Efficient Subspace Implementation of an Iterative Ensemble Smoother for Solving Inverse Problems

Geir Evensen

NORCE

NORCE–Norwegian Research Center

SIAM CSE 2021, MS301 Advances in Data Assimilation, March 5th, 2021

Parameter-estimation problem

Nonlinear model and measurements

$$\boldsymbol{y} = \boldsymbol{g}(\boldsymbol{x}) \qquad \boldsymbol{d} \leftarrow \boldsymbol{y} + \boldsymbol{e}$$

Bayesian formulation

$$f(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{d}) \propto f(\boldsymbol{d}|\boldsymbol{y})f(\boldsymbol{y}|\boldsymbol{x})f(\boldsymbol{x})$$

- Problems with a high-dimensional $\boldsymbol{x} \in \Re^n$ and vast datasets $\boldsymbol{d} \in \Re^m$.
- Standard History-Matching problem for oil-reservoir models.

“Indirect” DA update

Analog DA problem with a nonlinear model

$$\boldsymbol{x}_{i+1} = \boldsymbol{m}(\boldsymbol{x}_i)$$

Nonlinear measurement operator and measurements

$$\boldsymbol{y} = \boldsymbol{h}(\boldsymbol{x}_{i+1}) = \boldsymbol{h}\big(\boldsymbol{m}(\boldsymbol{x}_i)\big) = \boldsymbol{g}(\boldsymbol{x}_i), \qquad \boldsymbol{d} \leftarrow \boldsymbol{y} + \boldsymbol{e}$$

Bayesian formulation

$$f(\boldsymbol{x}_i, \boldsymbol{y}|\boldsymbol{d}) \propto f(\boldsymbol{d}|\boldsymbol{y}) f(\boldsymbol{y}|\boldsymbol{x}_i) f(\boldsymbol{x}_i)$$

- Smoother update step in sequential data assimilation

Marginal pdf

Nonlinear model and measurements

$$\boldsymbol{y} = \boldsymbol{g}(\boldsymbol{x}) \qquad \boldsymbol{d} \leftarrow \boldsymbol{y} + \boldsymbol{e}$$

Perfect-model pdf

$$f(\boldsymbol{y}|\boldsymbol{x}) = \delta\big(\boldsymbol{y} - \boldsymbol{g}(\boldsymbol{x})\big)$$

Bayesian formulation

$$f(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{d}) \propto f(\boldsymbol{d}|\boldsymbol{y}) f(\boldsymbol{y}|\boldsymbol{x}) f(\boldsymbol{x})$$

Marginal pdf

$$f(\boldsymbol{x}|\boldsymbol{d}) \propto \int f(\boldsymbol{d}|\boldsymbol{y}) f(\boldsymbol{y}|\boldsymbol{x}) f(\boldsymbol{x}) d\boldsymbol{y} = f\big(\boldsymbol{d}|\boldsymbol{g}(\boldsymbol{x})\big) f(\boldsymbol{x})$$

Gaussian priors

Model and observations

$$\boldsymbol{y} = \boldsymbol{g}(\boldsymbol{x}) \qquad \boldsymbol{d} \leftarrow \boldsymbol{y} + \boldsymbol{e}$$

Bayes

$$f(\boldsymbol{x} \,|\, \boldsymbol{d}) \propto f(\boldsymbol{x}) f(\boldsymbol{d} \,|\, \boldsymbol{g}(\boldsymbol{x})).$$

MAP estimate

$$\mathcal{J}(\boldsymbol{x}) = \big(\boldsymbol{x} - \boldsymbol{x}^{\mathrm{f}}\big)^{\mathrm{T}} \boldsymbol{C}_{zz}^{-1} \big(\boldsymbol{x} - \boldsymbol{x}^{\mathrm{f}}\big) + \big(\boldsymbol{g}(\boldsymbol{x}) - \boldsymbol{d}\big)^{\mathrm{T}} \boldsymbol{C}_{dd}^{-1} \big(\boldsymbol{g}(\boldsymbol{x}) - \boldsymbol{d}\big).$$

Ensemble formulation for approximate sampling of $f(\boldsymbol{x} \,|\, \boldsymbol{d})$ (normal priors)

$$\mathcal{J}(\boldsymbol{x}_j) = \big(\boldsymbol{x}_j - \boldsymbol{x}^{\mathrm{f}}_j\big)^{\mathrm{T}} \boldsymbol{C}_{zz}^{-1} \big(\boldsymbol{x}_j - \boldsymbol{x}^{\mathrm{f}}_j\big) + \big(\boldsymbol{g}(\boldsymbol{x}_j) - \boldsymbol{d}_j\big)^{\mathrm{T}} \boldsymbol{C}_{dd}^{-1} \big(\boldsymbol{g}(\boldsymbol{x}_j) - \boldsymbol{d}_j\big).$$

Ensemble subspace Randomized Maximum Likelihood

Ensemble representation of prior error covariances (Chen and Oliver, 2013)

$$\mathcal{J}(\boldsymbol{x}_j) = \left(\boldsymbol{x}_j - \boldsymbol{x}_j^{\mathrm{f}}\right)^{\mathrm{T}} \overline{\boldsymbol{C}}_{xx}^{-1} \left(\boldsymbol{x}_j - \boldsymbol{x}_j^{\mathrm{f}}\right) + \left(\boldsymbol{g}(\boldsymbol{x}_j) - \boldsymbol{d}_j\right)^{\mathrm{T}} \boldsymbol{C}_{dd}^{-1} \left(\boldsymbol{g}(\boldsymbol{x}_j) - \boldsymbol{d}_j\right).$$

Solution contained in the ensemble subspace (Evensen et al., 2019, Raanes et al., 2019).

$$\boldsymbol{x}_j^{\mathrm{a}} = \boldsymbol{x}_j^{\mathrm{f}} + \boldsymbol{A}\boldsymbol{w}_j,$$

with $\boldsymbol{A} = \boldsymbol{X}\boldsymbol{\Pi}$ being the ensemble anomalies and we get,

$$\mathcal{J}(\boldsymbol{w}_j) = \boldsymbol{w}_j^{\mathrm{T}} \boldsymbol{w}_j + \Big(\boldsymbol{g}\big(\boldsymbol{x}_j^{\mathrm{f}} + \boldsymbol{A}\boldsymbol{w}_j\big) - \boldsymbol{d}_j\Big)^{\mathrm{T}} \boldsymbol{C}_{dd}^{-1} \Big(\boldsymbol{g}\big(\boldsymbol{x}_j^{\mathrm{f}} + \boldsymbol{A}\boldsymbol{w}_j\big) - \boldsymbol{d}_j\Big)$$

Reduces dimension of problem from state size to ensemble size (Hunt et al., 2007).

Gradient and Hessian of cost function

Gradient

$$\nabla \mathcal{J}(\boldsymbol{w}_j) = 2\boldsymbol{w}_j + 2\big(\boldsymbol{G}_j \boldsymbol{A}\big)^{\mathrm{T}} \boldsymbol{C}_{dd}^{-1} \Big(\boldsymbol{g}\big(\boldsymbol{x}_j^{\mathrm{f}} + \boldsymbol{A}\boldsymbol{w}_j\big) - \boldsymbol{d}_j\Big),$$

Hessian (approximate)

$$\nabla\nabla \mathcal{J}(\boldsymbol{w}_j) \approx 2\boldsymbol{I} + 2\big(\boldsymbol{G}_j \boldsymbol{A}\big)^{\mathrm{T}} \boldsymbol{C}_{dd}^{-1} \big(\boldsymbol{G}_j \boldsymbol{A}\big)$$

Gauss-Newton iterations

$$\boxed{\begin{aligned}\boldsymbol{w}_j^{i+1} = \boldsymbol{w}_j^i - \gamma\Bigg\{&\boldsymbol{w}_j^i - \left(\boldsymbol{G}_j^i\boldsymbol{A}\right)^{\mathrm{T}}\Big(\left(\boldsymbol{G}_j^i\boldsymbol{A}\right)\left(\boldsymbol{G}_j^i\boldsymbol{A}\right)^{\mathrm{T}} + \boldsymbol{C}_{dd}\Big)^{-1} \\ &\times \Big(\left(\boldsymbol{G}_j^i\boldsymbol{A}\right)\boldsymbol{w}_j^i + \boldsymbol{d}_j - \boldsymbol{g}\left(\boldsymbol{x}_j^{\mathrm{f}} + \boldsymbol{A}\boldsymbol{w}_j^i\right)\Big)\Bigg\}.\end{aligned}}$$

with

$$\boldsymbol{G}_j^i = \left(\nabla \boldsymbol{g}|_{\boldsymbol{x}_j^{\mathrm{f}} + \boldsymbol{A}\boldsymbol{w}_j^i}\right)^{\mathrm{T}}.$$

$G^i_j A$

Replace G^i_j with average sensitivity G^i

$\boldsymbol{G}_j^i \boldsymbol{A}$

Replace $\boldsymbol{G}_j^i$ with average sensitivity $\boldsymbol{G}^i$

Define the linear regression

$$\overline{\boldsymbol{G}}_i = \boldsymbol{Y}_i \boldsymbol{A}_i^{+}$$

$$\boldsymbol{Y}_i = \boldsymbol{g}\big(\boldsymbol{X}_i\big)\boldsymbol{\Pi}$$

$G^i_j A$

Replace G^i_j with average sensitivity G^i

Define the linear regression

$$\overline{G}_i = Y_i A_i^+$$

Write

$$G^i_j A \approx \overline{G}^i A = Y_i A_i^+ A$$

$$Y_i = g(X_i)\Pi$$

$\boldsymbol{G}_j^i \boldsymbol{A}$

Replace $\boldsymbol{G}_j^i$ with average sensitivity $\overline{\boldsymbol{G}}^i$

Define the linear regression

$$\overline{\boldsymbol{G}}_i = \boldsymbol{Y}_i \boldsymbol{A}_i^+$$

Write

$$\begin{aligned} \boldsymbol{G}_j^i \boldsymbol{A} \approx \overline{\boldsymbol{G}}^i \boldsymbol{A} &= \boldsymbol{Y}_i \boldsymbol{A}_i^+ \boldsymbol{A} \\ &= \boldsymbol{Y}_i \boldsymbol{A}_i^+ \boldsymbol{A}_i \boldsymbol{\Omega}_i^{-1} \end{aligned}$$

$$\boldsymbol{A} = \boldsymbol{A}_i \boldsymbol{\Omega}_i^{-1}$$

$$\boldsymbol{Y}_i = \boldsymbol{g}(\boldsymbol{X}_i)\boldsymbol{\Pi}$$

$$\boldsymbol{\Omega}_i = \boldsymbol{I} + \boldsymbol{W}_i \boldsymbol{\Pi}$$

$G_j^i A$

Replace G_j^i with average sensitivity G^i

Define the linear regression

$$\overline{G}_i = Y_i A_i^+$$

Write

$$\begin{aligned} G_j^i A \approx \overline{G}^i A &= Y_i A_i^+ A \\ &= Y_i A_i^+ A_i \Omega_i^{-1} && A = A_i \Omega_i^{-1} \\ &= Y_i \Omega_i^{-1} = S_i && \text{when } n \geq N-1 \end{aligned}$$

$$Y_i = g(X_i)\Pi$$

$$\Omega_i = I + W_i \Pi$$

Iteration formula for $\boldsymbol{W}_i$

Standard form ($\mathcal{O}(m^3)$)

$$\boldsymbol{W}_{i+1} = \boldsymbol{W}_i - \gamma\Big(\boldsymbol{W}_i - \boldsymbol{S}_i^{\mathrm{T}}\left(\boldsymbol{S}_i\boldsymbol{S}_i^{\mathrm{T}} + \boldsymbol{C}_{dd}\right)^{-1}\left(\boldsymbol{S}_i\boldsymbol{W}_i + \boldsymbol{D} - \boldsymbol{g}(\boldsymbol{X}_i)\right)\Big)$$

From Woodbury, rewrite as

$$\boldsymbol{W}_{i+1} = \boldsymbol{W}_i - \gamma\Big\{\boldsymbol{W}_i - \left(\boldsymbol{S}_i^{\mathrm{T}}\boldsymbol{C}_{dd}^{-1}\boldsymbol{S}_i + \boldsymbol{I}_N\right)^{-1}\boldsymbol{S}_i^{\mathrm{T}}\boldsymbol{C}_{dd}^{-1}\left(\boldsymbol{S}_i\boldsymbol{W}_i + \boldsymbol{D} - \boldsymbol{g}(\boldsymbol{X}_i)\right)\Big\}$$

For $\boldsymbol{C}_{dd} = \boldsymbol{I}_m$ we have ($\mathcal{O}(mN^2)$)

$$\boldsymbol{W}_{i+1} = \boldsymbol{W}_i - \gamma\Big\{\boldsymbol{W}_i - \left(\boldsymbol{S}_i^{\mathrm{T}}\boldsymbol{S}_i + \boldsymbol{I}_N\right)^{-1}\boldsymbol{S}_i^{\mathrm{T}}\left(\boldsymbol{S}_i\boldsymbol{W}_i + \boldsymbol{D} - \boldsymbol{g}(\boldsymbol{X}_i)\right)\Big\}$$

Subspace inversion represents $C_{dd} \approx \boldsymbol{E}\boldsymbol{E}^{\mathrm{T}}$

- Algorithm by Evensen (2004) works directly with $\boldsymbol{E}$.

$$
\begin{aligned}
&(\boldsymbol{S}\boldsymbol{S}^{\mathrm{T}} + \boldsymbol{E}\boldsymbol{E}^{\mathrm{T}}) \\
&\qquad \approx \boldsymbol{S}\boldsymbol{S}^{\mathrm{T}} + (\boldsymbol{S}\boldsymbol{S}^{+})\boldsymbol{E}\boldsymbol{E}^{\mathrm{T}}(\boldsymbol{S}\boldsymbol{S}^{+})^{\mathrm{T}} \\
&\qquad = \boldsymbol{U}\boldsymbol{\Sigma}\left(\boldsymbol{I}_N + \boldsymbol{\Sigma}^{+}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{E}\boldsymbol{E}^{\mathrm{T}}\boldsymbol{U}(\boldsymbol{\Sigma}^{+})^{\mathrm{T}}\right)\boldsymbol{\Sigma}^{\mathrm{T}}\boldsymbol{U}^{\mathrm{T}} \\
&\qquad = \boldsymbol{U}\boldsymbol{\Sigma}\left(\boldsymbol{I}_N + \boldsymbol{Z}\boldsymbol{\Lambda}\boldsymbol{Z}^{\mathrm{T}}\right)\boldsymbol{\Sigma}^{\mathrm{T}}\boldsymbol{U}^{\mathrm{T}} \\
&\qquad = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{Z}\left(\boldsymbol{I}_N + \boldsymbol{\Lambda}\right)\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\Sigma}^{\mathrm{T}}\boldsymbol{U}^{\mathrm{T}}.
\end{aligned}
$$

$$
\left(\boldsymbol{S}\boldsymbol{S}^{\mathrm{T}} + \boldsymbol{E}\boldsymbol{E}^{\mathrm{T}}\right)^{-1} \approx \boldsymbol{U}(\boldsymbol{\Sigma}^{+})^{\mathrm{T}}\boldsymbol{Z}\left(\boldsymbol{I}_N + \boldsymbol{\Lambda}\right)^{-1}\left(\boldsymbol{U}(\boldsymbol{\Sigma}^{+})^{\mathrm{T}}\boldsymbol{Z}\right)^{\mathrm{T}}
$$

Computational cost is $\mathcal{O}(mN^2)$.

Subspace EnRML algorithm: (Evensen et al., 2019)

Input: $\boldsymbol{X}_0 \in \Re^{n \times N}$ (prior model ensemble)

Input: $\boldsymbol{D} \in \Re^{m \times N}$ (perturbed measurements)

$\boldsymbol{W}_0 = 0$ $\qquad \boldsymbol{W} \in \Re^{N \times N}$

$\boldsymbol{\Pi} = \left(\boldsymbol{I} - \frac{1}{N}\boldsymbol{1}\boldsymbol{1}^{\mathrm{T}}\right)/\sqrt{N-1}$ $\qquad \boldsymbol{\Pi} \in \Re^{N \times N}$

$\boldsymbol{E} = \boldsymbol{D}\boldsymbol{\Pi}$ $\qquad \boldsymbol{E} \in \Re^{m \times N}$

i=0

repeat

$\quad \boldsymbol{Y}_i = \boldsymbol{g}(\boldsymbol{X}_i)\boldsymbol{\Pi}$ $\qquad \boldsymbol{Y} \in \Re^{m \times N}$

$\quad \boldsymbol{\Omega}_i = \boldsymbol{I} + \boldsymbol{W}_i\boldsymbol{\Pi}$ $\qquad \boldsymbol{\Omega} \in \Re^{N \times N}$

$\quad \boldsymbol{S}_i = \boldsymbol{Y}_i\boldsymbol{\Omega}_i^{-1}$ $\qquad \boldsymbol{S} \in \Re^{m \times N}$

$\quad \boldsymbol{H}_i = \boldsymbol{S}_i\boldsymbol{W}_i + \boldsymbol{D} - \boldsymbol{g}(\boldsymbol{X}_i)$ $\qquad \boldsymbol{H} \in \Re^{m \times N}$

$\quad \boldsymbol{W}_{i+1} = \boldsymbol{W}_i - \gamma\left(\boldsymbol{W}_i - \boldsymbol{S}_i^{\mathrm{T}}\left(\boldsymbol{S}_i\boldsymbol{S}_i^{\mathrm{T}} + \boldsymbol{E}\boldsymbol{E}^{\mathrm{T}}\right)^{-1}\boldsymbol{H}_i\right)$

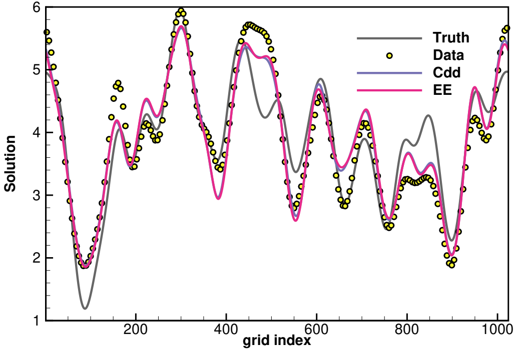
$\quad \boldsymbol{T}_i = \left(\boldsymbol{I} + \boldsymbol{W}_{i+1}\Big/\sqrt{N-1}\right)$ $\qquad \boldsymbol{T} \in \Re^{N \times N}$

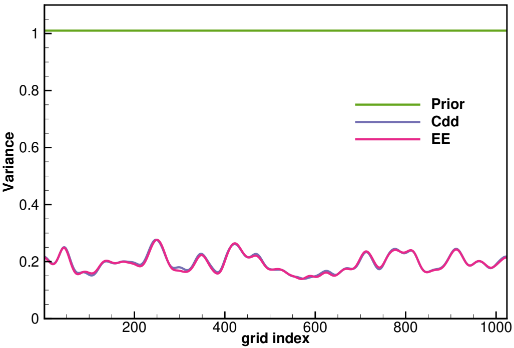
$\quad \boldsymbol{X}_{i+1} = \boldsymbol{X}\boldsymbol{T}_i$

$\quad i = i + 1$

until convergence

Subspace inversion with many measurements (Evensen, 2021)



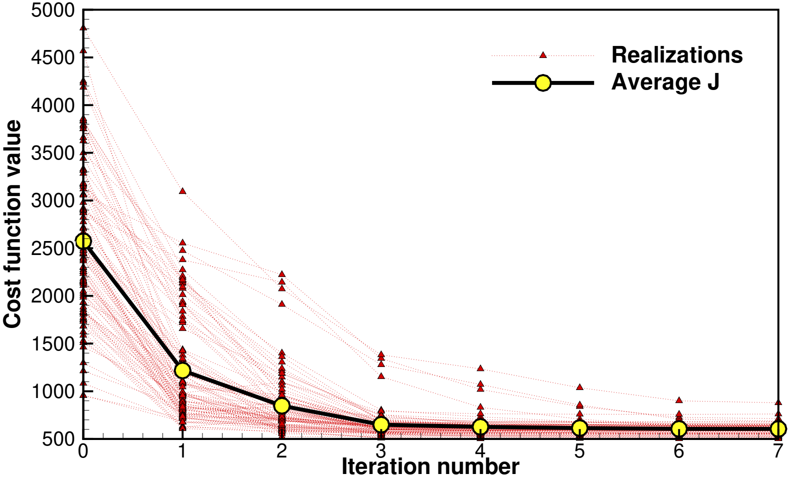


- Ensemble size $N = 100$, $\boldsymbol{E} \in \Re^{m \times 10N}$.
- Number of measurements $m = 200$ with correlated errors.
- State size $n = 1000$.

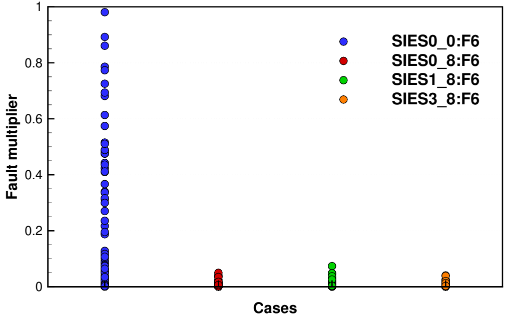
Reservoir case

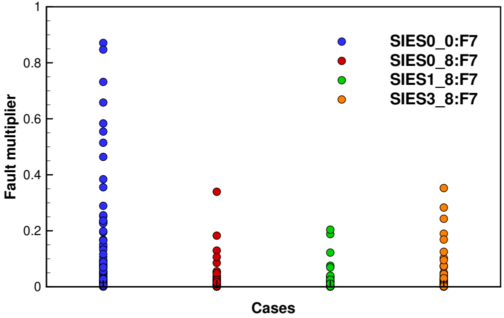
- $n = 28000$ model parameters (3D porosity field and fault multipliers)
- $m = 453$ measured production rates from six wells.
- $N = 100$ ensemble size.

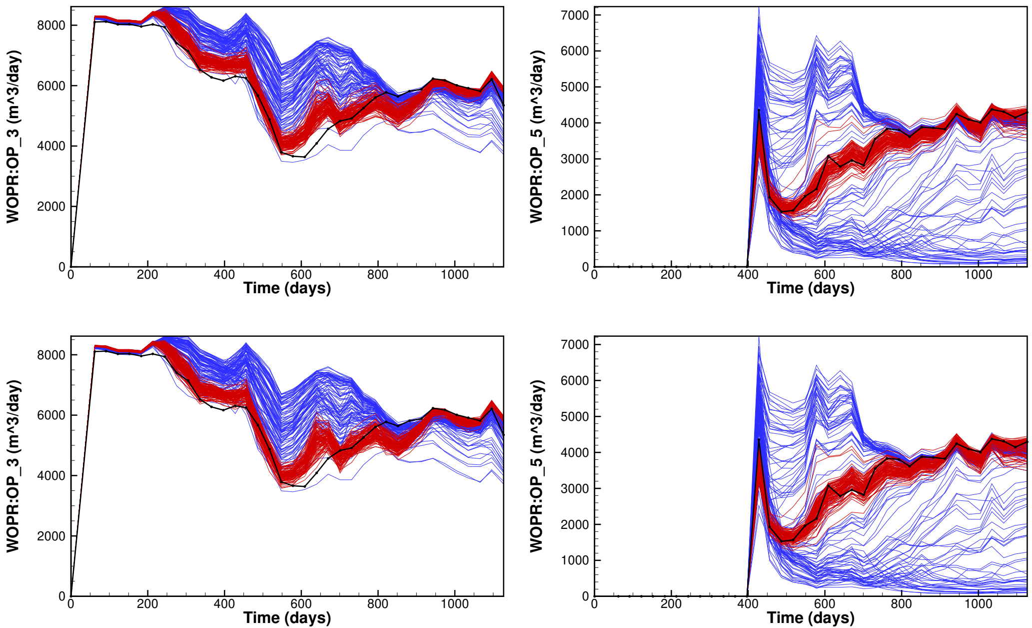
Ensemble of cost functions converges very fast



Fault multipliers F6 and F7






WOPR:OP_3 (m^3/day)
WOPR:OP_5 (m^3/day)
Time (days)
0
200
400
600
800
1000
2000
3000
4000
5000
6000
7000
8000

Summary: Ensemble subspace RML

- Approximately samples posterior pdf for nonlinear problems.
- Avoids model adjoints by using an ensemble averaged model sensitivity.
- Computationally efficient formulation and implementation.
- Cost is linear in number of measurements and state size.
- No inversion or factorization of large matrices.
- Allows for correlated measurement errors and a nondiagonal $\boldsymbol{C}_{dd}$.
- Allows for localization through local analysis Neto et al. (2020).
- For the case with additional model errors see Evensen (2019).
- For the case with additional controls or forcing see Evensen (2021).

References

Chen, Y. and D. S. Oliver. Levenberg-Marquardt forms of the iterative ensemble smoother for efficient history matching and uncertainty quantification. *Computat Geosci*, 17:689–703, 2013. doi:10.1007/s10596-013-9351-5.

Evensen, G. Sampling strategies and square root analysis schemes for the EnKF. *Ocean Dynamics*, 54:539–560, 2004. doi:10.1007/s10236-004-0099-2.

Evensen, G. Accounting for model errors in iterative ensemble smoothers. *Computat Geosci*, 23(4):pp. 761–775, 2019. doi:10.1007/s10596-019-9819-z.

Evensen, G. Formulating the history matching problem with consistent error statistics. *Computat Geosci*, 2021. doi:10.1007/s10596-021-10032-7.

Evensen, G., P. Raanes, A. Stordal, and J. Hove. Efficient implementation of an iterative ensemble smoother for data assimilation and reservoir history matching. *Frontiers in Applied Mathematics and Statistics*, 5:47, 2019. doi:10.3389/fams.2019.00047.

Hunt, B. R., E. J. Kostelich, and I. Szunyogh. Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D*, 230:112–126, 2007. doi:10.1016/j.physd.2006.11.008.

Neto, G. M. S., R. V. Soares, G. Evensen, A. Davolioa, and D. J. Schiozer. Subspace ensemble randomized maximum likelihood with local analysis for time-lapse-seismic-data assimilation. *SPE Journal*, SPE-205029-PA:21, 2020. doi:10.2118/205029-PA.

Raanes, P. N., A. S. Stordal, and G. Evensen. Revising the stochastic iterative ensemble smoother. *Nonlin. Processes Geophys*, 26:325–338, 2019. doi:10.5194/npg-2019-10.