Data assimilation using continuous ensemble Kalman filters

Sebastian Reich in collaboration with Georg Gottwald (University of Sydney)
and Kay Bergemann (Uni Potsdam)

1. Problem statement

Consider a differential equation

$$\dot{\mathbf{x}} = f(\mathbf{x}), \qquad \mathbf{x} \in \mathbb{R}^n,$$

for which we have some (but not precise) prior knowledge about its initial state $\mathbf{x}(0)$ and for which we can collect observations

$$\mathbf{y}(t_j) \approx \mathbf{H}\mathbf{x}(t_j), \qquad \mathbf{y} \in \mathbb{R}^k, \quad \mathbf{H} \in \mathbb{R}^{k \times n},$$

$k < n$, at discrete times $t_j > 0$, $j = 1, \ldots, J$, subject to some measurement errors.

We wish to find a solution $\mathbf{x}(t)$, $t \in [0, t_J]$, that makes optimal use of the available information in terms of initial data and observations.

2. Kalman Filter

Propagation step: Between observations propagate the mean and the covariance under the differential equation, i.e. prior to each observation, we have a most likely state $\mathbf{x}_f(t_j)$ and a covariance matrix $\mathbf{P}_f(t_j)$ and we make the simplifying (often highly questionable) assumption that

$$\mathbf{x}(t_j, \mathbf{x}_0) \sim \mathsf{N}(\mathbf{x}_f(t_j), \mathbf{P}_f(t_j)).$$

Kalman analysis step: Feed in the observations $\mathbf{y}(t_j)$ to obtain an improved most likely state $\mathbf{x}_a(t_j)$ and a covariance matrix $\mathbf{P}_a(t_j)$. We continue with a propagation step using the analyzed states

$$\mathbf{x}(t_j; \mathbf{x}_0) \sim \mathsf{N}(\mathbf{x}_a(t_j), \mathbf{P}_a(t_j)).$$

3 Ensemble propagation and ensemble Kalman filter

We now consider a collection

$$\mathbf{X}(t) = [\mathbf{x}_1(t) \,|\, \mathbf{x}_2(t) \,|\, \cdots \,|\, \mathbf{x}_m(t)] \in \mathbb{R}^{n\times m}$$

of m independent solutions of the differential equation

$$\dot{\mathbf{x}} = f(\mathbf{x}).$$

From this matrix we extract the time-dependent ensemble

$$\begin{aligned}
\text{mean:} \quad & \bar{\mathbf{x}} = m^{-1} \sum_i \mathbf{x}_i \in \mathbb{R}^n, \\
\text{deviations:} \quad & \mathbf{Y} = [\mathbf{x}_1 - \bar{\mathbf{x}} \,|\, \mathbf{x}_2 - \bar{\mathbf{x}} \,|\, \cdots \,|\, \mathbf{x}_m - \bar{\mathbf{x}}] \in \mathbb{R}^{n\times m}, \\
\text{covariance:} \quad & \mathbf{P} = \frac{1}{m-1} \mathbf{Y}\mathbf{Y}^T \in \mathbb{R}^{n\times n}.
\end{aligned}$$

The ensemble Kalman filter of Evensen (1996), combines ensemble propagation with the classical Kalman analysis step, i.e., at t_j there is a discontinuous change in $\mathbf{X}(t)$, i.e.

$$\mathbf{X}_f := \mathbf{X}(t_j - \varepsilon) \quad \rightarrow \quad \mathbf{X}(t_j + \varepsilon) = \mathbf{X}_a,$$

where $\mathbf{X}_a$ is the ensemble generated from the assimilated ensemble mean and deviation matrix:

$$\mathbf{X}_a := \mathbf{x}_a \mathbf{e}^T + \mathbf{Y}_a, \qquad \mathbf{e} = [1, 1, \ldots, 1]^T \in \mathbb{R}^m.$$

How to formulate the Kalman analysis step in terms of the ensemble deviations $\mathbf{Y}$ from the mean? ==> deterministic and stochastic formulations.

Kalman square root filters rely on a presentation/approximation of a covariance matrix $\mathbf{P} \in \mathbb{R}^{n \times n}$ in terms of the $n \times m$ ensemble deviations matrix $\mathbf{Y}$, i.e.

$$\mathbf{P} = \frac{1}{m-1} \mathbf{Y}\mathbf{Y}^T.$$

The Kalman analysis becomes

$$\mathbf{Y}_a = \mathbf{Y}_f \mathbf{T}_r$$

or equivalently

$$\mathbf{Y}_a = \mathbf{T}_l \mathbf{Y}_f.$$

The matrices $\mathbf{T}_r \in \mathbb{R}^{m \times m}$ and $\mathbf{T}_l \in \mathbb{R}^{n \times n}$ are given in terms of square roots of symmetric matrices involving $\mathbf{P}$, $\mathbf{R}$, and $\mathbf{H}$.

See book by Evensen for details on the available methods and their history.

4. Continuous update formulation

The Kalman analysis step can be formulated in terms of an gradient flow ODE (Bergemann/Reich, 2009)

$$\frac{\mathrm{d}\mathbf{x}_i}{\mathrm{d}s} = -\mathbf{P}\nabla_{\mathbf{x}_i}\mathcal{V}(\mathbf{X}), \qquad s \in [0,1],$$

in the ensemble members $\mathbf{x}_i$, $i = 1, \ldots, m$, with the ensemble induced covariance matrix $\mathbf{P}$, potential

$$\mathcal{V}(\mathbf{X}) = \frac{m}{2}\left\{ S(\bar{\mathbf{x}}) + \frac{1}{m}\sum_{i=1}^{m} S(\mathbf{x}_i) \right\},$$

and observation cost function

$$S(\mathbf{x}) = \frac{1}{2}\left(\mathbf{h}(\mathbf{x}) - \mathbf{y}\right)^T \mathbf{R}^{-1}\left(\mathbf{h}(\mathbf{x}) - \mathbf{y}\right).$$

Each observation $\mathbf{y}_j = \mathbf{y}(t_j)$ gives rise to an associated potential $\mathcal{V}_j(\mathbf{X}(t_j))$. The forward operator can also be non-linear, i.e., $\mathbf{y} = \mathbf{h}(\mathbf{x})$.

The complete assimilation system can now be formulated concisely as

$$\frac{\mathrm{d}\mathbf{x}_i}{\mathrm{d}t} = f(\mathbf{x}_i) - \sum_{j=1}^{J} \delta(t - t_j)\, \mathbf{P}\nabla_{\mathbf{x}_i}\mathcal{V}_j(\mathbf{X}),$$

$i = 1, \ldots, m$, where $\delta(\cdot)$ denotes the Dirac delta function.

5. A couple of extensions

The discontinuous effect of the filtering step can be mollified by

$$\frac{\mathrm{d}\mathbf{x}_i}{\mathrm{d}t} = f(\mathbf{x}_i) - \sum_{j=1}^{J} \delta_\varepsilon(t - t_j)\, \mathbf{P}\nabla_{\mathbf{x}_i}\mathcal{V}_j(\mathbf{X}),$$

where $\delta_\varepsilon(\cdot)$ denotes an approximation such as

$$\delta_\varepsilon(t) = \frac{1}{\varepsilon}\psi(t/\varepsilon)$$

and $\psi(s)$ is, e.g., the standard hat function (B-spline). The mollified EnKF gives rise to a sophisticated form of nudging (Anthes, 1974) and might be particularly beneficial for multi-scale assimilation since it involves temporal smoothing.

An ensemble Kalman filter step with perturbed observations can be interpreted as a stochastic ODE

$$\frac{\mathrm{d}\mathbf{x}_i}{\mathrm{d}s} = -\mathbf{P}\left[\nabla_{\mathbf{x}_i} S(\mathbf{x}_i) + \nabla_{\mathbf{x}_i} h(\mathbf{x}_i)\mathbf{R}^{-1/2}\frac{\mathrm{d}\mathbf{W}(s)}{\mathrm{d}s}\right],$$

where $\mathbf{W}(s)$ is standard k-dimensional Brownian motion.

There is a continuous embedding of both approaches using

$$\begin{aligned}\frac{\mathrm{d}\mathbf{x}_i}{\mathrm{d}s} \;=\; -\frac{1}{2}\mathbf{P}\,[&(2-\alpha)\nabla_{\mathbf{x}_i} S(\mathbf{x}_i) + \alpha m \nabla_{\mathbf{x}_i} S(\overline{\mathbf{x}}) + \\ &2\max\{0,(1-\alpha)\}\nabla_{\mathbf{x}_i} h(\mathbf{x}_i)\mathbf{R}^{-1/2}\frac{\mathrm{d}\mathbf{W}(s)}{\mathrm{d}s}\Big],\end{aligned}$$

for $\alpha \in [0,1]$ ($\alpha = 0$: perturbed observations, $\alpha = 1$: square root filter). We may also consider $\alpha \in (1,2)$, which implies a form of ensemble inflation.

6. Localization

The empirical covariance matrix $\mathbf{P}$ contains spurious long-distance correlations due to under-sampling ($m \ll n$). This problem has led to the idea of localization, i.e., $\mathbf{P}_f$ in the Kalman analysis step is replaced by

$$\mathbf{P}_f \to \mathbf{C} \circ \mathbf{P}_f,$$

where $\mathbf{C}$ is a "local" (in some distance) covariance matrix and $\circ$ denotes the Schur product of two matrices $\mathbf{C}, \mathbf{A} \in \mathbb{R}^{n \times k}$, i.e. $(\mathbf{C} \circ \mathbf{A})_{i,j} = (\mathbf{C})_{i,j} \, (\mathbf{A})_{i,j}$.

While it is straightforward to apply localization to the update of the ensemble mean, localized updates of the deviation matrix $\mathbf{Y}_f \to \mathbf{Y}_a$ are the subject of ongoing research.

Based on our continuous update formulation, we can implement localization easily.

For example, we take the (linearized) ODE update equation

$$\frac{\mathrm{d}\mathbf{x}_i}{\mathrm{d}s} = -\mathbf{P}_f \nabla_{\mathbf{x}_i} \mathcal{V}(\mathbf{X})$$

with $\mathbf{P}_f$ constant and temper/localize it to

$$\frac{\mathrm{d}\mathbf{x}_i}{\mathrm{d}s} = -\left(\mathbf{C} \circ \mathbf{P}_f\right) \nabla_{\mathbf{x}_i} \mathcal{V}(\mathbf{X}).$$

This is a linear, constant coefficient ODE in the ensemble members $\mathbf{x}_i(s)$.

See Bergemann and Reich, 2009, for details.

7. Numerical results for a barotropic fluid model

We use a 1.5 layer reduced-gravity quasi-geostrophic model with double-gyre wind forcing and biharmonic friction:

$$q_t = -\psi_x - \varepsilon J(\psi, q) - A\Delta^3\psi + 2\pi \sin(2\pi y),$$

where $q = \Delta\psi - F\psi$, $J(\psi, q) = \psi_x q_y - \psi_y q_x$. See Sakov & Oke, Tellus A, 2008, for details.

The number of degrees of freedom (phase space) after spatial discretization is $n = 127 \times 127 = 16129$, the number of observables at each t_j is $k = 300$, and the size of the ensemble is $m = 25$. The variance of the observation error is 4, i.e., $\mathbf{R} = 4\mathbf{I}$. We also use

$$\mathbf{C}_{ij,i'j'} = \exp\left(-(i-i')^2/r_0^2 - (j-j')^2/r_0^2\right)$$

for grid points x_{ij} and $x_{i'j'}$; r_0 the localization radius.

Observations are obtained from a reference numerical trajectory with added noise of variance $\mathbf{R}$, i.e., we treat our numerical model as "perfect". Simulations are run over 4000/5000 time-steps with data assimilated every fourth/fifth time-step ($\Delta t_{\text{obs}} = 5$).

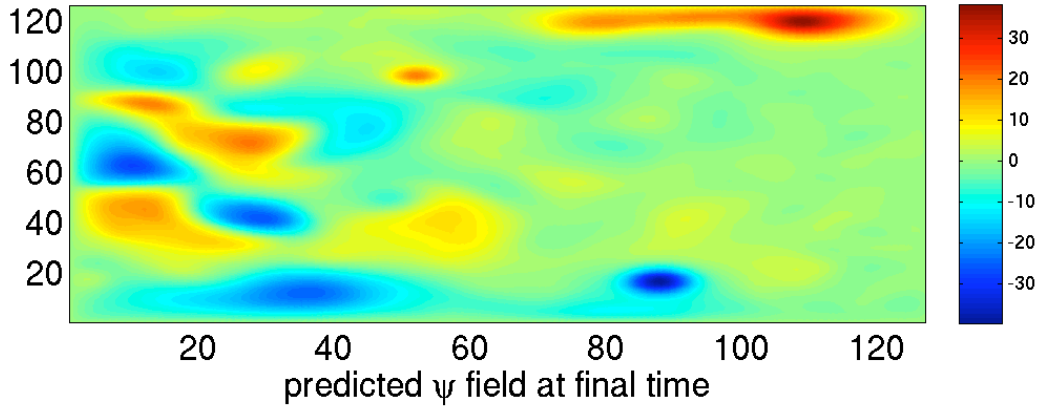
We compute the standard deviation (STD) of the difference between the "true" (unperturbed) trajectory and the ensemble mean at each observation point t_i. Roughly speaking the filter yields "skill" if the STD is less than 2 on average.

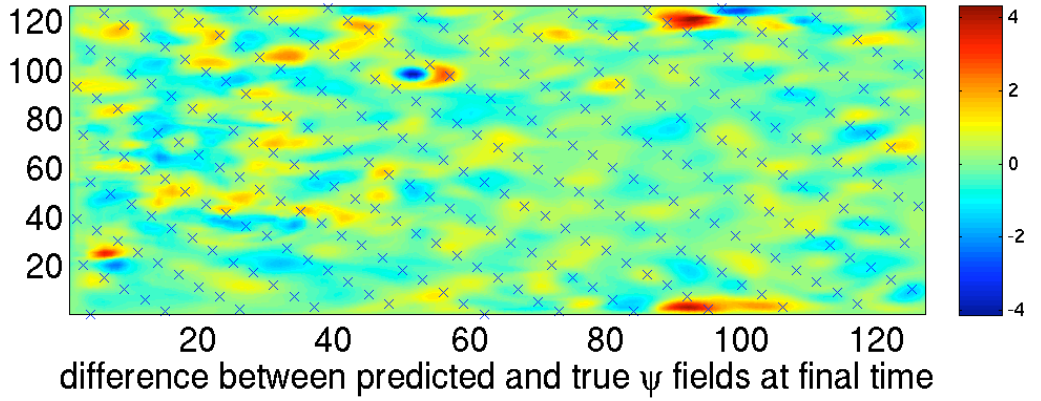
We study the behavior of the localized filter for different values of the localization radius r_0. Some experiments also use ensemble inflation

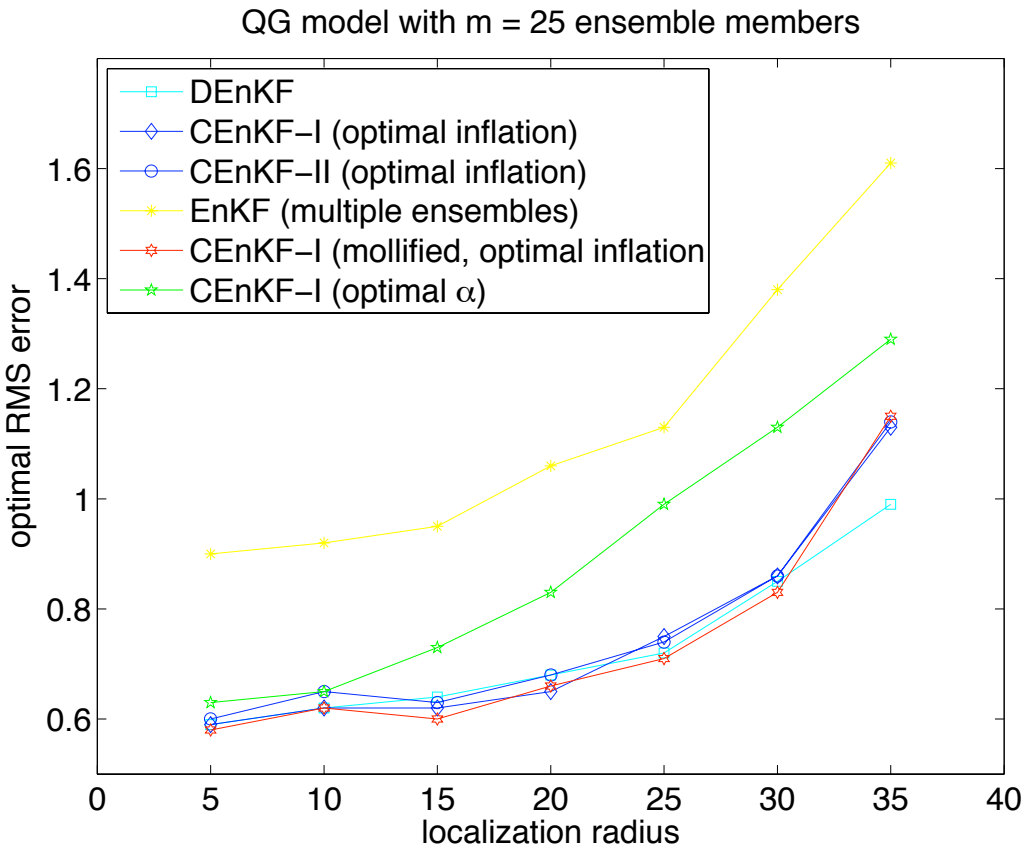
$$\mathbf{Y}(t_i) \to \delta\, \mathbf{Y}(t_i), \qquad \delta > 1$$

after each assimilation step.

A typical results at final time:







Mean RMS error for the ensemble mean update over the last 3000/4000 time steps as a functions of the localization radius r_0 (over grid point indices) optimized over the inflation factor δ and/or embedding parameter $\alpha \in (1, 2)$.

8. Numerical results from the Lorenz 96 model

The standard implementation of the Lorenz-96 model has state vector $\mathbf{x} = (x_1, \ldots, x_n)^T$ with $n = 40$ and its time evolution is given by the equation

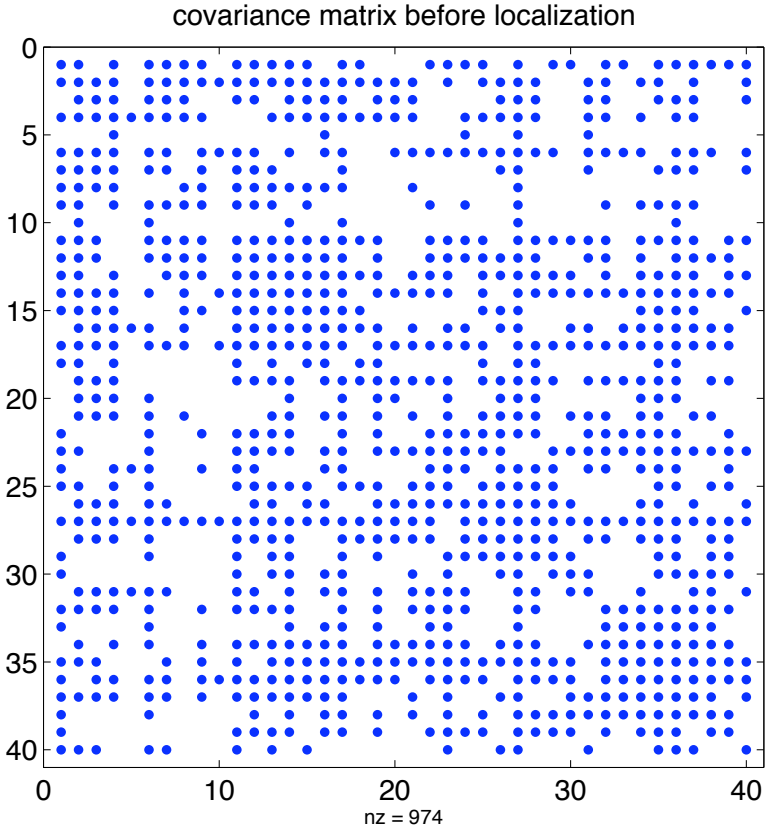
$$\dot{x}_j = (x_{j+1} - x_{j-2})x_{j-1} - x_j + 8$$

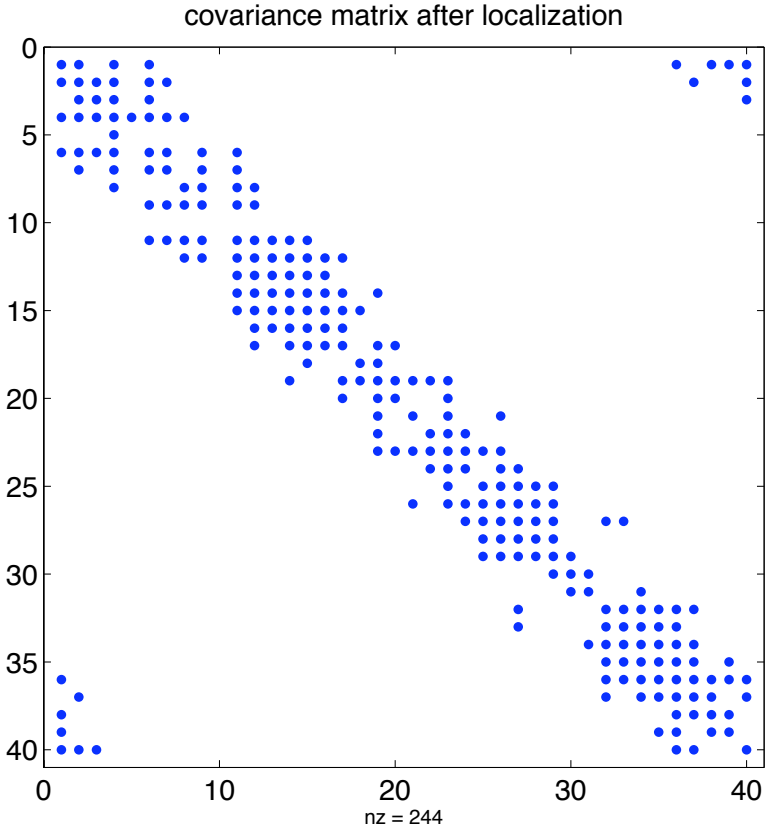
with periodic boundary conditions.

We observer every second grid point, i.e., $k = 20$ and every 0.05 units of time (≈ 6 hours real time) with measurement variance $\mathbf{R} = \mathbf{I}$.

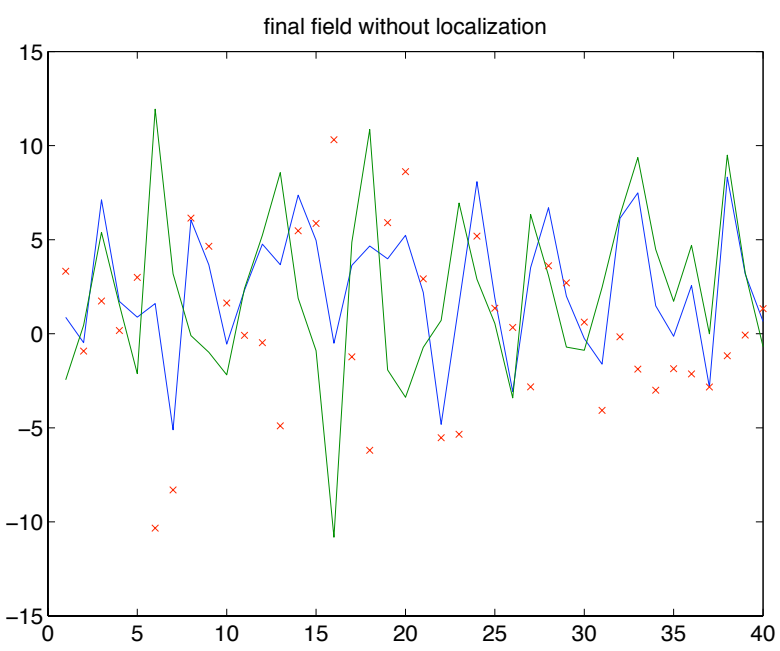
The model is chaotic with 13 positive Lyapunov exponents. One would expect that the necessary ensemble size m should be larger than the number of positive Lyapunov exponents. But this is not the case if localization is used! We show results for $m = 10$.

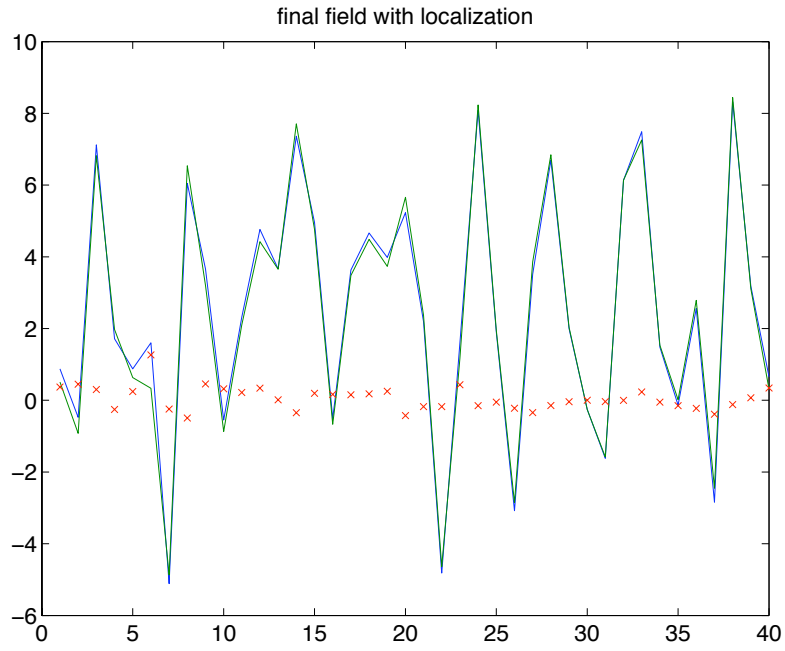
Impact of localization on 40 x 40 covariance matrix. We indicate all entries that are above a certain threshold.





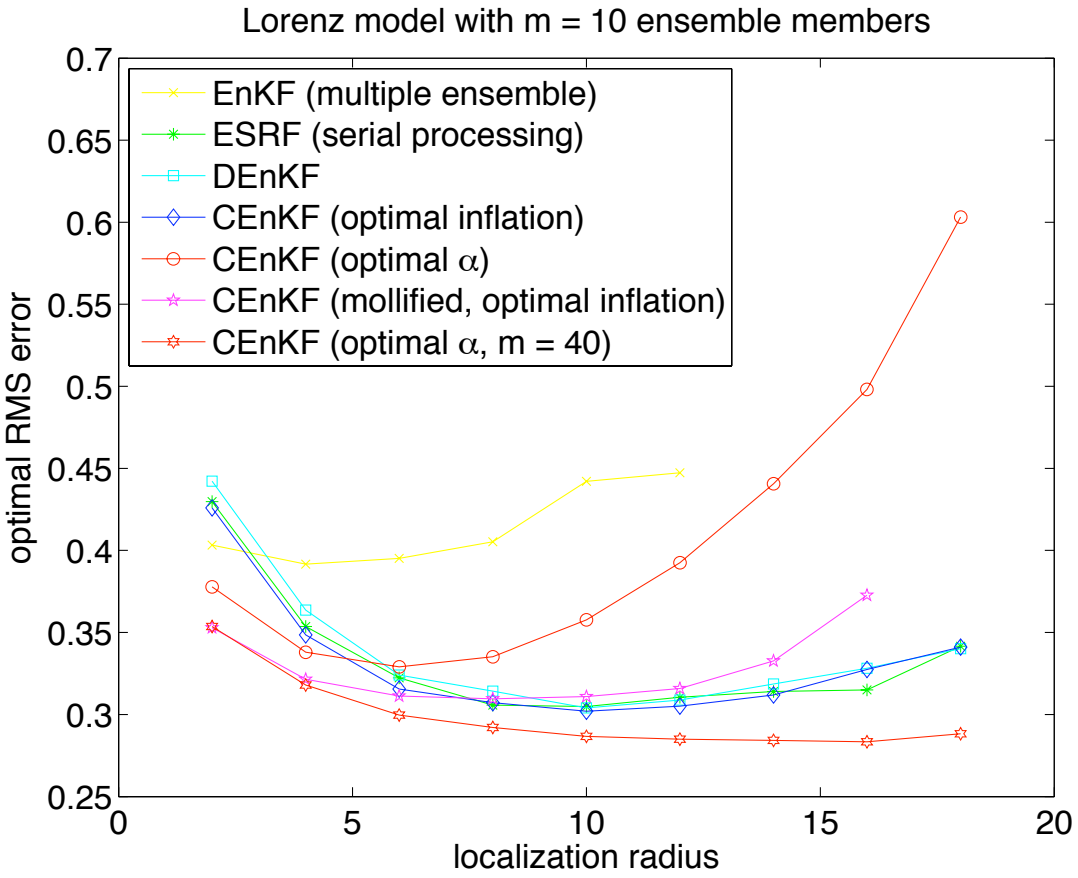
Final fields with and without localization (ensemble size $m = 10$):





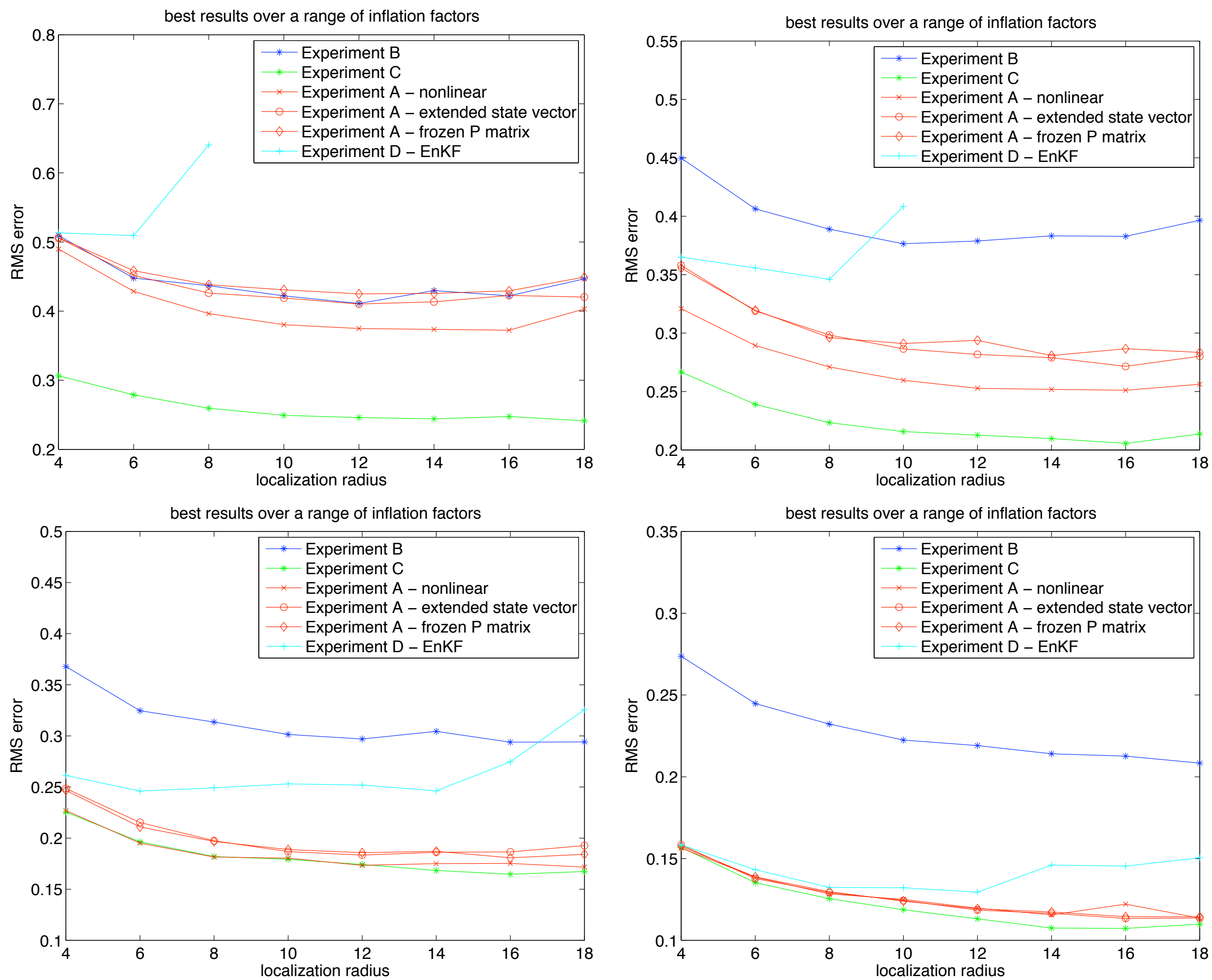
(truth green, prediction blue, difference red)

Comparison with other standard ensemble Kalman filter implementations.



We now look at slighly modified assimilation setting. We still take observations in time increments of Δt_{obs}. But instead of assimilating them directly, we assimilate only in intervals of $\Delta t_{\text{ass}} = 3\Delta t_{\text{obs}}$ by batching the closest observation together into a single Kalman filter step.

We consider $\Delta t_{\text{ass}} = 0.1, 0.075, 0.05, 0.025$. We compare the results from batched assimilation (Experiments A) to simulations with all observations used in a standard Kalman fashion (Experiment C) and to simulations using only observations at $t_{\text{obs}} = t_{\text{ass}}$ (Experiment B).


best results over a range of inflation factors
RMS error
localization radius
Experiment B
Experiment C
Experiment A – nonlinear
Experiment A – extended state vector
Experiment A – frozen P matrix
Experiment D – EnKF
0.8
0.7
0.6
0.5
0.4
0.3
0.2
4
6
8
10
12
14
16
18
best results over a range of inflation factors
RMS error
localization radius
Experiment B
Experiment C
Experiment A – nonlinear
Experiment A – extended state vector
Experiment A – frozen P matrix
Experiment D – EnKF
0.55
0.5
0.45
0.4
0.35
0.3
0.25
0.2
4
6
8
10
12
14
16
18
best results over a range of inflation factors
RMS error
localization radius
Experiment B
Experiment C
Experiment A – nonlinear
Experiment A – extended state vector
Experiment A – frozen P matrix
Experiment D – EnKF
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
4
6
8
10
12
14
16
18
best results over a range of inflation factors
RMS error
localization radius
Experiment B
Experiment C
Experiment A – nonlinear
Experiment A – extended state vector
Experiment A – frozen P matrix
Experiment D – EnKF
0.35
0.3
0.25
0.2
0.15
0.1
4
6
8
10
12
14
16
18

9. Numerical results from a slow-fast Lorenz 96 model

We consider two dynamic variables in an idealized slow-fast Lorenz 96 model

$$\begin{aligned} \dot{x}_j &= (1-\delta)(x_{j-1}x_{j+1} - x_{j-2}x_{j-1}) - x_j + 8 + \\ &\qquad \delta\left(x_{j-1}h_{j+1} - x_{j-2}h_{j-1}\right), \\ \varepsilon^2\ddot{h}_j &= \frac{1}{4}\left[h_{j+1} - 2h_j + h_{j-1}\right] - h_j + x_j \end{aligned}$$

with $\varepsilon = 0.0025$, $\delta = 0.1$, $j = 1, \ldots, 40$.

We note the "geostrophic"-type balance relation

$$h_j - \frac{1}{4}\left[h_{j+1} - 2h_j + h_{j-1}\right] = x_j$$

and the wave dynamics is undamped in h.

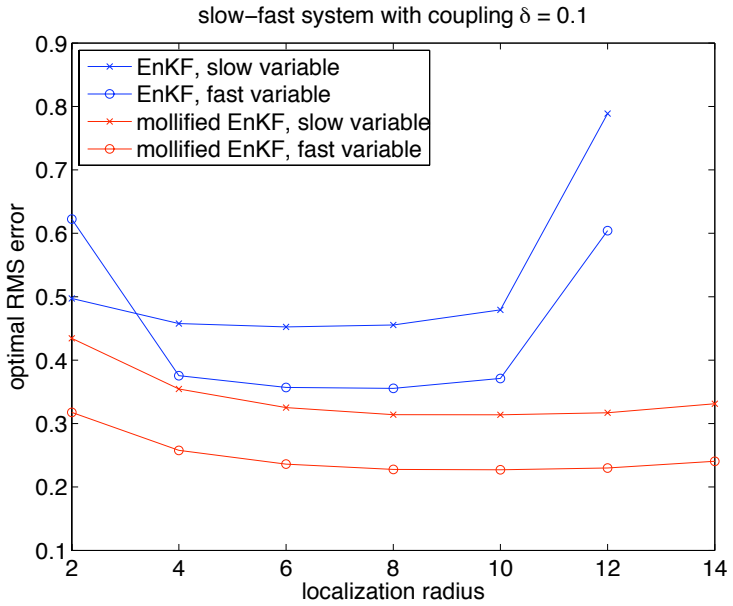
We only (partially) observe $\{x_j\}$ is it was done for the standard Lorenz-96 model in the previous epxeriments.

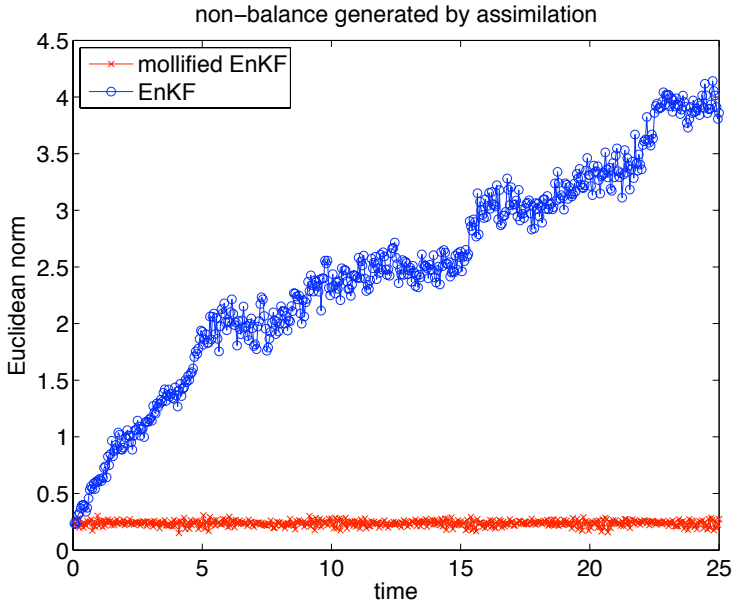
Ensemble members are initialized such that the "geostrophic" balance relation holds at initial time.

We use $\Delta t_{\text{obs}} = 0.05$ and $\Delta t = \Delta t_{\text{obs}}/40$.

We compare impulse like assimilation with a mollified implementation using hat functions spread out over $[t_j - \Delta t_{\text{obs}}/2, t_j + \Delta t_{\text{obs}}/2]$.

RMS errors in x and h separately optimized over the ensemble inflation factor as a function of localization radius and generation of unbalanced motion:





10. Conclusion

- Ensemble Kalman filters are now being implemented operationally. However, observations are batched into 6 or 12 hour cycles. In my opinion, observation should be inputed as they arrive.

- Problems arise for small ensemble sizes due to underestimation of variances and spurious covariances.

- Ensemble inflation and localization are common approaches to overcome these limitations.

- These techniques can be implemented efficiently and robustly within the ODE Kalman formulations.

- Localization and inflation are problematic for multi-scale problems and are subject of ongoing research. Mollification might provide an interesting solution *ansatz*.

- The assumption of Gaussian distributed solutions in the Kalman analysis step is questionable. EnKF with perturbed observations might be easier to extend (in the sense of mollified DA/nudging) to multi-modal behavior since no definition of a mean is required.

Reference:

[1] Evensen, 2009, Data assimilation, the ensemble Kalman filter, Springer-Verlag, 2nd edition.

[2] Bergemann, Gottwald, Reich, 2009, Ensemble propagation and continuous matrix factorization algorithms, Quarterly Journal Royal Meteorological Society, 135, 1560-1572, 2009.

[3] Bergemann, Reich, 2009, Localization techniques for ensemble transform Kalman filters, Quarterly Journal Royal Meteorological Society, in press.