A Parallel Ensemble Kalman Filter Implementation Based On Modified Cholesky Decomposition

Elias D. Nino-Ruiz and Adrian Sandu
Computational Science Laboratory,
Virginia Polytechnic Institute and State University,
Blacksburg, VA 24060, USA
enino@vt.edu, sandu@cs.vt.edu

November 16, 2015

Table of contents

Problem Formulation I

- We want to estimate $\mathbf{x}^* \in \mathbb{R}^{n \times 1}$,

$$\mathbf{x}_k^* = \mathcal{M}_{t_{k-1} \to t_k} \left(\mathbf{x}_{k-1}^* \right),$$

  model dimension: n.
- Based on (assuming Gaussian errors):
  - A prior estimate (best estimate prior measurements):

$$\mathbf{x}^b = \mathbf{x}^* + \boldsymbol{\xi}, \text{ with } \boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}_n, \mathbf{B})$$

    where $\mathbf{B} \in \mathbb{R}^{n \times n}$ is unknown.
  - A noisy observation:

$$\mathbf{y} = \mathcal{H}(\mathbf{x}^*) + \boldsymbol{\epsilon}, \text{ with } \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}_m, \mathbf{R}),$$

    number of observed model components: m. $\mathbf{R}^{m \times m}$ is the data error covariance matrix. $\mathcal{H} : \mathbb{R}^{n \times 1} \to \mathbb{R}^{m \times 1}$. With $m \ll n$ or $m < n$.

Problem Formulation II

- Bayesian approximation (posterior state):

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{B} \cdot \mathbf{H}^T \cdot \left[\mathbf{R} + \mathbf{H} \cdot \mathbf{B} \cdot \mathbf{H}^T \right]^{-1} \cdot \mathbf{d} \in \mathbb{R}^{n \times 1}$$

  where $\mathbf{d} = \mathbf{y} - \mathcal{H}\left(\mathbf{x}^b\right) \in \mathbb{R}^{m \times 1}$ and $\mathcal{H}' \approx \mathbf{H} \in \mathbb{R}^{m \times N}$.
- How can we estimate **B**?
- Background error statistics of any model state $\mathbf{x} \in \mathbb{R}^{n \times 1}$:

$$\mathbf{x} \sim \mathcal{N}\left(\mathbf{x}^b, \, \mathbf{B}\right) \, .$$

Problem Formulation III

- Empirical moments of an ensemble:

$$\mathbf{X}^b = \left[\mathbf{x}^{b[1]}, \mathbf{x}^{b[2]}, \dots, \mathbf{x}^{b[N]}\right] \in \mathbb{R}^{n \times N}.$$

$$\mathbf{x}^b \approx \overline{\mathbf{x}}^b = \frac{1}{N} \sum_{i=1}^{N} \mathbf{x}^{b[i]} \in \mathbb{R}^{n \times 1}, \quad \mathbf{B} \approx \mathbf{P}^b = \mathbf{S} \cdot \mathbf{S}^T \in \mathbb{R}^{n \times n}$$

where $\mathbf{S} = \frac{1}{\sqrt{N-1}} \cdot \left[\mathbf{X}^b - \overline{\mathbf{x}}^b \otimes \mathbf{1}_N^T\right] \in \mathbb{R}^{n \times N}$.

Problem Formulation IV

- The posterior (analysis) ensemble:

$$\mathbf{X}^a = \mathbf{X}^b + \mathbf{P}^b \cdot \mathbf{H}^T \cdot \left[\mathbf{R} + \mathbf{H} \cdot \mathbf{P}^b \cdot \mathbf{H}^T\right]^{-1} \cdot \mathbf{D} \in \mathbb{R}^{n \times N},$$

the i-th column of $\mathbf{D} \in \mathbb{R}^{m \times N}$ reads:

$$\mathbf{d}^{[i]} = \mathbf{y} + \boldsymbol{\epsilon}^{[i]} - \mathbf{H} \cdot \mathbf{x}^{b[i]} \in \mathbb{R}^{m \times 1}, \text{ for } 1 \leq i \leq N,$$

with (stochastic version of the filter):

$$\boldsymbol{\epsilon}^{[i]} \sim \mathcal{N}\left(\mathbf{0}_m, \mathbf{R}\right) .$$

Problem Formulation V

- Unfortunately, the number of samples N is much lower than the model dimension $n \gg N$.
  - $\mathbf{P}^b$ is low-rank. (Spurious correlations)
  - $\mathbf{X}^a$ is computed in the ensemble space (few degrees of freedom)
- Model dimensions are in the order of billions, while ensemble sizes in the order of hundreds.
- Model propagations are computationally expensive.
- Computational effort of the analysis is high.
- We do need HPC not only to speedup computations but to have enough memory to represent ensemble members and to perform linear algebra computations.

Local Ensemble Transform Kalman Filter (LETKF) I

- One of the best parallel ensemble based implementations.
- Analysis equations:
  - Perturbations: $\mathbf{U} = \mathbf{X}^b - \overline{\mathbf{x}}^b \otimes \mathbf{1}_N^T \in \mathbb{R}^{n \times N}$.
  - Optimality in ensemble space: $\mathbf{Q} = \mathbf{H} \cdot \mathbf{U} \in \mathbb{R}^{m \times N}$,

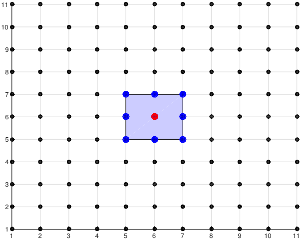
$$
\begin{aligned}
\widetilde{\mathbf{P}}^a &= \left[(N-1) \cdot \mathbf{I}_{N \times N} + \mathbf{Q}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{Q}\right]^{-1} \in \mathbb{R}^{N \times N} \\
\mathbf{w}^a &= \widetilde{\mathbf{P}}^a \cdot \mathbf{Q}^T \cdot \mathbf{R}^{-1} \cdot \left[\mathbf{y} - \mathbf{H} \cdot \overline{\mathbf{x}}^b\right] \\
\mathbf{W} &= \mathbf{w}^a \otimes \mathbf{1}_N^T + \mathbf{W}^a \in \mathbb{R}^{N \times N},\ \mathbf{W}^a = \left[(N-1) \cdot \widetilde{\mathbf{P}}^a\right]^{1/2} \in \mathbb{R}^{N \times N}
\end{aligned}
$$

  - Analysis ensemble:

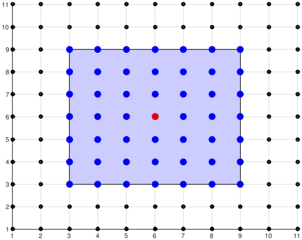
$$
\mathbf{X}^a = \overline{\mathbf{x}}^b \otimes \mathbf{1}_N^T + \mathbf{U} \cdot \mathbf{W} \in \mathbb{R}^{n \times N}.
$$

- Domain localization [OHS+04, Kep00]:

Local Ensemble Transform Kalman Filter (LETKF) II



(d) $r = 1$



(e) $r = 3$

Estimating $\mathbf{B}^{-1}$ I

- When m_i and m_j are conditionally independent, ${\mathbf{C}^{-1}}_{m_i,m_j} = 0$.
- We want to estimate $\mathbf{B}^{-1}$:
  - Recall $\mathbf{U} = \mathbf{X}^b - \overline{\mathbf{x}}^b \otimes \mathbf{1}_N^T \in \mathbb{R}^{n \times N}$. Thus, $\mathbf{u}^{[i]} \sim \mathcal{N}\left(\mathbf{0}_n,\, \mathbf{B}\right)$, for $1 \le i \le N$.
  - Let $\mathbf{x}^{[i]} \in \mathbb{R}^{N \times 1}$ the vector holding the i-th row across all columns of $\mathbf{U}$, for $1 \le i \le n$.
  - Then, the approximation of $\mathbf{B}^{-1}$ arises from:

$$\mathbf{x}^{[i]} = \sum_{j=1}^{i-1} \mathbf{x}^{[j]} \cdot \beta_{i,j} + \boldsymbol{\xi}^{[i]} \in \mathbb{R}^{N \times 1}$$

Estimating $\mathbf{B}^{-1}$ II

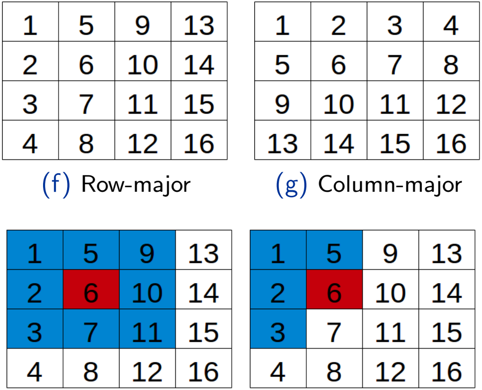
- By the modified Cholesky (MC) decomposition for inverse covariance matrix estimation:

$$\mathbf{B}^{-1} \approx \widehat{\mathbf{B}}^{-1} = \mathbf{T}^T \cdot \mathbf{D}^{-1} \cdot \mathbf{T} \in \mathbb{R}^{n \times n}$$
$$\mathbf{B} \approx \widehat{\mathbf{B}} = \mathbf{T}^{-1} \cdot \mathbf{D} \cdot \mathbf{T}^{-1^T} \in \mathbb{R}^{n \times n}$$

  where $\mathbf{T} \in \mathbb{R}^{n \times n}$ is an unitary lower triangular matrix with $\{\mathbf{T}\}_{i,j} = -\beta_{i,j}$ and $\mathbf{D} \in \mathbb{R}^{n \times n}$ is a diagonal matrix with $\{\mathbf{D}\}_{i,i} = \textbf{var}\left(\xi^{[i]}\right)$, for $1 \leq j < i \leq n$.I

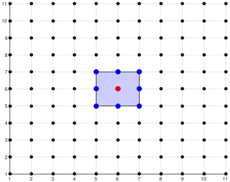
- $\widehat{\mathbf{B}}^{-1}$ can be sparse, $\widehat{\mathbf{B}}$ is not necessarily sparse. Structure of $\widehat{\mathbf{B}}^{-1}$ depends on $\mathbf{T}$.

Choosing the predecessors

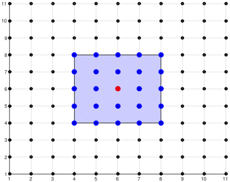


(f) Row-major (g) Column-major

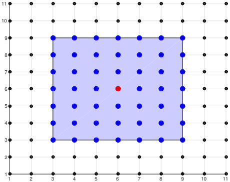
(h) $r = 1$ (i) Predecess.



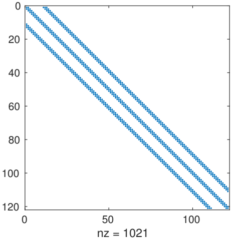
(a) $r = 1$



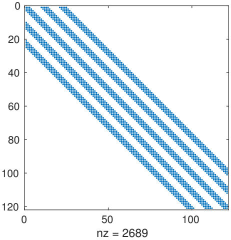
(b) $r = 2$



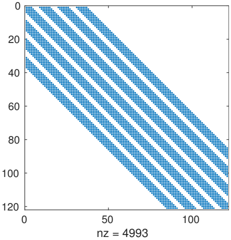
(c) $r = 3$



(d) $\widehat{\mathbf{B}}^{-1}$ for $r = 1$



(e) $\widehat{\mathbf{B}}^{-1}$ for $r = 2$



(f) $\widehat{\mathbf{B}}^{-1}$ for $r = 3$

EnKF formulations based on modified Cholesky decomposition for inverse background error estimation

- Primal:

$$\mathbf{X}^a = \mathbf{X}^b + \left[\widehat{\mathbf{B}}^{-1} + \mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{H}\right]^{-1} \cdot \mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot \left[\mathbf{Y}^s - \mathbf{H} \cdot \mathbf{X}^b\right]$$

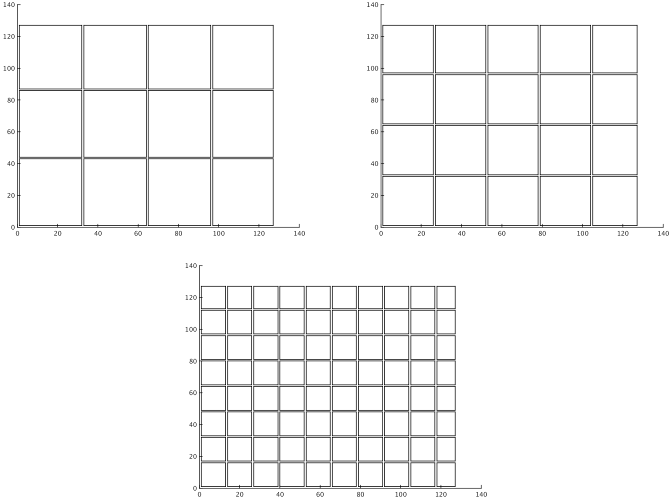
- Dual:

$$\mathbf{X}^a = \mathbf{X}^b + \mathbf{X} \cdot \mathbf{V}^T \cdot \left[\mathbf{R} + \mathbf{V} \cdot \mathbf{V}^T\right]^{-1} \cdot \left[\mathbf{Y}^s - \mathbf{H} \cdot \mathbf{H} \cdot \mathbf{X}^b\right]$$

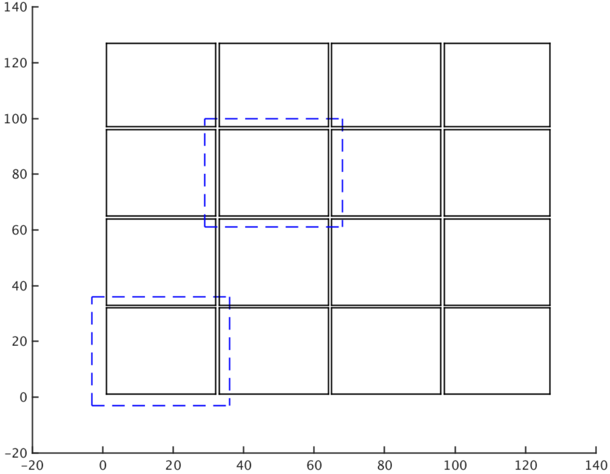
  where $\mathbf{T} \cdot \mathbf{X} = \mathbf{D}^{1/2} \in \mathbb{R}^{n \times n}$ and $\mathbf{V} = \mathbf{H} \cdot \mathbf{X} \in \mathbb{R}^{m \times n}$.

- Efficient implementations [NRSA14, NRS15].

Domain Decomposition



Boundary Information



AT-GCM - SPEEDY (Numerical Model)

- SPEEDY is a simplified GCM developed at ICTP by Franco Molteni and Fred Kucharski.
- Nicknamed SPEEDY, for "Simplified Parameterizations, privitivE-Equation DYnamics"
- It is a hydrostatic, s-coordinate, spectral-transform model in the vorticity-divergence form, with semi-implicit treatment of gravity waves.
- 8 layers, u, v, T and sh.
- T-63 resolution (96 x 192)

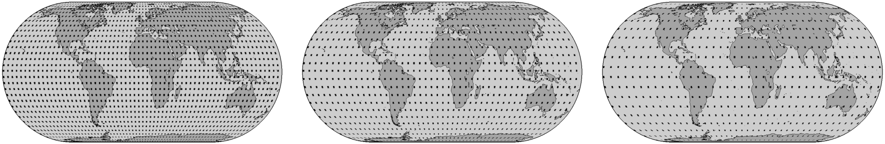
Blueridge Super Computer @ VT

- BlueRidge is a 408-node Cray CS-300 cluster.
- Each node is outfitted with two octa-core Intel Sandy Bridge CPUs and 64 GB of memory.
- Total of 6,528 cores and 27.3 TB of memory systemwide.
- Eighteen nodes have 128 GB of memory.
- In addition, 130 nodes are outfitted with two Intel MIC (Xeon Phi) coprocessors.

Experimental settings

- Number of ensemble members 96.
- 3 radius of influence are considered: 3, 4, 5.
- Model is propagated for 2 days and then observations are assimilated.
- Number of processors: 6 computing nodes (96 processors) up to 128 computing nodes (2048 processors)
- Fortran 90 and 77, MPI, LAPACK and BLAS.
- 3 different observational networks.

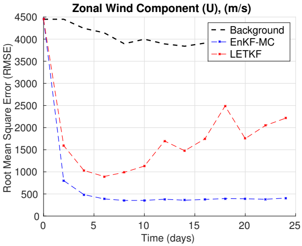
Observational networks



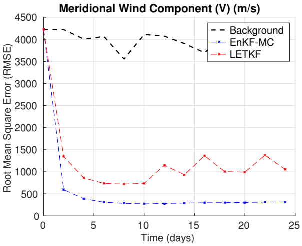
(a) $\mathbf{H}^{[1]}$, $p \sim 12\%$ (b) $\mathbf{H}^{[2]}$, $p \sim 6\%$ (c) $\mathbf{H}^{[3]}$, $p \sim 4\%$

Figure : Sparse observational networks. Observed components in black. p denotes percentage of observed model components.

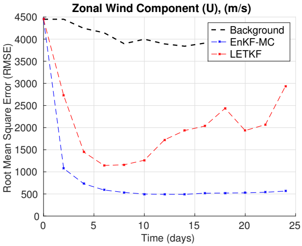
RMSE for some configurations.



(a) $r = 3,\ p \sim 12\%,\ u$

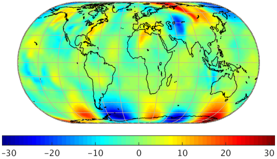


(b) $r = 4,\ p \sim 12\%,\ v$

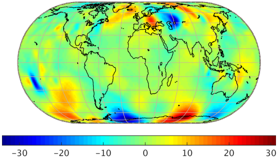


(c) $r = 5,\ p \sim 6\%,\ u$

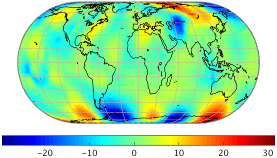
Initial snapshots for $r = 5$ and $p \sim 4\%$ for v



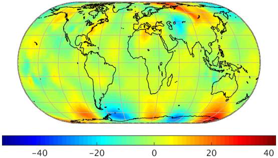
(a) Reference



(b) Background

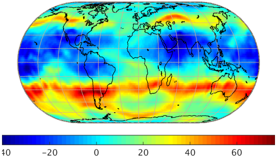


(c) EnKF-MC

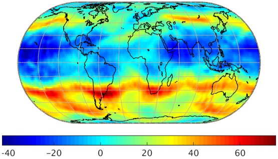


(d) LETKF

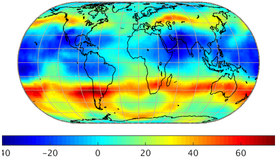
Initial snapshots for $r = 5$ and $p \sim 4\%$ for u



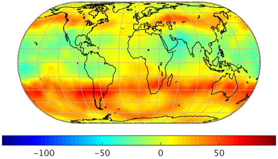
(a) Reference



(b) Background

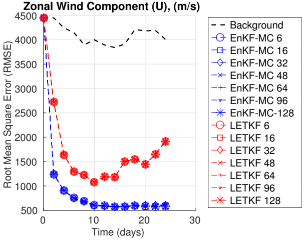


(c) EnKF-MC

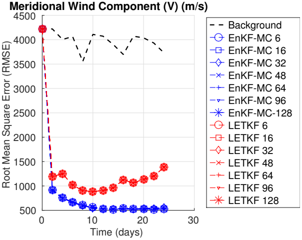


(d) LETKF

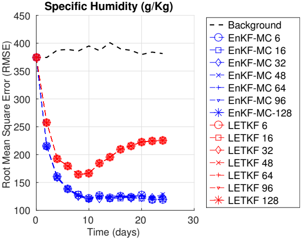
RMSE for different variables and # of computing nodes.



(a) $r = 5,\ p \sim 4\%,\ u$

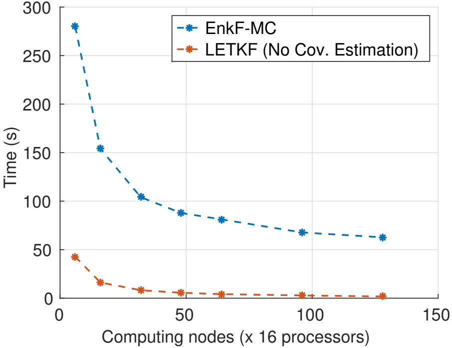


(b) $r = 5,\ p \sim 4\%,\ v$



(c) $r = 5,\ p \sim 4\%,\ sh$

Elapsed time.



Conclusions

- The proposed implementations outperforms the LETKF under the RMSE metric.
- Parallel resources and domain decompositions can be exploited in order to speedup the assimilation process.
- Localization is implicit. Domain decomposition is used just for computational reasons.
- The computational effort of the proposed method makes it attractive for the use under realistic scenarios.

Thank You.

(John 3:16) For God so loved the world that he gave his one and only Son, that whoever believes in him shall not perish but have eternal life.

[Kep00] Christian L. Keppenne. Data Assimilation into a Primitive-Equation Model with a Parallel Ensemble Kalman Filter. Monthly Weather Review, 128(6):1971–1981, 2000.

[NRS15] EliasD. Nino-Ruiz and Adrian Sandu. Ensemble kalman filter implementations based on shrinkage covariance matrix estimation. Ocean Dynamics, 65(11):1423–1439, 2015.

[NRSA14] EliasD. Nino Ruiz, Adrian Sandu, and Jeffrey Anderson. An Efficient Implementation of the Ensemble Kalman Filter Based on an Iterative ShermanMorrison Formula. Statistics and Computing, pages 1–17, 2014.

[OHS+04] Edward Ott, Brian R. Hunt, Istvan Szunyogh, Aleksey V. Zimin, Eric J. Kostelich, Matteo Corazza, Eugenia Kalnay, D. J. Patil, and James A. Yorke. A local ensemble kalman filter for atmospheric data assimilation. Tellus A, 56(5):415–428, 2004.

$$
\begin{aligned}
\left[\widehat{\mathbf{B}}^{-1} + \mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{H}\right]^{-1} &= \left[\mathbf{X}^T \cdot \mathbf{X} + \mathbf{H}^T \cdot \mathbf{R}^{-1}\mathbf{H}\right]^{-1} \\
&= \left\{\mathbf{X}^T \cdot \left[\mathbf{I}_{n\times n} + \mathbf{Q} \cdot \mathbf{Q}^T\right] \cdot \mathbf{X}\right\}^{-1} \\
&= \mathbf{X}^{-1} \cdot \left[\mathbf{I}_{n\times n} + \mathbf{Q} \cdot \mathbf{Q}^T\right]^{-1} \cdot \mathbf{X}^{-T}
\end{aligned}
$$

where $\mathbf{X}^T \cdot \mathbf{Q} = \mathbf{H}^T \cdot \mathbf{R}^{-1/2}$.