Linear ensemble transform filters: A unified perspective on ensemble Kalman and particle filters

Yuan Cheng & Sebastian Reich

University of Potsdam and University of Reading

EnKF workshop 2014
Bergen, 23 June 2014

Stochastic processes (here discrete time)

$$Z^{0:N} = (Z^0, Z^1, \dots, Z^N)$$

May depend on parameters, *i.e.* $Z^{0:N}|\lambda$.

Subject them to partial observations

$$Y^{1:K} = (Y^1, Y^2, \dots, Y^K\}$$

in order to **assess** and **calibrate** models.

$K < N$ (prediction), $N = K$ (filtering), $K > N$ (smoothing).

Conditional PDFs $\pi_{Z^{0:N}}(z^{0:N}|y^{1:K}, \lambda)$ or $\pi_\Lambda(\lambda|y^{1:K})$ through **Bayesian inference** and **Monte Carlo** methods.

Stochastic processes (here discrete time)

$$Z^{0:N} = (Z^0, Z^1, \ldots, Z^N)$$

May depend on parameters, *i.e.* $Z^{0:N}|\lambda$.

Subject them to partial observations

$$Y^{1:K} = (Y^1, Y^2, \ldots, Y^K\}$$

in order to **assess** and **calibrate** models.

$K < N$ (prediction), $N = K$ (filtering), $K > N$ (smoothing).

Conditional PDFs $\pi_{Z^{0:N}}(z^{0:N}|y^{1:K}, \lambda)$ or $\pi_\Lambda(\lambda|y^{1:K})$ through **Bayesian inference** and **Monte Carlo** methods.

Stochastic processes (here discrete time)

$$Z^{0:N} = (Z^0, Z^1, \ldots, Z^N)$$

May depend on parameters, *i.e.* $Z^{0:N}|\lambda$.

Subject them to partial observations

$$Y^{1:K} = (Y^1, Y^2, \ldots, Y^K\}$$

in order to **assess** and **calibrate** models.

$K < N$ (prediction), $N = K$ (filtering), $K > N$ (smoothing).

Conditional PDFs $\pi_{Z^{0:N}}(z^{0:N}|y^{1:K}, \lambda)$ or $\pi_\Lambda(\lambda|y^{1:K})$ through **Bayesian inference** and **Monte Carlo** methods.

A typical scenario

Shadow or **track** an unknown **reference solution**

$$z_{\rm ref}^{n+1} = \Psi(z_{\rm ref}^n),$$

accessible through **partial and noisy observations**

$$y_{\rm obs}^n = h(z_{\rm ref}^n) + \xi^n, \qquad n \geq 1.$$

We only know that $z_{\rm ref}^0$ is drawn from a **random variable** Z^0.

Ensemble prediction relies on M independent realizations $z_i^0 = Z^0(\omega_i)$ (MC or quasi-MC) from the initial Z^0 and associated trajectories

$$z_i^{n+1} = \Psi(z_i^n;\lambda), \qquad n \ge 0, \qquad i = 1,\ldots,M.$$

Analysis step transforms the **forecast ensemble** $\{z_i^f = z_i^{n+1}\}$ into an **analysis ensemble** $\{z_i^a\}$ using **Bayes theorem**:

$$\pi_{Z^a}(z|y_{\rm obs}) = \frac{\pi_Y(y_{\rm obs}|z)\,\pi_{Z^f}(z)}{\pi_Y(y_{\rm obs})}.$$

Continue ensemble prediction with $\{z_i^{n+1} = z_i^a\}$.

Ensemble prediction relies on M independent realizations $z_i^0 = Z^0(\omega_i)$ (MC or quasi-MC) from the initial Z^0 and associated trajectories

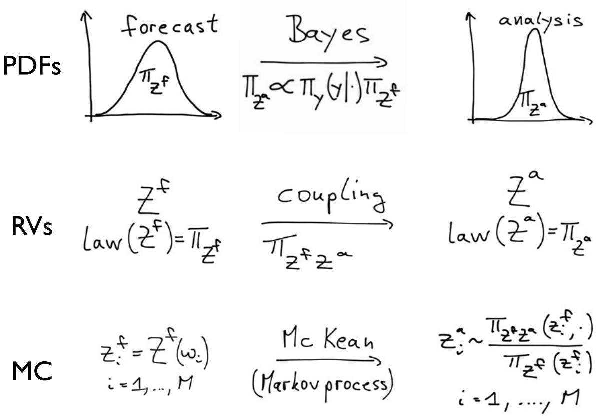
$$z_i^{n+1} = \Psi(z_i^n; \lambda), \qquad n \geq 0, \qquad i = 1, \ldots, M.$$

Analysis step transforms the **forecast ensemble** $\{z_i^f = z_i^{n+1}\}$ into an **analysis ensemble** $\{z_i^a\}$ using **Bayes theorem**:

$$\pi_{Z^a}(z|y_{\rm obs}) = \frac{\pi_Y(y_{\rm obs}|z)\,\pi_{Z^f}(z)}{\pi_Y(y_{\rm obs})}.$$

Continue ensemble prediction with $\{z_i^{n+1} = z_i^a\}$.

Summary of the **McKean approach** to the **analysis step**:



Ref.: Del Moral (2004), CJC & SR (2013), YC & SR (2014).

Parametric statistics: The Gaussian choice

(A) Fit a Gaussian $\mathrm{N}(\bar{z}^f, P^f)$ to the forecast ensemble $\{z_i^f\}$ and assume that h is linear. Then the analysis is also **Gaussian** $\mathrm{N}(\bar{z}^a, P^a)$ with

$$\bar{z}^a = \bar{z}^f - K(H\bar{z}^f - y_{\mathrm{obs}}), \qquad P^a = P^f - KHP^f.$$

Here K denotes the **Kalman gain matrix**.

Non-parametric statistics: Empirical measures

(B) Use the **empirical measure**

$$\pi^f(z) = \frac{1}{M}\sum_{i=1}^{M}\delta(z - z_i^f)$$

to define the analysis measure

$$\pi^a(z) = \sum_{i=1}^{M} w_i\delta(z - z_i^f)$$

with **importance weights**

$$w_i = \frac{\exp\left(-\frac{1}{2}(h(z_i^f) - y_{\rm obs})^T R^{-1}(h(z_i^f) - y_{\rm obs})\right)}{\sum_{j=1}^{M}\exp\left(-\frac{1}{2}(h(z_j^f) - y_{\rm obs})^T R^{-1}(h(z_j^f) - y_{\rm obs})\right)}$$

Implementation of the McKean approach then either requires **coupling** two Gaussians (approach A) or two empirical measures (approach B).

Approach A: **ensemble Kalman filters** (Evensen, 2006)

Approach B: **particle filters** (Doucet et al, 2001).

Optimal couplings in the sense of minimizing some cost function are known in both cases (CJC & SR, 2013).

We next provide a **unifying mathematical framework** in form of **linear ensemble transform filters** (LETFs) (YC & SR, 2014).

Implementation of the McKean approach then either requires **coupling** two Gaussians (approach A) or two empirical measures (approach B).

Approach A: **ensemble Kalman filters** (Evensen, 2006)

Approach B: **particle filters** (Doucet et al, 2001).

Optimal couplings in the sense of minimizing some cost function are known in both cases (CJC & SR, 2013).

We next provide a **unifying mathematical framework** in form of **linear ensemble transform filters** (LETFs) (YC & SR, 2014).

Implementation of the McKean approach then either requires **coupling** two Gaussians (approach A) or two empirical measures (approach B).

Approach A: **ensemble Kalman filters** (Evensen, 2006)

Approach B: **particle filters** (Doucet et al, 2001).

Optimal couplings in the sense of minimizing some cost function are known in both cases (CJC & SR, 2013).

We next provide a **unifying mathematical framework** in form of **linear ensemble transform filters** (LETFs) (YC & SR, 2014).

The analysis steps of an **ensemble Kalman filter** (EnKF) as well as the resampling step of a **particle filter** are of the form

$$z_j^a = \sum_{i=1}^{M} z_i^f s_{ij},$$

where $\{z_i^f\}_{i=1}^M$ is the **forecast ensemble** and $\{z_i^a\}_{i=1}^M$ is the **analysis ensemble**.

(i) The matrix $S = \{s_{ij}\} \in \mathbb{R}^{M\times M}$ depends on $y_{\rm obs}$ and the forecast ensemble.

(ii) S can be the realization of a matrix-valued RV $\mathcal{S} : \Omega \to \mathbb{R}^{M\times M}$, *i.e.* $S = \mathcal{S}(\omega)$.

The analysis steps of an **ensemble Kalman filter** (EnKF) as well as the resampling step of a **particle filter** are of the form

$$z_j^a = \sum_{i=1}^{M} z_i^f s_{ij},$$

where $\{z_i^f\}_{i=1}^M$ is the **forecast ensemble** and $\{z_i^a\}_{i=1}^M$ is the **analysis ensemble**.

(i) The matrix $S = \{s_{ij}\} \in \mathbb{R}^{M\times M}$ depends on $y_{\rm obs}$ and the forecast ensemble.

(ii) S can be the realization of a matrix-valued RV $\mathcal{S} : \Omega \to \mathbb{R}^{M\times M}$, *i.e.* $S = \mathcal{S}(\omega)$.

The **ensemble transform particle filter** (ETPF) (SR, 2013) is determined by a **coupling** $T \in \mathbb{R}^{M\times M}$ between the **discrete random variables**

$$Z^f : \Omega \to \{z_1^f, \ldots, z_M^f\} \quad \text{with} \quad \mathbb{P}[z_i^f] = 1/M$$

and

$$Z^a : \Omega \to \{z_1^f, \ldots, z_M^f\} \quad \text{with} \quad \mathbb{P}[z_i^f] = w_i,$$

respectively.

A coupling T has to satisfy $t_{ij} \geq 0$,

$$\sum_{i=1}^M t_{ij} = 1/M, \quad \sum_{j=1}^M t_{ij} = w_i.$$

The **ensemble transform particle filter** (ETPF) (SR, 2013) is determined by a **coupling** $T \in \mathbb{R}^{M\times M}$ between the **discrete random variables**

$$Z^f : \Omega \to \{z_1^f, \ldots, z_M^f\} \quad \text{with} \quad \mathbb{P}[z_i^f] = 1/M$$

and

$$Z^a : \Omega \to \{z_1^f, \ldots, z_M^f\} \quad \text{with} \quad \mathbb{P}[z_i^f] = w_i,$$

respectively.

A coupling T has to satisfy $t_{ij} \geq 0$,

$$\sum_{i=1}^M t_{ij} = 1/M, \quad \sum_{j=1}^M t_{ij} = w_i.$$

Chosing a coupling that maximizes the correlation between forecast and analysis leads to an **optimal transport problem** with cost

$$J(\{t_{ij}\}) = \sum_{i,j} \|z_i^f - z_j^f\|^2 t_{ij}.$$

Leads to the celebrated **Monge-Kantorovich problem**:

$$\pi^*_{Z^f Z^a}(z^f, z^a) = \arg \inf_{\pi_{Z^f Z^a}(z^f, z^a) \in \Pi(\pi_{Z^f}, \pi_{Z^a})} \mathbb{E}_{Z^f Z^a}\left[\|z^f - z^a\|^2\right]$$

as $M \to \infty$ (McCann, 1996, SR, 2013).

Let us denote the minimize by T^*, then the ETPF is given by

$$z_j^a = M \sum_{i=1}^{M} z_i^f t_{ij}^*.$$

Chosing a coupling that maximizes the correlation between forecast and analysis leads to an **optimal transport problem** with cost

$$J(\{t_{ij}\}) = \sum_{i,j} \|z_i^f - z_j^f\|^2 t_{ij}.$$

Leads to the celebrated **Monge-Kantorovich problem**:

$$\pi^*_{Z^f Z^a}(z^f, z^a) = \arg \inf_{\pi_{Z^f Z^a}(z^f, z^a) \in \Pi(\pi_{Z^f}, \pi_{Z^a})} \mathbb{E}_{Z^f Z^a}\left[\|z^f - z^a\|^2\right]$$

as $M \to \infty$ (McCann, 1996, SR, 2013).

Let us denote the minimize by T^*, then the ETPF is given by

$$z_j^a = M \sum_{i=1}^{M} z_i^f t_{ij}^*.$$

Chosing a coupling that maximizes the correlation between forecast and analysis leads to an **optimal transport problem** with cost

$$J(\{t_{ij}\}) = \sum_{i,j} \|z_i^f - z_j^f\|^2 t_{ij}.$$

Leads to the celebrated **Monge-Kantorovich problem**:

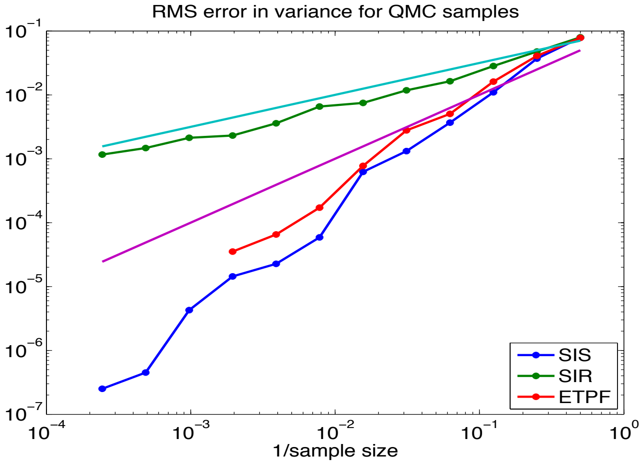
$$\pi^*_{Z^f Z^a}(z^f, z^a) = \arg \inf_{\pi_{Z^f Z^a}(z^f, z^a) \in \Pi(\pi_{Z^f}, \pi_{Z^a})} \mathbb{E}_{Z^f Z^a}\left[\|z^f - z^a\|^2\right]$$

as $M \to \infty$ (McCann, 1996, SR, 2013).

Let us denote the minimize by T^*, then the ETPF is given by

$$z_j^a = M \sum_{i=1}^{M} z_i^f t_{ij}^*.$$

Convergence rate for a single analysis step. The prior is two-dimensional uniform and **quasi-MC samples** are being used.

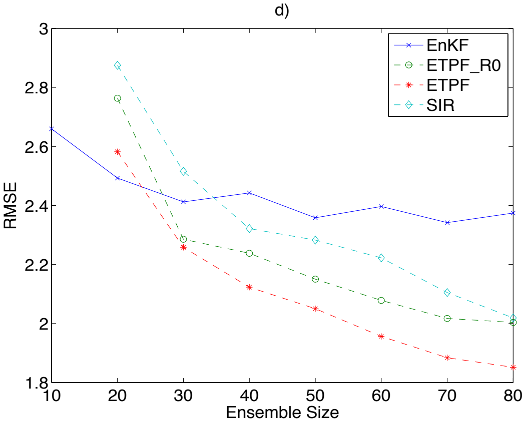


Lorenz-63 model with outputs generated every 0.12 units of time. Only the x variable is observed with measurement error variance equal to $R = 8$.

Each DA algorithm is implemented either with **ensemble inflation** or **particle rejuvenation**. A total of 20,000 assimilation steps are performed.

We compare the resulting time-averaged RMSEs:

$$\sqrt{\sum_{n=1}^{20000} \frac{1}{20000} \|\bar{z}^{a,n} - z^n_{\mathrm{ref}}\|^2}.$$


d)
RMSE
Ensemble Size
EnKF
ETPF_R0
ETPF
SIR
3
2.8
2.6
2.4
2.2
2
1.8
10
20
30
40
50
60
70
80

On the curse of dimensionality

Dynamical system

$$z^{n+1} = z^n$$

with initial PDF $\mathrm{N}(0, I)$, dimension of state space N_z, reference solution $z^n_{\mathrm{ref}} \equiv 0$.

At iteration index n we observe the nth component of the state vector, *i.e.*

$$y^n_{\mathrm{obs}} = e_n^T z^n_{\mathrm{ref}} + \xi^n, \qquad \xi \sim \mathrm{N}(0, R)$$

with $R = 0.16$, $e_n^T = (0, \ldots, 0, 1, 0, \ldots, 0)$ the nth unit vector in $\mathbb{R}^{N_z}$, and $K = N_z$.

On the curse of dimensionality

Dynamical system

$$z^{n+1} = z^n$$

with initial PDF $\mathrm{N}(0, I)$, dimension of state space N_z, reference solution $z^n_{\mathrm{ref}} \equiv 0$.

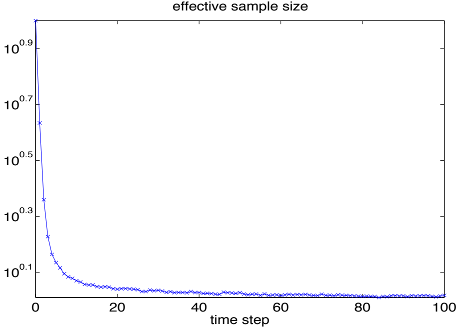
At iteration index n we observe the nth component of the state vector, *i.e.*

$$y^n_{\mathrm{obs}} = e_n^T z^n_{\mathrm{ref}} + \xi^n, \qquad \xi \sim \mathrm{N}(0, R)$$

with $R = 0.16$, $e_n^T = (0, \ldots, 0, 1, 0, \ldots, 0)$ the nth unit vector in $\mathbb{R}^{N_z}$, and $K = N_z$.

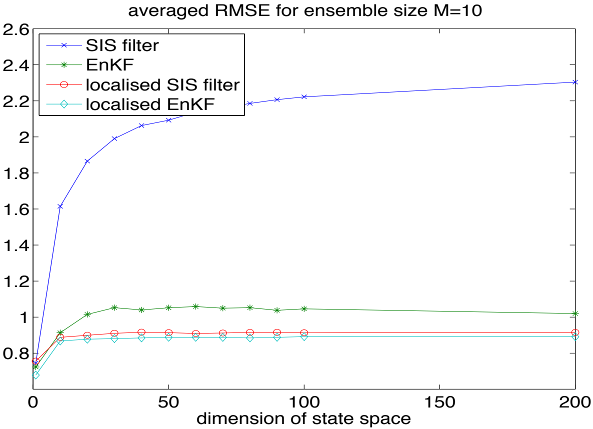
A SIS particle filter leads to the following simple update for the weights and particles

$$w_i^n \propto w_i^{n-1} e^{-\frac{1}{2R}(e_n^T z_i^0 - y_{\rm obs}^n)^2}, \qquad z_i^n = z_i^0.$$



Effective sample size

$$M_{\rm eff}^n = \frac{1}{\sum_i (w_i^n)^2}, \qquad M_{\rm off}^0 = 10.$$



RMSEs (normalised by $\sqrt{R}$) are based on either

$$\bar{z}^n = \sum_{i=1}^{M} w_i^n z_i^0 \quad \text{or} \quad \bar{z}^n = \sum_{l=1}^{N_z} \left\{ \sum_{i=1}^{M} w_i^n(l) e_l^T z_i^0 \right\} e_l .$$

Lessons to learn

1) Monte Carlo methods generate spurious correlations/dependencies between dynamic variables.

2) Correlation structures need to be explicitly built into a particle filter. This can be achieved via **localization** or appropriate model hierarchies.

3) Localization effectively increases the sample size.

Spatially extended dynamical systems

Spatially extended system with $x \in \mathbb{R}$ taking the role of the spatial variable. The forecast ensemble is now $\{z_i^f(x)\}$ and the LETF becomes

$$z_i^a(x) = \sum_{i=1}^{M} z_i^f(x)\, s_{ij}.$$

This does not work unless M is huge. Instead one uses **localization**:

$$z_i^a(x) = \sum_{i=1}^{M} z_i^f(x)\, s_{ij}(x).$$

Analysis fields need to have **sufficient spatial regularity**, *i.e.*, $z_i^f \in \mathcal{H}$ should imply $z_i^a \in \mathcal{H}$!

Spatially extended dynamical systems

Spatially extended system with $x \in \mathbb{R}$ taking the role of the spatial variable. The forecast ensemble is now $\{z_i^f(x)\}$ and the LETF becomes

$$z_i^a(x) = \sum_{i=1}^M z_i^f(x)\, s_{ij}.$$

This does not work unless M is huge. Instead one uses **localization**:

$$z_i^a(x) = \sum_{i=1}^M z_i^f(x)\, s_{ij}(x).$$

Analysis fields need to have **sufficient spatial regularity**, *i.e.*, $z_i^f \in \mathcal{H}$ should imply $z_i^a \in \mathcal{H}$!

R-localization for the ETPF:

Define a localization function with **localization radius** $r_{\rm loc} > 0$, e.g.

$$\rho(x - x') = \begin{cases} 1 - |x - x'|/r_{\rm loc} & \text{for } |x - x'| \le r_{\rm loc}, \\ 0 & \text{else.} \end{cases}$$

Depending on the spatial location $x \in \mathbb{R}$, the error variance R_k of an observation at x_k is modified to

$$\tilde{R}_k^{-1}(x) := \rho(x - \mathrm{x}_k)\, R_k^{-1}$$

and gives rise to **localized importance weights**

$$w_i(x) \propto \sum_k \exp\left(-\frac{1}{2}(z_i^f(\mathrm{x}_k) - z_{\rm obs}(\mathrm{x}_k))\tilde{R}_k^{-1}(x)(z_i^f(\mathrm{x}_k) - z_{\rm obs}(\mathrm{x}_k))\right).$$

An optimal transport problem is now solved for each computational grid point $x = x_i$ with **localized transport cost**

$$d(z^f, z^a)(x_i) := \int_{\mathbb{R}} \rho(x_i - x')\|z^f(x') - z^a(x')\|^2 \,\mathrm{d}x'.$$

R-localization for the ETPF:

Define a localization function with **localization radius** $r_{\rm loc} > 0$, e.g.

$$\rho(x - x') = \begin{cases} 1 - |x - x'|/r_{\rm loc} & \text{for } |x - x'| \le r_{\rm loc}, \\ 0 & \text{else.} \end{cases}$$

Depending on the spatial location $x \in \mathbb{R}$, the error variance R_k of an observation at x_k is modified to

$$\tilde{R}_k^{-1}(x) := \rho(x - \mathrm{x}_k)\, R_k^{-1}$$

and gives rise to **localized importance weights**

$$w_i(x) \propto \sum_k \exp\left(-\frac{1}{2}(z_i^f(\mathrm{x}_k) - z_{\rm obs}(\mathrm{x}_k))\tilde{R}_k^{-1}(x)(z_i^f(\mathrm{x}_k) - z_{\rm obs}(\mathrm{x}_k))\right).$$

An optimal transport problem is now solved for each computational grid point $x = x_i$ with **localized transport cost**

$$d(z^f, z^a)(x_i) := \int_{\mathbb{R}} \rho(x_i - x')\|z^f(x') - z^a(x')\|^2 \,\mathrm{d}x'.$$

R-localization for the ETPF:

Define a localization function with **localization radius** $r_{\rm loc} > 0$, e.g.

$$\rho(x - x') = \begin{cases} 1 - |x - x'|/r_{\rm loc} & \text{for } |x - x'| \le r_{\rm loc}, \\ 0 & \text{else.} \end{cases}$$

Depending on the spatial location $x \in \mathbb{R}$, the error variance R_k of an observation at x_k is modified to

$$\tilde{R}_k^{-1}(x) := \rho(x - \mathrm{x}_k)\, R_k^{-1}$$

and gives rise to **localized importance weights**

$$w_i(x) \propto \sum_k \exp\left(-\frac{1}{2}(z_i^f(\mathrm{x}_k) - z_{\rm obs}(\mathrm{x}_k))\tilde{R}_k^{-1}(x)(z_i^f(\mathrm{x}_k) - z_{\rm obs}(\mathrm{x}_k))\right).$$

An optimal transport problem is now solved for each computational grid point $x = x_i$ with **localized transport cost**

$$d(z^f, z^a)(x_i) := \int_{\mathbb{R}} \rho(x_i - x')\|z^f(x') - z^a(x')\|^2\, \mathrm{d}x'.$$

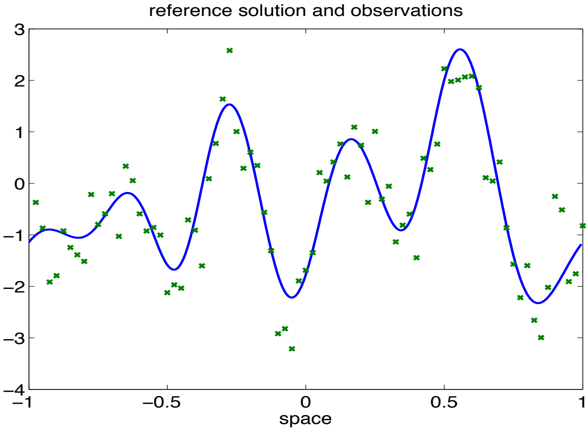
Example. **Random field** (superposition of Gaussians):

$$z(x) = \sum_i \xi_i \, \mathrm{n}(x; x_i, \sigma^2), \qquad x \in [-1, 1],$$

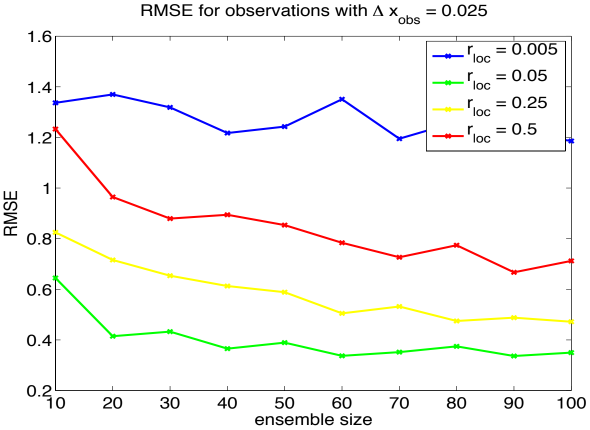
with mesh-size $\Delta x = 0.005$, grid points $x_i = i\Delta x$, random coefficients $\xi_i \sim N(0, \Delta x)$, and $\sigma^2 = 0.1$.

Observations are taken in intervals of $\Delta x_{\rm obs} = 0.025$ (every 5 grid points). The measurement errors are i.i.d. Gaussian with variance $R = 0.4$.

Typical field and observations:



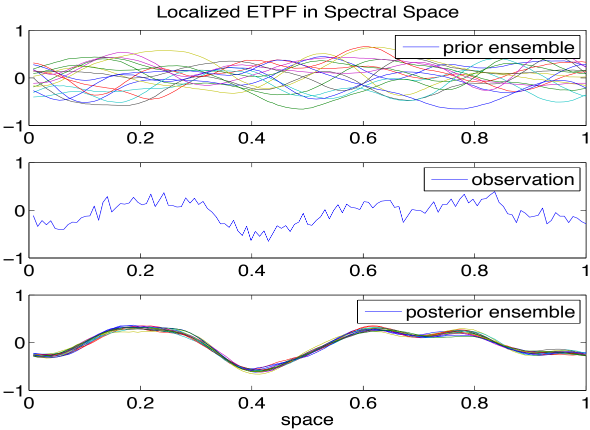
Root mean square errors (RMSE) for varying ensemble sizes and localization radii:



Note: $R^{1/2} \approx 0.63$.

Another example but now with **localization in spectral space**.

Signals are periodic and weakly correlated in spectral space, $N_z = 128$ grid points and $M = 16$ ensemble members, every grid point observed.



Example. The **Lorenz-96 ODE model**

$$\frac{\mathrm{d}u_j}{\mathrm{d}t} = -\frac{u_{j-1}u_{j+1} - u_{j-2}u_{j-1}}{3\Delta x} - u_j + F, \qquad j = 1, \ldots, 40,$$

can be thought of as the discretization of the forced-damped advection equation

$$\frac{\partial u}{\partial t} = -\frac{1}{2}\frac{\partial (u)^2}{\partial x} - u + F.$$

Every other grid point is observed in intervals of $\Delta t = 0.12$. The error variance is $R = 8$.

Example. The **Lorenz-96 ODE model**

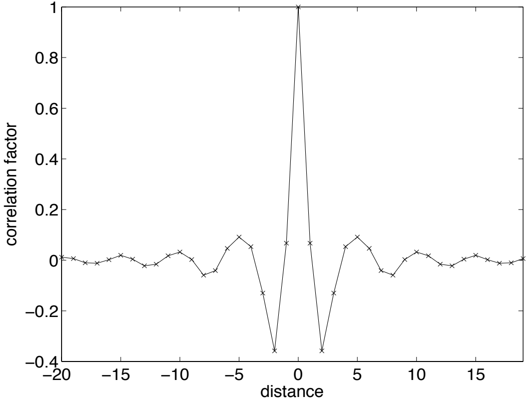
$$\frac{\mathrm{d}u_j}{\mathrm{d}t} = -\frac{u_{j-1}u_{j+1} - u_{j-2}u_{j-1}}{3\Delta x} - u_j + F, \qquad j = 1, \ldots, 40,$$

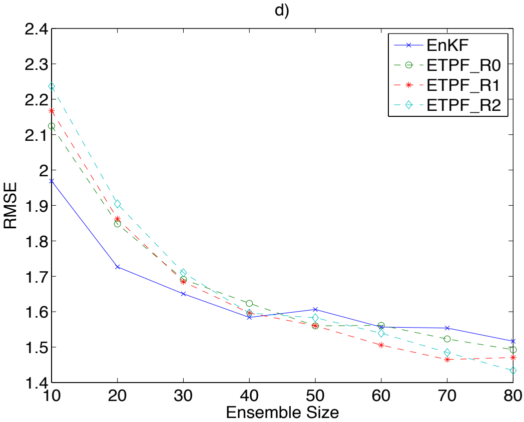
can be thought of as the discretization of the forced-damped advection equation

$$\frac{\partial u}{\partial t} = -\frac{1}{2}\frac{\partial (u)^2}{\partial x} - u + F.$$

Every other grid point is observed in intervals of $\Delta t = 0.12$. The error variance is $R = 8$.

Time averaged **spatial correlation** of solutions to the Lorenz-96 ODE:




d)
2.4
2.3
2.2
2.1
2
1.9
1.8
1.7
1.6
1.5
1.4
RMSE
10
20
30
40
50
60
70
80
Ensemble Size
EnKF
ETPF_R0
ETPF_R1
ETPF_R2

A few of topics for future work:

- replace linear transport by approximations (Earth mover's distance) such as Sinkhorn (Doucet, Cuturi, 2013), space filling Hilbert curves (Chopin, 2014), or hierarchical approaches
- time-continuous LETF formulations

$$\mathrm{d}z_j = f(z_j)\mathrm{d}t + \sum_{i=1}^{M} z_i \mathrm{d}s_{ij} + \mathrm{d}\Xi_j$$

  (Crisan et al, 2010, Sean Meyn et al, 2013, CR, 2013).
- Choice of localization function: For linear systems perfect localization can be achieved in spectral space (Harlim & Majda, 2012).
- Gaussian mixture models, ensemble smoother, adaptive methods, ...

References:

1. C.J. Cotter and S. Reich, Ensemble filter techniques for intermittent data assimilation, in: M. Cullen et al (eds.), Large scale inverse problems, Radon Ser. Comput. Appl. Math. Vol. 13, pages 91-134, 2013.
2. S. Reich, A non-parametric ensemble transform method for Bayesian inference, SIAM SISC, Vol. 35, pages A2013-A2014, 2013.
3. Y. Cheng and S. Reich, A McKean optimal transportation perspective on Feynman-Kac formulae with application to data assimilation, Frontiers in Dynamical Systems, Springer-Verlag, to appear.
4. S. Reich and C.J. Cotter, Uncertainty quantification and Bayesian data assimilation: An Introduction, Cambridge University Press, to appear in early 2015