar{\mathbf{x}})^T \widehat{\boldsymbol{\Sigma}}^{\dagger} (\mathbf{x_i} - \bar{\mathbf{x}})\right)$$

Maximum likelihood estimation when $n < m$

- Maximize the likelihood function with respect to the degenerate normal distribution:

$$\mathcal{L}(\widehat{\mathbf{\Sigma}}|\mathbf{X}) \propto \prod_{i=1}^{n} \det{}^*(\widehat{\mathbf{\Sigma}})^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{x_i} - \bar{\mathbf{x}})^T \widehat{\mathbf{\Sigma}}^{\dagger} (\mathbf{x_i} - \bar{\mathbf{x}})\right)$$

where

▷ $\det^*(\widehat{\mathbf{\Sigma}})$ is the pseudo-determinant

$$\det{}^*(\widehat{\mathbf{\Sigma}}) = \det{}^*(\mathbf{QLL}^T\mathbf{Q}^T) = \mathbf{Q}\det(\mathbf{LL}^T)\mathbf{Q}^T$$

▷ $\widehat{\mathbf{\Sigma}}^{\dagger}$ is the More-Penrose generalized inverse

$$\widehat{\mathbf{\Sigma}}^{\dagger} = (\mathbf{QLL}^T\mathbf{Q}^T)^{\dagger} = \mathbf{Q}(\mathbf{LL}^T)^{-1}\mathbf{Q}^T$$

Maximum likelihood estimation when $n < m$

- Minimize the negative log-likelihood:

$$g(\mathbf{L}) = n \log \det(\mathbf{L}) + \frac{1}{2}\mathrm{trace}\big((\mathbf{L}\mathbf{L}^T)^{-1}\mathbf{Q}^T\mathbf{A}\mathbf{A}^T\mathbf{Q}\big)$$

where $\mathbf{A}$ is the matrix of anomalies.

Maximum likelihood estimation when $n < m$

- Minimize the negative log-likelihood:

$$g(\mathbf{L}) = n \log \det(\mathbf{L}) + \frac{1}{2}\text{trace}\big((\mathbf{L}\mathbf{L}^T)^{-1}\mathbf{Q}^T\mathbf{A}\mathbf{A}^T\mathbf{Q}\big)$$

where $\mathbf{A}$ is the matrix of anomalies.

- First order optimality conditions:

$$\mathbf{L}\mathbf{L}^T = \frac{1}{n}\mathbf{Q}^T\mathbf{A}\mathbf{A}^T\mathbf{Q}$$

$$\widehat{\mathbf{\Sigma}} = \mathbf{Q}\mathbf{L}\mathbf{L}^T\mathbf{Q}^T = \frac{1}{n}\mathbf{Q}\mathbf{Q}^T\mathbf{A}\mathbf{A}^T\mathbf{Q}\mathbf{Q}^T$$

Maximum likelihood estimation when $n < m$

- Minimize the negative log-likelihood:

$$g(\mathbf{L}) = n \log \det(\mathbf{L}) + \frac{1}{2}\text{trace}\big((\mathbf{L}\mathbf{L}^T)^{-1}\mathbf{Q}^T\mathbf{A}\mathbf{A}^T\mathbf{Q}\big)$$

where $\mathbf{A}$ is the matrix of anomalies.

- First order optimality conditions:

$$\mathbf{L}\mathbf{L}^T = \frac{1}{n}\mathbf{Q}^T\mathbf{A}\mathbf{A}^T\mathbf{Q}$$
$$\widehat{\mathbf{\Sigma}} = \mathbf{Q}\mathbf{L}\mathbf{L}^T\mathbf{Q}^T = \frac{1}{n}\mathbf{Q}\mathbf{Q}^T\mathbf{A}\mathbf{A}^T\mathbf{Q}\mathbf{Q}^T$$

Remarks:

- ▷ For $\mathbf{L}$ to be nonsingular: $p \le n - 1$.
- ▷ $\widehat{\mathbf{\Sigma}}$ is a projection of the sample covariance matrix.
- ▷ If $\text{range}(\mathbf{Q}) = \text{range}(\mathbf{A})$ then $\widehat{\mathbf{\Sigma}}$ is the sample covariance matrix.

Numerical optimization

- To compute the entries of the optimal **L** numerically, we use the one-to-one correspondence between

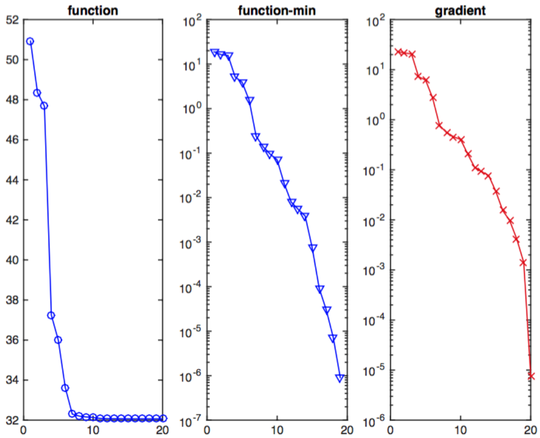
$$\mathbf{L} = \begin{bmatrix} \ell_{1,1} & & & \\ \ell_{2,1} & \ell_{2,2} & & \\ \vdots & & \ddots & \\ \ell_{p,1} & \dots & \dots & \ell_{p,p} \end{bmatrix} \qquad \longleftrightarrow \qquad \boldsymbol{\ell} = \begin{bmatrix} \ell_{1,1} \\ \vdots \\ \ell_{p,1} \\ \ell_{2,2} \\ \vdots \\ \ell_{p,p} \end{bmatrix}$$

Numerical optimization

- To compute the entries of the optimal $\mathbf{L}$ numerically, we use the one-to-one correspondence between

$$\mathbf{L} = \begin{bmatrix} \ell_{1,1} & & & \\ \ell_{2,1} & \ell_{2,2} & & \\ \vdots & & \ddots & \\ \ell_{p,1} & \dots & \dots & \ell_{p,p} \end{bmatrix} \qquad \longleftrightarrow \qquad \boldsymbol{\ell} = \begin{bmatrix} \ell_{1,1} \\ \vdots \\ \ell_{p,1} \\ \ell_{2,2} \\ \vdots \\ \ell_{p,p} \end{bmatrix}$$

- We use BFGS to minimize the negative log-likelihood $g(\boldsymbol{\ell})$.
  - ▷ Find a search direction $\mathbf{p}_k$ by solving $\mathbf{B}_k \mathbf{p}_k = -\nabla g(\boldsymbol{\ell}_k)$.
  - ▷ Perform a linesearch to find an acceptable step-size α_k.
  - ▷ Take a step: $\boldsymbol{\ell}_{k+1} = \boldsymbol{\ell}_k + \alpha_k \mathbf{p}_k$.
  - ▷ Use a low-rank update to obtain $\mathbf{B}_{k+1}$ from $\mathbf{B}_k$.



Example from a L40 simulation

$m = 40,\ n = 5,\ p = 4$

Q chosen to be the left singular vectors of **A**

Outline

- Motivation
- A framework for maximum likelihood estimation
- Maximum likelihood estimation with multi-ensembles
- Application to ensemble-based Kalman filters
- Numerical experiments

Setting (III)

- Random variables:

$$\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \text{ with } \boldsymbol{\mu} \in \mathbb{R}^m \text{ and } \boldsymbol{\Sigma} \in \mathbb{R}^{m \times m} \text{ unknown.}$$
$$\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma} + \mathbf{C}) \text{ with } \mathbf{C} \in \mathbb{R}^{m \times m} \text{ known.}$$

- Very few samples available:

$$\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_{n_x}] \in \mathbb{R}^{m \times n_x}$$
$$\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_{n_y}] \in \mathbb{R}^{m \times n_y}$$

with $n_x + n_y < m$.

- Goal: Compute estimates $\hat{\boldsymbol{\mu}} \approx \boldsymbol{\mu}$ and $\widehat{\boldsymbol{\Sigma}} \approx \boldsymbol{\Sigma}$

Maximum likelihood estimation with multi-ensembles

- Given samples "high-fidelity" samples $\mathbf{X}$ and "noisy" samples $\mathbf{Y}$, the likelihood function with respect to the degenerate normal distribution is

$$\begin{aligned}&\mathcal{L}(\widehat{\boldsymbol{\Sigma}}|\mathbf{X},\mathbf{Y})\\ &\propto \prod_{i=1}^{n_x} \det{}^*(\widehat{\boldsymbol{\Sigma}})^{-1/2} \exp\left(-\tfrac{1}{2}(\mathbf{x_i}-\bar{\mathbf{x}})^T \widehat{\boldsymbol{\Sigma}}^{\dagger}(\mathbf{x_i}-\bar{\mathbf{x}})\right)\\ &\qquad \times \prod_{j=1}^{n_y} \det{}^*(\widehat{\boldsymbol{\Sigma}}+\mathbf{C})^{-1/2} \exp\left(-\tfrac{1}{2}(\mathbf{y_j}-\bar{\mathbf{x}})^T (\widehat{\boldsymbol{\Sigma}}+\mathbf{C})^{\dagger}(\mathbf{y_j}-\bar{\mathbf{x}})\right)\end{aligned}$$

where $\det^*(\cdot)$ is the pseudo-determinant and $(\cdot)^\dagger$ denotes the More-Penrose generalized inverse.

Maximum likelihood estimation with multi-ensembles

- Search for an estimate of the covariance matrix of the form

$$\widehat{\mathbf{\Sigma}} = \mathbf{QMQ}^T = \mathbf{QLL}^T\mathbf{Q}^T, \quad \text{where} \quad \mathbf{Q} \in \mathbb{R}^{m \times p}, \; \mathbf{Q}^T\mathbf{Q} = \mathbf{I}.$$

- As in the single-ensemble case:

$$\det{}^*(\widehat{\mathbf{\Sigma}}) = \mathbf{Q}\det(\mathbf{LL}^T)\mathbf{Q}^T, \quad \widehat{\mathbf{\Sigma}}^\dagger = \mathbf{Q}(\mathbf{LL}^T)^{-1}\mathbf{Q}^T$$

Maximum likelihood estimation with multi-ensembles

- Search for an estimate of the covariance matrix of the form

$$\widehat{\mathbf{\Sigma}} = \mathbf{QMQ}^T = \mathbf{QLL}^T\mathbf{Q}^T, \quad \text{where} \quad \mathbf{Q} \in \mathbb{R}^{m\times p}, \; \mathbf{Q}^T\mathbf{Q} = \mathbf{I}.$$

- As in the single-ensemble case:

$$\det{}^*(\widehat{\mathbf{\Sigma}}) = \mathbf{Q}\det(\mathbf{LL}^T)\mathbf{Q}^T, \quad \widehat{\mathbf{\Sigma}}^\dagger = \mathbf{Q}(\mathbf{LL}^T)^{-1}\mathbf{Q}^T$$

- We approximate $\widehat{\mathbf{\Sigma}} + \mathbf{C}$ by projection:

$$\widehat{\mathbf{\Sigma}} + \mathbf{C} = \mathbf{QLL}^T\mathbf{Q}^T + \mathbf{C} \approx \mathbf{Q}(\mathbf{LL}^T + \mathbf{Q}^T\mathbf{CQ})\mathbf{Q}^T$$

Maximum likelihood estimation with multi-ensembles

- Search for an estimate of the covariance matrix of the form

$$\widehat{\mathbf{\Sigma}} = \mathbf{QMQ}^T = \mathbf{QLL}^T\mathbf{Q}^T, \quad \text{where} \;\; \mathbf{Q} \in \mathbb{R}^{m \times p}, \;\; \mathbf{Q}^T\mathbf{Q} = \mathbf{I}.$$

- As in the single-ensemble case:

$$\det{}^*(\widehat{\mathbf{\Sigma}}) = \mathbf{Q}\det(\mathbf{LL}^T)\mathbf{Q}^T, \quad \widehat{\mathbf{\Sigma}}^\dagger = \mathbf{Q}(\mathbf{LL}^T)^{-1}\mathbf{Q}^T$$

- We approximate $\widehat{\mathbf{\Sigma}} + \mathbf{C}$ by projection:

$$\widehat{\mathbf{\Sigma}} + \mathbf{C} = \mathbf{QLL}^T\mathbf{Q}^T + \mathbf{C} \approx \mathbf{Q}(\mathbf{LL}^T + \mathbf{Q}^T\mathbf{CQ})\mathbf{Q}^T$$

- Then

$$\det{}^*(\widehat{\mathbf{\Sigma}} + \mathbf{C}) \approx \mathbf{Q}\det(\mathbf{LL}^T + \mathbf{Q}^T\mathbf{CQ})\mathbf{Q}^T$$

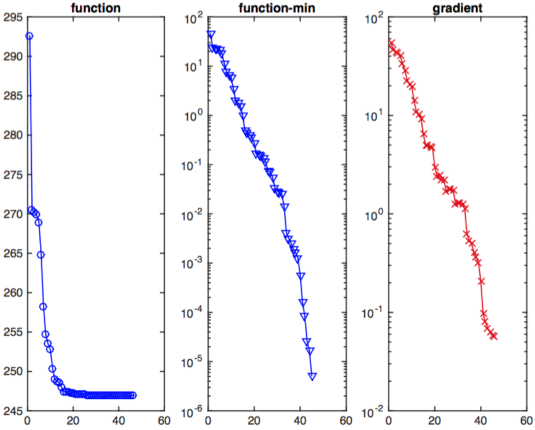
$$(\widehat{\mathbf{\Sigma}} + \mathbf{C})^\dagger \approx \mathbf{Q}(\mathbf{LL}^T + \mathbf{Q}^T\mathbf{CQ})^{-1}\mathbf{Q}^T$$

Maximum likelihood estimation with multi-ensembles

- Minimize the negative log-likelihood:

$$g(\mathbf{L}) = n_x \log\det(\mathbf{L}) + \frac{n_y}{2} \log\det(\mathbf{L}\mathbf{L}^T + \mathbf{Q}^T\mathbf{C}\mathbf{Q})$$
$$+ \operatorname{trace}\big((\mathbf{L}\mathbf{L}^T)^{-1}\mathbf{Q}^T\mathbf{A}_x\mathbf{A}_x^T\mathbf{Q}\big)$$
$$+ \operatorname{trace}\big((\mathbf{L}\mathbf{L}^T + \mathbf{Q}^T\mathbf{C}\mathbf{Q})^{-1}\mathbf{Q}^T\mathbf{A}_y\mathbf{A}_y^T\mathbf{Q}\big)$$

with $\mathbf{A}_x$ and $\mathbf{A}_y$ the anomalies of samples $\mathbf{X}$ and $\mathbf{Y}$, respectively.



Example from a L40 simulation

$m = 40,\ n_x = 5,\ n_y = 15,\ p = 19$

noise: truncation of the vector components to 10^{-6}, $\mathbf{C} = 10^{-12}\mathbf{I}$

$\mathbf{Q}$ chosen to be the left singular vectors of matrix of anomalies

convergence of $\hat{\boldsymbol{\Sigma}} = \mathbf{Q}\mathbf{L}\mathbf{L}^T\mathbf{Q}^T$ close to the sample covariance $\sim$ variance of the noise

Outline

- Motivation
- A framework for maximum likelihood estimation
- Maximum likelihood estimation with multi-ensembles
- Application to ensemble-based Kalman filters
- Numerical experiments

Anomalies-based formulation (Sakov et. al., 2010)

- Notation:
  - ▷ ensemble $\mathbf{E} = [\mathbf{X}, \mathbf{Y}] \in \mathbb{R}^{m \times n}$
  - ▷ ensemble mean $\mathbf{x} = \frac{1}{n}\mathbf{E}\mathbf{e} \in \mathbb{R}^m$, where $\mathbf{e} = [1, \ldots, 1]^T$
  - ▷ anomalies $\mathbf{A} = \mathbf{E} - \mathbf{x}\mathbf{e}^T$

Anomalies-based formulation (Sakov et. al., 2010)

- Notation:
  - ▷ ensemble $\mathbf{E} = [\mathbf{X}, \mathbf{Y}] \in \mathbb{R}^{m \times n}$
  - ▷ ensemble mean $\mathbf{x} = \frac{1}{n}\mathbf{E}\mathbf{e} \in \mathbb{R}^m$, where $\mathbf{e} = [1, \ldots, 1]^T$
  - ▷ anomalies $\mathbf{A} = \mathbf{E} - \mathbf{x}\mathbf{e}^T$

- The analysis step updates the mean and the anomalies instead of the ensemble members.

Anomalies-based formulation (Sakov et. al., 2010)

- Notation:
  - ▷ ensemble $\mathbf{E} = [\mathbf{X}, \mathbf{Y}] \in \mathbb{R}^{m \times n}$
  - ▷ ensemble mean $\mathbf{x} = \frac{1}{n}\mathbf{E}\mathbf{e} \in \mathbb{R}^{m}$, where $\mathbf{e} = [1, \ldots, 1]^T$
  - ▷ anomalies $\mathbf{A} = \mathbf{E} - \mathbf{x}\mathbf{e}^T$
- The analysis step updates the mean and the anomalies instead of the ensemble members.
- To recover the analysis ensemble: $\mathbf{E} = \mathbf{x}\mathbf{e}^T + \mathbf{A}$

Anomalies-based formulation (Sakov et. al., 2010)

- Update of the **ensemble mean**:

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{A}^f \mathbf{G} \mathbf{R}^{-\frac{1}{2}} (\mathbf{z} - \mathbf{H}\mathbf{x}^f)/\sqrt{n-1}$$

with

$$\mathbf{G} = \mathbf{S}^T (\mathbf{I} + \mathbf{S}\mathbf{S}^T)^{-1}, \qquad \mathbf{S} = \mathbf{R}^{-\frac{1}{2}} \mathbf{H} \mathbf{A}^f / \sqrt{n-1}.$$

Anomalies-based formulation (Sakov et. al., 2010)

- Update of the **ensemble mean**:

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{A}^f \mathbf{G} \mathbf{R}^{-\frac{1}{2}} (\mathbf{z} - \mathbf{H}\mathbf{x}^f)/\sqrt{n-1}$$

with

$$\mathbf{G} = \mathbf{S}^T(\mathbf{I} + \mathbf{S}\mathbf{S}^T)^{-1}, \qquad \mathbf{S} = \mathbf{R}^{-\frac{1}{2}} \mathbf{H} \mathbf{A}^f / \sqrt{n-1}.$$

- Update of the **ensemble anomalies**: $\mathbf{A}^a = \mathbf{A}^f + \mathbf{A}^f \mathbf{T}$

Anomalies-based formulation (Sakov et. al., 2010)

- Update of the **ensemble mean**:

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{A}^f \mathbf{G}\mathbf{R}^{-\frac{1}{2}}(\mathbf{z} - \mathbf{H}\mathbf{x}^f)/\sqrt{n-1}$$

with

$$\mathbf{G} = \mathbf{S}^T(\mathbf{I} + \mathbf{S}\mathbf{S}^T)^{-1}, \qquad \mathbf{S} = \mathbf{R}^{-\frac{1}{2}}\mathbf{H}\mathbf{A}^f/\sqrt{n-1}.$$

- Update of the **ensemble anomalies**: $\mathbf{A}^a = \mathbf{A}^f + \mathbf{A}^f\mathbf{T}$
  - ▷ **EnKF**: $\mathbf{T} = \mathbf{G}(\mathbf{D} - \mathbf{S})$ where the columns of $\mathbf{D}$ are Gaussian samples
  - ▷ **ETKF**: $\mathbf{T} = (\mathbf{I}_n + \mathbf{S}^T\mathbf{S})^{-\frac{1}{2}} - \mathbf{I}_n$
  - ▷ **DEnKF**: $\mathbf{T} = -\frac{1}{2}\mathbf{G}\mathbf{S}$

QL-based formulation

- Deriving the filters using **QL** instead of the anomalies $\mathbf{A}^f$:

$$\mathbf{P}^f = \left(\frac{\mathbf{A}^f}{\sqrt{n-1}}\right)\left(\frac{\mathbf{A}^f}{\sqrt{n-1}}\right)^T \quad \longrightarrow \quad \mathbf{P}^f = \mathbf{QL}(\mathbf{QL})^T$$

QL-based formulation

- Deriving the filters using **QL** instead of the anomalies $\mathbf{A}^f$:

$$\mathbf{P}^f = \left(\frac{\mathbf{A}^f}{\sqrt{n-1}}\right)\left(\frac{\mathbf{A}^f}{\sqrt{n-1}}\right)^T \quad \longrightarrow \quad \mathbf{P}^f = \mathbf{QL}(\mathbf{QL})^T$$

- **QL** not similar to an anomalies matrix:
  - ▷ $\mathbf{QL} \in \mathbb{R}^{m\times p}$, with $p \leq n-1$ and $\mathbf{A} \in \mathbb{R}^{m\times n}$ of rank $n-1$.
  - ▷ $\mathbf{QLe} \neq 0$.

QL-based formulation

- Deriving the filters using **QL** instead of the anomalies $\mathbf{A}^f$:

$$\mathbf{P}^f = \left(\frac{\mathbf{A}^f}{\sqrt{n-1}}\right)\left(\frac{\mathbf{A}^f}{\sqrt{n-1}}\right)^T \quad \longrightarrow \quad \mathbf{P}^f = \mathbf{QL}(\mathbf{QL})^T$$

- **QL** not similar to an anomalies matrix:
  - ▷ $\mathbf{QL} \in \mathbb{R}^{m\times p}$, with $p \leq n-1$ and $\mathbf{A} \in \mathbb{R}^{m\times n}$ of rank $n-1$.
  - ▷ $\mathbf{QLe} \neq 0$.

- To interpret **QL** as an anomalies matrix:

$$\mathbf{P}^f = \mathbf{QL}(\mathbf{QL})^T = \mathbf{QLV}^T(\mathbf{QLV}^T)^T$$

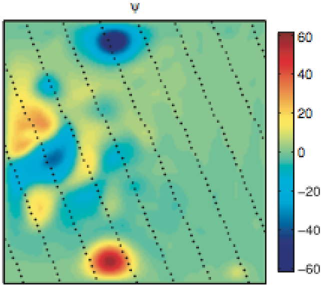
where $\mathbf{V} \in \mathbb{R}^{n\times p}$, $\mathbf{V}^T\mathbf{V} = \mathbf{I}$, $\mathbf{Ve} = 0$, e.g.,
  - ▷ right singular vectors from the SVD of $\mathbf{A}^f$
  - ▷ random orthogonal matrices with columns orthogonal to **e** (SEIK filter; Hoteit et. al., 2002).

Outline

- Motivation
- A framework for maximum likelihood estimation
- Maximum likelihood estimation with multi-ensembles
- Application to ensemble-based Kalman filters
- Numerical experiments

A quasi-geostrophic model

- Model
  - ▷ a 1.5-layer reduced-gravity QG model with double-gyre wind forcing and biharmonic friction
  - ▷ nonlinear, 129×129 grid points
  - ▷ dimension $m = 1.6 \times 10^4$
  - ▷ EnKF-Matlab toolbox (Sakov, 2013).
- Ensembles
  - ▷ "Accurate" ensemble $\mathbf{X}$: $n_x = 5$ members.
  - ▷ "Noisy" ensemble $\mathbf{Y}$: $n_y = 20$ members, truncation to 1 digit before the analysis.
  - ▷ Noise covariance matrix $\mathbf{C}$: diagonal (variance estimated from a large sample).

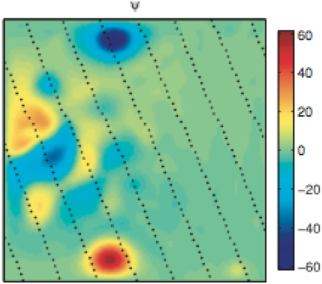


Water height, observations.

Sakov and Oke (2008)

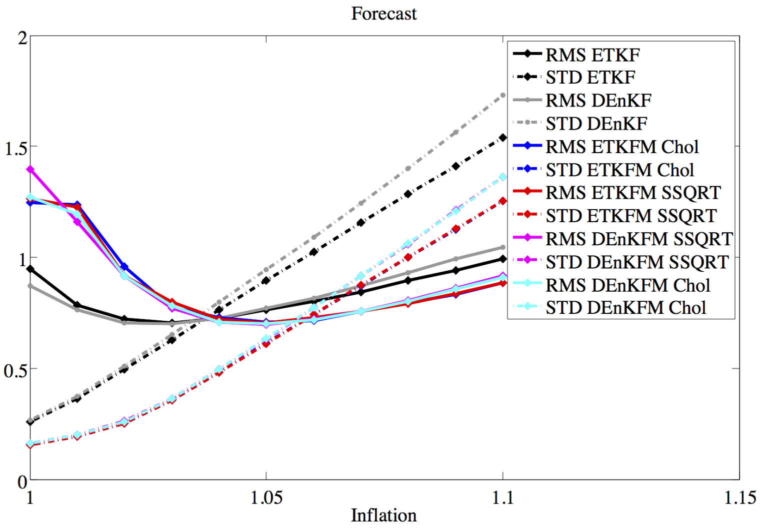
A quasi-geostrophic model

- Observations
  - ▷ Normal distributed additive noise: $\mathcal{N}(0, \sigma_o^2 \mathbf{I})$.
  - ▷ $\mathbf{R} = \sigma_o^2 \mathbf{I}$.
  - ▷ $\sigma_0^2 = 4$, equidistant tracks, every fourth time step.
- Moderation
  - ▷ Inflation: 1.0 : 0.01 : 1.1.
  - ▷ Local analysis: localization radius $= 10$.
- Anomalies-like ensemble-based Kalman filters.
- Configuration similar to Sakov and Oke (2008).



Water height, observations.

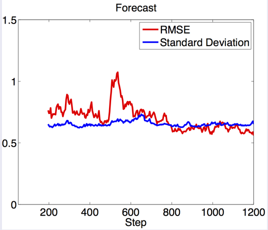
Sakov and Oke (2008)



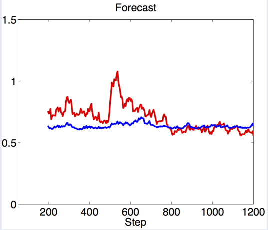
	ETKF	ETKFM (C)	ETKFM (S)	DEnKF	DEnKFM (C)	DEnKFM (S)
RMS	0.7044	0.7067	0.7043	0.7025	0.6999	0.6991
STD	0.6293	0.6138	0.6121	0.6534	0.6343	0.6323

Temporal evolution: errors

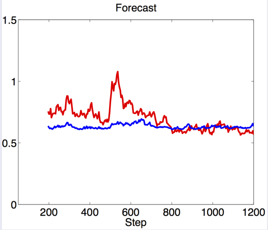
DEnKF



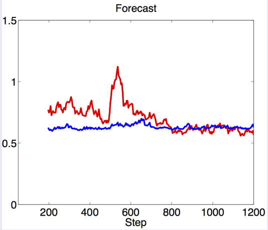
DEnKFM (Chol)



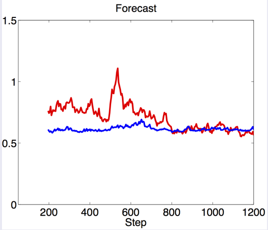
DEnKFM (SSQRT)



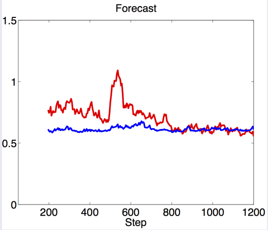
ETKF



ETKFM (Chol)

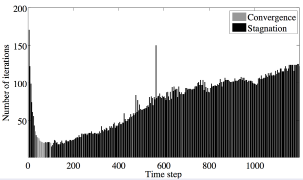


ETKFM (SSQRT)



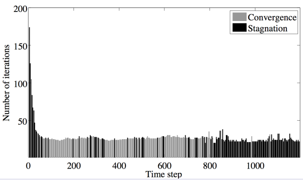
Temporal evolution: BFGS iterations [DEnKFM (SSQRT)]

$infl = 1.00$



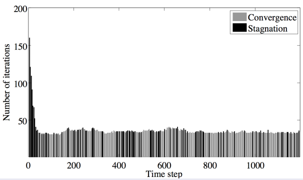
$Mean = 73.7, \; Conv = 4.3\%$

$infl = 1.03$



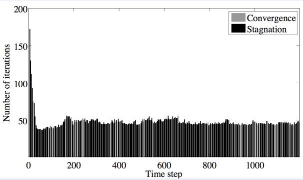
$Mean = 28.1, \; Conv = 51.7\%$

$infl = 1.05$



$Mean = 36.6, \; Conv = 67\%$

$infl = 1.08$



$Mean = 48.5, \; Conv = 12.3\%$

Summary

- A strategy for estimating mean and covariance matrix of multi-ensembles has been suggested.
  - ▷ Normally distributed random vectors: same mean, covariance matrices of the additional "noise" known.
  - ▷ Estimation in a subspace of smaller dimension.
  - ▷ Local minima can be computed numerically.
- Application to ensemble-based Kalman filters.
  - ▷ Derivation of a **QL**-based EnKF.
  - ▷ Adapted to specific variants of EnKF.
  - ▷ Anomalies-like **QL**-based formulation
- Preliminary results
  - ▷ Similar performances compared to anomalies-based filters, with a shift in the inflation.
  - ▷ Issues to be worked out: convergence of BFGS, choice of subspace, modelling of noise, choice of parameters.

Perspectives

- Ongoing work and perspectives.
  - ▷ Longer experiments (more analysis cycles), sensitivity to the dimension of the problem, the noise,...
  - ▷ Different strategies for building anomalies-like **QL**-based filters.
  - ▷ Nonlinear observation operators.
  - ▷ Numerical optimization on manifolds: fixed-rank symmetric positive semidefinite matrices.
  - ▷ Biased estimator?
- Towards multigrid strategies in ensemble-based Kalman filters.

References

- Gratton S., Simon E., Titley-Peloquin D.: Covariance matrix estimation with noisy samples. In preparation.
- Hoteit I., Pham D.-T., Blum J.: A simplified reduced order Kalman filtering and application to altimetric data assimilation in tropical Pacific, *Journal of Marine Systems*, 36, 101-127, 2002.
- Sakov P., Evensen G., Bertino L.: Asynchronous data assimilation with the EnKF, *Tellus*, 62A, 24-29, 2010.
- Sakov P., Oke P. R.: A deterministic formulation of the ensemble Kalman filter: an alternative to ensemble square root filters, *Tellus*, 60A, 361-371, 2008.
- Sakov P.: EnKF-Matlab toolbox.

Thank you for your attention!