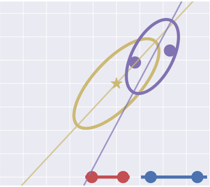
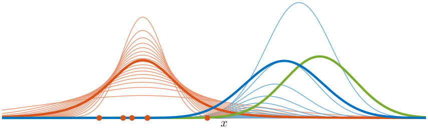
EnKF – FAQ

(Ensemble Kalman filter – Frequently asked questions)





Patrick N. Raanes, Geir Evensen, Andreas S. Stordal

Marc Bocquet, Alberto Carrassi

EnKF workshop, Voss, June 4, 2019

Received: 25 January 2018 | Revised: 03 July 2018 | Accepted: 18 July 2018
DOI: 10.1002/qj.3386

RESEARCH ARTICLE

Quarterly Journal of the Royal Meteorological Society RMetS

Adaptive covariance inflation in the ensemble Kalman filter by Gaussian scale mixtures

Patrick N. Raanes[1] | Marc Bocquet[2] | Alberto Carrassi[1]

[1]Nansen Environmental and Remote Sensing Center, Bergen, Norway
[2]CEREA, Joint laboratory of École des Ponts ParisTech and EDF R&D, Université Paris-Est, Champs-sur-Marne, France

Correspondence
Patrick N. Raanes, NERSC, Thormøhlensgate 47, Bergen N-5006, Norway.
Email: patrick.n.raanes@gmail.com

This paper studies multiplicative inflation: the complementary scaling of the state covariance in the ensemble Kalman filter (EnKF). Firstly, error sources in the EnKF are catalogued and discussed in relation to inflation; nonlinearity is given particular attention as a source of sampling error. In response, the "finite-size" refinement known as the EnKF-*N* is re-derived via a Gaussian scale mixture, again demonstrating how it yields adaptive inflation. Existing methods for adaptive inflation estimation are reviewed, and several insights are gained from a comparative analysis. One such adaptive inflation method is selected to complement the EnKF-*N* to make a hybrid that is suitable for contexts where model error is present and imperfectly parametrized. Benchmarks are obtained from experiments with the two-scale Lorenz model and its slow-scale truncation. The proposed hybrid EnKF-*N* method of adaptive inflation is found to yield systematic accuracy improvements in comparison with the existing methods, albeit to a moderate degree.

KEYWORDS
adaptive filtering, Bayesian inference, covariance inflation, data assimilation, [illegible], scale mixture

Funding information
EmblAUS project of the Nordic countries funding agency NordForsk, by DIGIRES sponsored by PETROMAKS [illegible] Norway and industry partners, and project REDDA of the Norwegian Research Council.

Also answered these questions about the EnKF:

- Why do we use $(N-1)$ in $\frac{1}{N-1}\sum_n (x_n - \bar{x})^2$?
- About nonlinearity:
  - Why does it create sampling error?
  - Why does it cause divergence?

1 | INTRODUCTION

Revising the stochastic iterative ensemble smoother

Patrick N. Raanes*[1,2], Geir Evensen[1,2], and Andreas S. Stordal[1]

[1]NORCE, Pb. 22 Nygårdstangen, 5838 Bergen, Norway
[2]NERSC, Thormøhlens gate 47, 5006 Bergen, Norway

February 4, 2019

Abstract

Ensemble randomized maximum likelihood (EnRML) is an iterative (stochastic) ensemble smoother, used for large and nonlinear inverse problems, such as history matching and data assimilation. Its current formulation is overly complicated and has issues with computational costs, noise, and covariance localization, even causing some practitioners to omit crucial prior information. This paper resolves these difficulties and streamlines the algorithm, without changing its output. These simplifications are achieved through the careful treatment of the linearizations and subspaces. For example, it is shown (a) how ensemble linearizations relate to average sensitivity, and (b) that the ensemble does not loose rank during updates. The paper also draws significantly on the theory of the (deterministic) iterative ensemble Kalman smoother (IEnKS). Comparative benchmarks are obtained with the Lorenz-96 model with these two smoothers and the ensemble smoother using multiple data assimilation (ES-MDA).

Also answered these questions about the EnKF:

- Why do we prefer the Kalman gain "form"?
- About ensemble linearizations:
  - What are they?
  - Why is this rarely discussed?
  - How does it relate to analytic derivatives?

Ensemble linearizations

Traditional EnKF presentation

Recall the KF gain:

$$\mathbf{K} = \mathbf{C}_{\boldsymbol{x}}\mathbf{H}^{\mathsf{T}}\Big(\mathbf{H}\mathbf{C}_{\boldsymbol{x}}\mathbf{H}^{\mathsf{T}} + \mathbf{R}\Big)^{-1} . \tag{1}$$

1st idea: substitute $\mathbf{C}_{\boldsymbol{x}} \leftarrow \bar{\mathbf{C}}_{\boldsymbol{x}} = \frac{1}{N-1}\mathbf{X}\mathbf{X}^{\mathsf{T}}$

$$\implies \bar{\mathbf{K}} = \bar{\mathbf{C}}_{\boldsymbol{x}}\mathbf{H}^{\mathsf{T}}\Big(\mathbf{H}\bar{\mathbf{C}}_{\boldsymbol{x}}\mathbf{H}^{\mathsf{T}} + \mathbf{R}\Big)^{-1} \tag{2}$$

$$= \mathbf{X}\mathbf{Y}^{\mathsf{T}}\Big(\mathbf{Y}\mathbf{Y} + (N-1)\mathbf{R}\Big)^{-1} \tag{3}$$

$$\text{with} \quad \mathbf{Y} = \mathbf{H}\mathbf{X} \tag{4}$$

$$= \mathcal{H}(\mathbf{X}) \tag{5}$$

$$= \mathcal{H}(\mathbf{E}) - \text{mean} . \tag{6}$$

2nd idea: use eqn. (6) also in nonlinear case (when $\nexists \mathbf{H}$).

What is the ensemble's linearization?

Recall $\bar{\mathbf{C}}_{\boldsymbol{x}} = \frac{1}{N-1}\mathbf{X}\mathbf{X}^{\mathsf{T}}$ and suppose $\mathcal{H}$ is nonlinear.

Question: Is there a matrix $\bar{\mathbf{H}}$ such that $\begin{cases} \frac{1}{N-1}\mathbf{X}\mathbf{Y}^{\mathsf{T}} = \bar{\mathbf{C}}_{\boldsymbol{x}}\bar{\mathbf{H}}^{\mathsf{T}} \\ \frac{1}{N-1}\mathbf{Y}\mathbf{Y}^{\mathsf{T}} = \bar{\mathbf{H}}\bar{\mathbf{C}}_{\boldsymbol{x}}\bar{\mathbf{H}}^{\mathsf{T}} \end{cases}$?

Answer: Yes (mostly): $\bar{\mathbf{H}} = \mathbf{Y}\mathbf{X}^{+}$.

Follow up questions:

- How come this is rarely discussed?
- Why $\mathbf{Y}\mathbf{X}^{+}$?
- Does it relate to the analytic derivative ($\mathcal{H}'$) ?

Does $\bar{\mathbf{H}} = \mathbf{Y}\mathbf{X}^{+}$ relate to $\mathcal{H}'$?

Theorem: $\lim_{N\to\infty} \bar{\mathbf{H}} = \mathbb{E}[\mathcal{H}'(\boldsymbol{x})]$

$$= \lim_{N\to\infty} \mathbf{Y}\mathbf{X}^{+} \qquad = \mathbf{C}_{\boldsymbol{yx}}\mathbf{C}_{\boldsymbol{x}}^{-1}$$

$$= \lim_{N\to\infty} \mathbf{Y}\mathbf{X}^{\mathsf{T}}(\mathbf{X}\mathbf{X}^{\mathsf{T}})^{-1} \qquad \text{(by Stein/IBP)}$$

$$= \lim_{N\to\infty} \bar{\mathbf{C}}_{\boldsymbol{yx}}\bar{\mathbf{C}}_{\boldsymbol{x}}^{-1}$$

$$= \mathbf{C}_{\boldsymbol{yx}}\mathbf{C}_{\boldsymbol{x}}^{-1} \quad \text{(a.s., by Slutsky, sub. to reg.)}$$

I.e. $\bar{\mathbf{H}}$ may (indeed) be called the "average" derivative.

Assumptions:

- Ensemble (behind $\bar{\mathbf{H}}$) and $\boldsymbol{x}$ share the same distribution.
- This is Gaussian.

How come $\bar{\mathbf{H}} = \mathbf{Y}\mathbf{X}^{+}$ is rarely discussed?

Substitute $\mathbf{H} \leftarrow \bar{\mathbf{H}}$ in $\bar{\mathbf{K}}$:

$$\bar{\mathbf{K}} = \bar{\mathbf{C}}_{\boldsymbol{x}}\bar{\mathbf{H}}^{\mathsf{T}}(\bar{\mathbf{H}}\bar{\mathbf{C}}_{\boldsymbol{x}}\bar{\mathbf{H}}^{\mathsf{T}} + \mathbf{R})^{-1} \tag{7}$$

$$= \mathbf{X}\mathbf{Y}^{\mathsf{T}}(\mathbf{Y}\boldsymbol{\Pi}_{\mathbf{X}^{\mathsf{T}}}\mathbf{Y}^{\mathsf{T}} + (N-1)\mathbf{R})^{-1}\,, \tag{8}$$

where $\boldsymbol{\Pi}_{\mathbf{X}^{\mathsf{T}}} = \mathbf{X}^{+}\mathbf{X}$, which is scary... But $\boldsymbol{\Pi}_{\mathbf{X}^{\mathsf{T}}}$

- is just a projection;
- vanishes if $\mathcal{H}$ is linear, or $(N-1) \leq M$;
- is present for any/all linearization of $\mathcal{H}$;

Why $\bar{\mathbf{H}} = \mathbf{Y}\mathbf{X}^{+}$?

$\bar{\mathbf{H}}$ is:

- Linear least-squares (LLS) estimate of $\mathcal{H}$ given $\mathbf{Y}$ and $\mathbf{X}$.
- BLUE ?
- MVUE ?

$\bar{\mathbf{H}}$ is LLS because $\bar{\mathbf{K}}$ is LLS, and the chain rule applies.

Why the “gain form”?

Not equivalent when $(N-1) < M$:

$$\bar{\mathbf{P}} = [\mathbf{I} - \bar{\mathbf{K}}\mathbf{H}]\bar{\mathbf{B}} \tag{9}$$

$$\bar{\mathbf{P}} = (\bar{\mathbf{B}}^{+} + \mathbf{H}^{\mathsf{T}}\mathbf{R}^{-1}\mathbf{H})^{-1} \tag{10}$$

Why is the Kalman gain form (9) better?

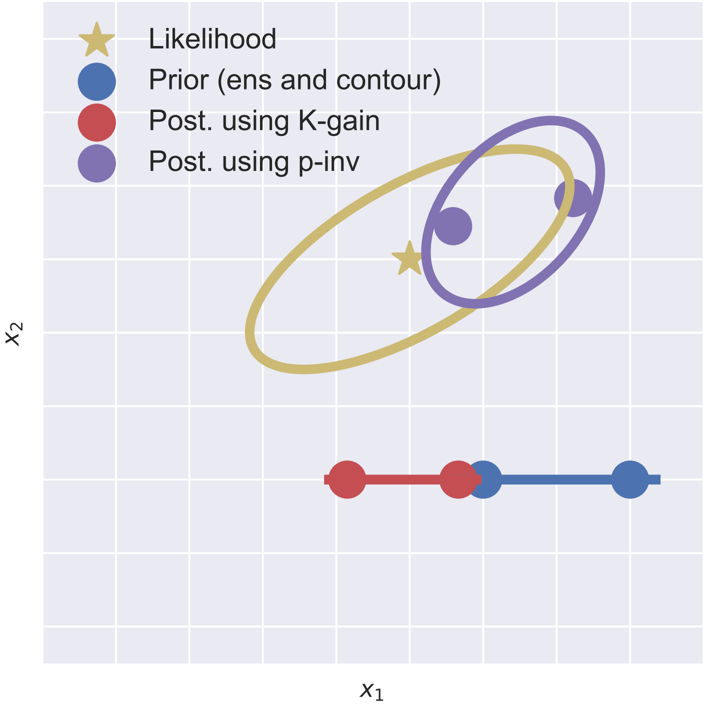
Note that eqn. (10) follows from

$$\text{prior} \propto \exp[-\tfrac{1}{2}(\boldsymbol{x} - \bar{\boldsymbol{x}})^{\mathsf{T}}\, \bar{\mathbf{B}}^{+}(\boldsymbol{x} - \bar{\boldsymbol{x}})]\,, \tag{11}$$

which is "flat" in the directions outside of $\mathrm{col}(\bar{\mathbf{B}})$.

$\Longrightarrow$ eqn. (10) yields "opposite" of the correct update.

Note: further complications in case $\bar{\mathbf{P}}$ not defined in eqn. (10).

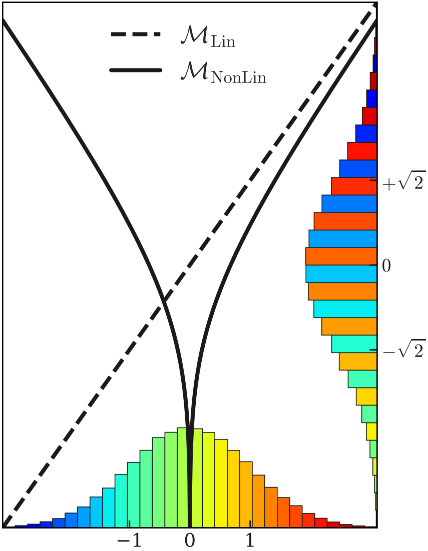

Likelihood
Prior (ens and contour)
Post. using K-gain
Post. using p-inv
x_2
x_1

Nonlinearity,

sampling error, and divergence.

Aim: study sampling error, due to nonlinearity, without worrying about non-Gaussianity.

$$\mathcal{M}_{\mathsf{Lin}}(x) = \sqrt{2}x\,,$$
$$\mathcal{M}_{\mathsf{NonLin}}(x) = \sqrt{2}F_{\mathcal{N}}^{-1}(F_{\chi}(x^2))$$



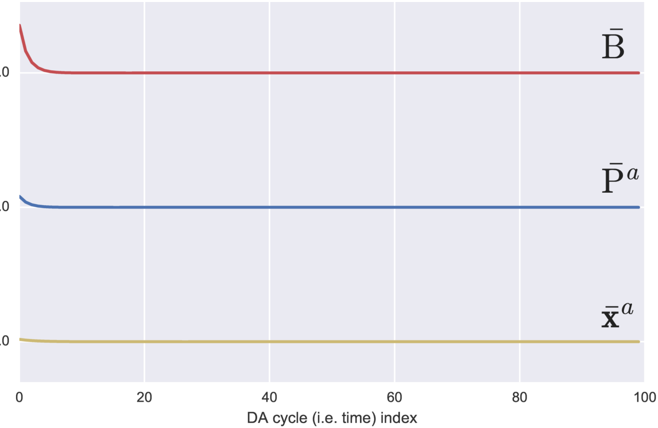
Motivational problem

$$\text{prior} = \mathcal{N}(x|0, \underline{2}),$$

$$\text{likelihood} = \mathcal{N}(0|x, 2),$$

$$\implies \text{posterior} = \mathcal{N}(x|\underline{0}, \underline{1}).$$

$$\text{dyn. model} = \mathcal{M}_{\text{Lin}}(x) = \sqrt{2}x \quad \text{dyn. model} = \cancel{\mathcal{M}_{\text{Lin}}(x)} = \cancel{\sqrt{2}x} \quad \text{dyn. model}$$



Sampling error from nonlinearity – why?

Consider the error
in the m-th sample moment
of the forecast (f) ensemble,
propagated by a degree-d model.
It can be shown that

$$\mathsf{Error}^{\mathsf{f}}_m = \sum_{i=1}^{md} C_{m,i} \mathsf{Error}^a_i \, , \tag{12}$$

i.e. the moments get coupled, which defeats moment-matching.

Riccati recursion

Assume constant, linear dynamics $(\mathbf{M})$, $\mathbf{Q} = 0$, $\mathbf{H} = \mathbf{I}$, and a deterministic EnKF.

The ensemble covariance obeys:

- Forecast:
$$\bar{\mathbf{B}}_k = \mathbf{M}^2 \bar{\mathbf{P}}_{k-1} \,. \tag{13}$$

- Analysis:
$$\bar{\mathbf{P}}_k = (\mathbf{I} - \bar{\mathbf{K}}_k)\bar{\mathbf{B}}_k \tag{14}$$
$$\iff \quad \bar{\mathbf{P}}_k^{-1} = \bar{\mathbf{B}}_k^{-1} + \mathbf{R}^{-1} \,. \tag{15}$$

$\implies$ The "Riccati recursion":

$$\bar{\mathbf{P}}_k^{-1} = (\mathbf{M}^2 \bar{\mathbf{P}}_{k-1})^{-1} + \mathbf{R}^{-1} \,. \tag{16}$$

Attenuation

Stationary Riccati:

$$\bar{\mathbf{P}}_\infty^{-1} = (\mathbf{M}^2 \bar{\mathbf{P}}_\infty)^{-1} + \mathbf{R}^{-1} \tag{17}$$

$$\iff \quad \bar{\mathbf{P}}_\infty = \bar{\mathbf{K}}_\infty \mathbf{R} \,, \quad \bar{\mathbf{K}}_\infty = \begin{cases} \mathbf{I} - \mathbf{M}^{-2} & \text{if } \mathbf{M} \geq 1 \,, \\ 0 & \text{otherwise.} \end{cases} \tag{18}$$

Initial conditions (ICs) don't appear
$\implies$ ICs are "forgotten".
$\implies$ Sampling error is attenuated.

Why $(N-1)$?

Suppose we re-define the EnKF algorithm to use a different normalization factor, i.e.

$$\tilde{\mathbf{P}}_k \tilde{\mathbf{P}}_0 = \frac{\alpha}{N-1} \mathbf{X}_k \mathbf{X}_k^{\mathsf{T}} \frac{\alpha}{N-1} \mathbf{X}_0 \mathbf{X}_0^{\mathsf{T}} = \alpha \bar{\mathbf{P}}_k \alpha \bar{\mathbf{P}}_0 \,. \tag{19}$$

But,

the ensemble forecast yields $\quad \tilde{\mathbf{B}}_k = \mathbf{M}^2 \tilde{\mathbf{P}}_{k-1} \,,$ (20)

the analysis using $\tilde{\mathbf{B}}_k$ yields $\quad \tilde{\mathbf{P}}_k^{-1} = \tilde{\mathbf{B}}_k^{-1} + \mathbf{R}^{-1} \,.$ (21)

$\Longrightarrow$ The Riccati recursion: $\quad \tilde{\mathbf{P}}_k^{-1} = (\mathbf{M}^2 \tilde{\mathbf{P}}_{k-1})^{-1} + \mathbf{R}^{-1}$ (22)

Note: α does not appear.

$\Longrightarrow$ its impact is attenuated, just like ICs.

$\Longrightarrow$ $\tilde{\mathbf{P}}_k \xrightarrow[k\to\infty]{} \bar{\mathbf{P}}_k$.

$\Longrightarrow$ $\tilde{\boldsymbol{x}}_k \xrightarrow[k\to\infty]{} \bar{\boldsymbol{x}}_k$.

Filter divergence

Recall Riccati:

$$\bar{\mathbf{P}}_k = \underbrace{(\,\mathbf{I} - \bar{\mathbf{K}}_k\,)}_{\xrightarrow[k\to\infty]{}\mathbf{M}^{-2}} \; \mathbf{M}^2 \bar{\mathbf{P}}_{k-1}\,. \tag{23}$$

Now consider $\delta\bar{\mathbf{P}}_k$. Its recursion is:

$$\delta\bar{\mathbf{P}}_k \approx (\,\mathbf{I} - \bar{\mathbf{K}}_k\,)^2\,[\mathbf{M}^2 + \mathcal{M}\mathcal{M}''\,]\,\delta\bar{\mathbf{P}}_{k-1}\,, \tag{24}$$

Yielding $\delta\bar{\mathbf{P}}_k \xrightarrow[k\to\infty]{} 0$ in the linear case ($\mathcal{M}'' = 0$), as we found previously.

By contrast, no such guarantee exists when $\mathcal{M}'' \neq 0$

$\implies$ filter divergence.

Also, $\mathcal{M}''$ may grow worse with k

$\implies$ vicious circle.

Revising

EnRML

EnRML issues

Gauss-Newton version:

(Reynolds et al., 2006; Gu and Oliver, 2007; Chen and Oliver, 2012):

- Requires "model sensitivity" matrix.
- . . . which requires pseudo-inverse of the "anomalies".

$\Longrightarrow$ Levenberg-Marquardt version (Chen and Oliver, 2013):

- Modification *inside* Hessian, simplifying *likelihood increment*.
- Further complicates *prior increment*, which is sometimes dropped!

$\Longrightarrow$ New version Raanes, Evensen, Stordal, 2019:

- No explicit computation of the model sensitivity matrix
- Computing its product with the prior covariance is efficient.
- Does not require any pseudo-inversions.

Algorithm simplification

Chen and Oliver (2013):

$$
\begin{array}{l}
d^{o} \leftarrow d_{\text{obs}} + \epsilon; \qquad \texttt{* perturb observations, } \epsilon \sim N[0, C_D]; \\
\Delta m_{\text{pr}} \leftarrow C_{\text{sc}}^{-1/2}\left(m_{\text{pr}} - \overline{m_{\text{pr}}}\right)/\sqrt{N_{\text{e}} - 1}; \\
U_{m_0}^{p_{m_0}} W_{m_0}^{p_{m_0}} (V_{m_0}^{p_{m_0}})^{\mathrm{T}} \leftarrow \Delta m_{\text{pr}}; \qquad \texttt{* truncated SVD;} \\
A_m \leftarrow U_{m_0}^{p_{m_0}} (W_{m_0}^{p_{m_0}})^{-1}; \\
m_0 \leftarrow m_{\text{pr}}; \quad d_0 \leftarrow g(m_0); \\
S_0 \leftarrow (d_0 - d^{o})^{\mathrm{T}} C_D^{-1} (d_0 - d^{o}); \\
\textbf{while } \ell < \ell_{\max} \textbf{ do} \\
\quad \Delta m \leftarrow C_{\text{sc}}^{-1/2}\left(m_{\ell-1} - \overline{m_{\ell-1}}\right)/\sqrt{N_{\text{e}} - 1}; \\
\quad \Delta d \leftarrow C_D^{-1/2}\left(d_{\ell-1} - \overline{d_{\ell-1}}\right)/\sqrt{N_{\text{e}} - 1}; \\
\quad U_d^{p_d} W_d^{p_d} (V_d^{p_d})^{\mathrm{T}} \leftarrow \Delta d; \qquad \texttt{* truncated SVD;} \\
\quad X_1 \leftarrow (U_d^{p_d})^{\mathrm{T}} C_D^{-1/2} (d_{\ell-1} - d^{o}); \\
\quad X_2 \leftarrow \left(\Box I_{p_d} + (W_d^{p_d})^2\right)^{-1} X_1; \\
\quad X_3 \leftarrow V_d^{p_d} W_d^{p_d} X_2; \\
\quad \delta m_1 \leftarrow -C_{\text{sc}}^{1/2} \Delta m X_3; \\
\quad X_4 \leftarrow A_m^{\mathrm{T}} C_{\text{sc}}^{-1/2} (m - m_{\text{pr}}); \\
\quad X_5 \leftarrow A_m X_4; \\
\quad X_6 \leftarrow \Delta m^{\mathrm{T}} X_5; \\
\quad X_7 \leftarrow V_d^{p_d} \left(\ I_{p_d} + (W_d^{p_d})^2\right)^{-1} (V_d^{p_d})^{\mathrm{T}} X_6; \\
\quad \delta m_2 \leftarrow -C_{\text{sc}}^{1/2} \Delta m_k X_7; \\
\quad m_\ell \leftarrow m_{\ell-1} + \delta m_1 + \delta m_2; \quad d_\ell \leftarrow g(m_\ell);
\end{array}
$$

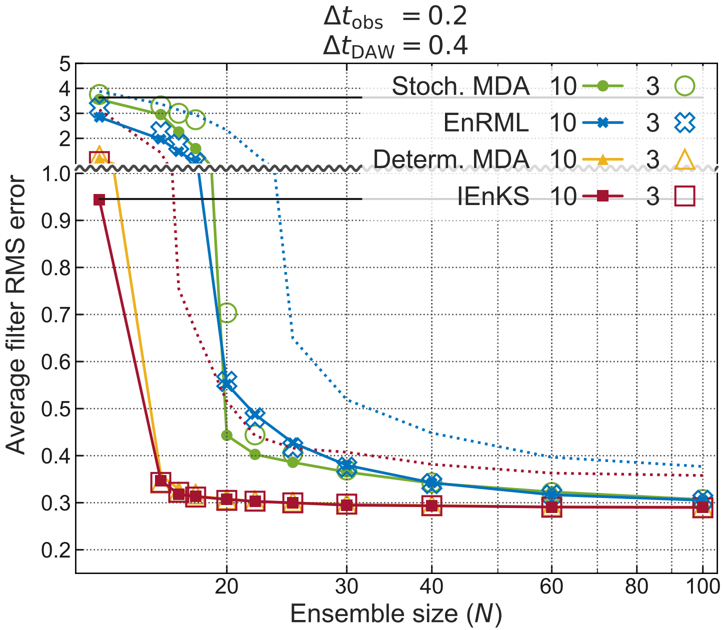
Raanes, Evensen, Stordal (2019):

$$
\begin{array}{l}
\textbf{Require: } \text{prior ens. } \mathbf{E}\text{, obs. perturb's } \mathbf{D} \\
\bar{\boldsymbol{x}} = \mathbf{E}\mathbb{1}/N \\
\mathbf{X} = \mathbf{E} - \bar{\boldsymbol{x}}\mathbb{1}^{\mathsf{T}} \\
\mathbf{W} = \mathbf{I}_N \\
\textbf{repeat:} \\
\quad \text{Run model (on each col.) to get } \mathcal{M}(\mathbf{E}) \\
\quad \mathbf{Y} = \mathcal{M}(\mathbf{E})\,(\mathbf{W}\boldsymbol{\Pi}_{\mathbb{1}}^{\perp})^{+} \\
\quad \nabla J_{\mathbf{w}}^{\text{lklhd}} = \mathbf{Y}^{\mathsf{T}}\mathbf{C}_{\mathbf{D}}^{-1}[\boldsymbol{y}\mathbb{1}^{\mathsf{T}} + \mathbf{D} - \mathcal{M}(\mathbf{E})] \\
\quad \nabla J_{\mathbf{w}}^{\text{prior}} = (N{-}1)[\mathbf{I}_N - \mathbf{W}] \\
\quad \bar{\mathbf{C}}_{\mathbf{w}} = \left(\mathbf{Y}^{\mathsf{T}}\mathbf{C}_{\mathbf{D}}^{-1}\mathbf{Y} + (N{-}1)\mathbf{I}_N\right)^{-1} \\
\quad \mathbf{W} = \mathbf{W} + \bar{\mathbf{C}}_{\mathbf{w}}[\nabla J_{\mathbf{w}}^{\text{prior}} + \nabla J_{\mathbf{w}}^{\text{lklhd}}] \\
\quad \mathbf{E} = \bar{\boldsymbol{x}}\mathbb{1}^{\mathsf{T}} + \mathbf{X}\mathbf{W}
\end{array}
$$

Available from github.com/nansencenter/DAPPER

```
if MDA: # View update as annealing (progressive assimilation).
    Cow1 = Cow1 @ T # apply previous update
    dw = dy @ Y.T @ Cow1
    if 'PertObs' in upd_a:          #== "ES-MDA". By Emerick/Reynolds.
      D     = mean0(randn(Y.shape)) * sqrt(nIter)
      T    -= (Y + D) @ Y.T @ Cow1
    elif 'Sqrt' in upd_a:           #== "ETKF-ish". By Raanes.
      T     = Cowp(0.5) * sqrt(za) @ T
    elif 'Order1' in upd_a:         #== "DEnKF-ish". By Emerick.
      T    -= 0.5 * Y @ Y.T @ Cow1
    # Tinv = eye(N) [as initialized] coz MDA does not de-condition.

else: # View update as Gauss-Newton optimzt. of log-posterior.
    grad  = Y0@dy - w*za            # Cost function gradient
    dw    = grad@Cow1               # Gauss-Newton step
    if 'Sqrt' in upd_a:             # =="ETKF-ish". By Bocquet/Sakov.
      T     = Cowp(0.5) * sqrt(N1)  # Sqrt-transforms
      Tinv  = Cowp(-.5) / sqrt(N1)  # Saves time [vs tinv(T)] when Nx<N
    elif 'PertObs' in upd_a:        # =="EnRML". By Oliver/Chen/Raanes/Evensen/Stordal.
      D     = mean0(randn(Y.shape)) if iteration==0 else D
      gradT = -(Y+D)@Y0.T + N1*(eye(N) - T)
      T     = T + gradT@Cow1
      # Tinv= tinv(T, threshold=N1) # unstable
      Tinv  = inv(T+1)              # the +1 is for stability.
    elif 'Order1' in upd_a:         #== "DEnKF-ish". By Raanes.
      # Included for completeness; does not make much sense.
      gradT = -0.5*Y@Y0.T + N1*(eye(N) - T)
      T     = T + gradT@Cow1
      Tinv  = tinv(T, threshold=N1)
```



$\Delta t_{obs} = 0.2$
$\Delta t_{DAW} = 0.4$
Stoch. MDA 10 3
EnRML 10 3
Determ. MDA 10 3
IEnKS 10 3
Average filter RMS error
5
4
3
2
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
20
30
40
60
80
100
Ensemble size (N)

Summary

- In the linear case, ICs are forgotten by Riccati.
  - $\implies$ Sampling error attenuates.
  - $\implies$ The covariance normalization factor is inconsequential.
- By contrast, nonlinearity
  - undoes the attenuation, causing filter divergence.
  - creates sampling error by cascading higher-order error down through the moments.
- Gain form $>$ Precision-matrix form.
- The ensemble linearizations
  - are LLS regression estimates.
  - converge to the average, analytic sensitivity.

References

Yan Chen and Dean S. Oliver. Ensemble randomized maximum likelihood method as an iterative ensemble smoother. *Mathematical Geosciences*, 44(1):1–26, 2012.

Yan Chen and Dean S. Oliver. Levenberg–Marquardt forms of the iterative ensemble smoother for efficient history matching and uncertainty quantification. *Computational Geosciences*, 17(4):689–703, 2013.

Yaqing Gu and Dean S. Oliver. An iterative ensemble Kalman filter for multiphase fluid flow data assimilation. *SPE Journal*, 12(04):438–446, 2007.

Patrick N. Raanes, Marc Bocquet, and Alberto Carrassi. Adaptive covariance inflation in the ensemble Kalman filter by Gaussian scale mixtures. *Quarterly Journal of the Royal Meteorological Society*, 145(718):53–75, 2019a. doi: 10.1002/qj.3386.

Patrick N. Raanes, Geir Evensen, and Andreas S. Stordal. Revising the stochastic iterative ensemble smoother. *Nonlinear Processes in Geophysics*, 0(0):0–0, 2019b. doi: 10.5194/npg-2019-10. arXiv preprint arXiv:1901.06570.

A. C. Reynolds, M. Zafari, and G. Li. Iterative forms of the ensemble Kalman filter. In *10th European Conference on the Mathematics of Oil Recovery*, 2006.

Appendix

Sampling error from nonlinearity – why?

- Consider the m-th "true" and "sample" moments:

$$\mu_m = \mathbb{E}[x^m]\,, \tag{25}$$

$$\hat{\mu}_m = N^{-1} \sum_{n=1}^{N} x_n^m\,. \tag{26}$$

- Define: $\text{Error}_m = \hat{\mu}_m - \mu_m\,.$
- Define: $\mu_m^{\mathsf{f}} = \mathbb{E}\big[\big(\mathcal{M}(x)\big)^m\big]$.
- Assume degree-d Taylor-exp. of $\mathcal{M}$ is accurate. Then

$$\mu_m^{\mathsf{f}} = \sum_{i=1}^{md} C_{m,i}\mu_i\,. \tag{27}$$

- Hence, Due to coupling of moments,

$$\text{Error}_m^{\mathsf{f}} = \sum_{i=1}^{md} C_{m,i}\text{Error}_i^a\text{Error}_i\,, \tag{28}$$

which defeats moment-matching.