Extended ensemble Kalman filters for high-dimensional hierarchical state-space models

Jonathan Stroud

McDonough School of Business
Georgetown University

Joint work with Matthias Katzfuss (Texas A&M) and Chris Wikle (Missouri)

Outline

1. SSMs and Existing DA Methods
2. Extended EnKFs
3. Numerical examples
4. Conclusions

Outline

The state-space model (SSM)

From a statistical perspective: DA is filtering in a SSM

SSM with additive Gaussian error in discrete time $t = 1, 2, \ldots$:

$$\begin{aligned} \mathbf{y}_t &= \mathbf{H}_t \mathbf{x}_t + \mathbf{v}_t, & \mathbf{v}_t &\sim \mathcal{N}_m(\mathbf{0}, \mathbf{R}_t), \\ \mathbf{x}_t &= \mathcal{M}_t(\mathbf{x}_{t-1}) + \mathbf{w}_t, & \mathbf{w}_t &\sim \mathcal{N}_n(\mathbf{0}, \mathbf{Q}_t), \end{aligned}$$

where

- $\mathbf{y}_t$ is $m \times 1$ observation vector.
- $\mathbf{x}_t$ is $n \times 1$ state vector.
- $\mathbf{H}_t$ is $m \times n$ observation matrix.
- $\mathcal{M}_t(\cdot)$ is evolution operator.
- $\mathbf{v}_t, \mathbf{w}_t$ are independent errors.
- initial state $\mathbf{x}_0 \sim \mathcal{N}_n(\mathbf{a}_0, \mathbf{P}_0)$
- assume parameters known (for now)

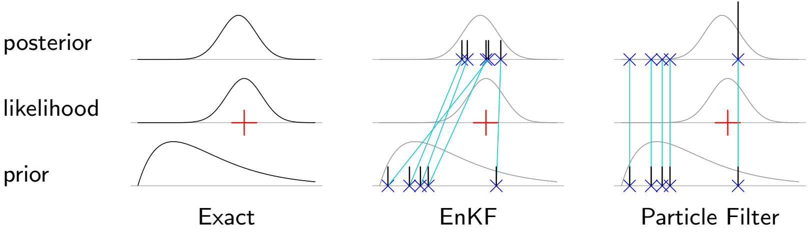
Sequential Data Assimilation Methods

Want filtering distribution $p(\mathbf{x}_t|\mathbf{y}_{1:t})$ for $t = 1, 2, \ldots, T$.

	Kalman Filter (KF) (Kalman, 1960)	**Ensemble Kalman Filter (EnKF)** (Evensen, 1994)	**Particle Filter (PF)** (Gordon et al., 1993)
Posterior at time $t-1$	$\mathcal{N}(\mathbf{a}_{t-1}, \mathbf{P}_{t-1})$	$\mathbf{x}_{t-1}^{(i)} \sim p(\mathbf{x}_{t-1}\|\mathbf{y}_{1:t-1})$	$\mathbf{x}_{t-1}^{(i)} \sim p(\mathbf{x}_{t-1}\|\mathbf{y}_{1:t-1})$
Prior at time t	$\mathcal{N}(\mathbf{a}_t^f, \mathbf{P}_t^f)$ $\mathbf{a}_t^f = \mathbf{M}_t \mathbf{a}_{t-1}$ $\mathbf{P}_t^f = \mathbf{M}_t \mathbf{P}_{t-1} \mathbf{M}_t' + \mathbf{Q}_t$	$\mathbf{x}_t^{f(i)} \sim p(\mathbf{x}_t\|\mathbf{x}_{t-1}^{(i)})$	$\mathbf{x}_t^{f(i)} \sim p(\mathbf{x}_t\|\mathbf{x}_{t-1}^{(i)})$
Posterior at time t	$\mathcal{N}(\mathbf{a}_t, \mathbf{P}_t)$ $\mathbf{a}_t = (\mathbf{I} - \mathbf{K}_t \mathbf{H}_t)\mathbf{a}_t^f + \mathbf{K}_t \mathbf{y}_t$ $\mathbf{P}_t = (\mathbf{I} - \mathbf{K}_t \mathbf{H}_t)\mathbf{P}_t^f$	$\mathbf{x}_t^{(i)} = (\mathbf{I} - \hat{\mathbf{K}}_t \mathbf{H}_t)\mathbf{x}_t^{f(i)} + \hat{\mathbf{K}}_t \mathbf{y}_t^{(i)}$ $\mathbf{y}_t^{(i)} \sim \mathcal{N}(\mathbf{y}_t, \mathbf{R}_t)$	Resample $\mathbf{x}_t^{f(i)}$ with weights $w_t^{(i)}$ where $w_t^{(i)} \propto p(\mathbf{y}_t\|\mathbf{x}_t^{f(i)})$ to obtain posterior samples $\mathbf{x}_t^{(i)}$

Illustration of the update step

Illustration of different updating schemes at a single time point



Particle weights degenerate for large n (Snyder et al., 2008; Slivinski and Snyder, 2016)

EnKF requires regularization of sample forecast covariance/gain matrix (Houtekamer and Mitchell, 1998, 2001; Ott et al., 2004; Hamill et al., 2001; Anderson and Anderson, 1999; Whitaker and Hamill, 2002; Zhang et al., 2004, ...)

Outline

Hierarchical state-space model (HSSM)

Starting with $\mathbf{x}_0 \sim \mathcal{N}_n(\mathbf{a}_0, \mathbf{P}_0)$, we assume for $t = 1, 2, \ldots$:

$$\begin{aligned}
\mathbf{z}_t | \mathbf{y}_t, \boldsymbol{\theta}_t &\sim g(\mathbf{y}_t; \boldsymbol{\theta}_t) \\
\mathbf{y}_t &= \mathbf{H}_t(\boldsymbol{\theta}_t)\,\mathbf{x}_t + \mathbf{v}_t, \qquad \mathbf{v}_t \sim \mathcal{N}_m(\mathbf{0}, \mathbf{R}_t(\boldsymbol{\theta}_t)) \\
\mathbf{x}_t &= \mathcal{M}_t(\mathbf{x}_{t-1}; \boldsymbol{\theta}_t) + \mathbf{w}_t, \quad \mathbf{w}_t \sim \mathcal{N}_n(\mathbf{0}, \mathbf{Q}_t(\boldsymbol{\theta}_t)) \\
\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1} &\sim p(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1})
\end{aligned}$$

where $\mathbf{z}_t \in \mathbb{R}^m$ are the actual measurements, $\mathbf{y}_t \in \mathbb{R}^m$ is a latent variable, $\boldsymbol{\theta}_t$ are model parameters, and we have added a transformation layer and a parameter layer to the additive Gaussian SSM (in black).

$\rightarrow$ Want filtering distribution $p(\mathbf{x}_t, \boldsymbol{\theta}_t | \mathbf{z}_{1:t})$ for $t = 1, 2, \ldots$.

Existing methods for parameter estimation

- State augmentation (Anderson, 2001). Most popular method, but does not work well if the states and parameters are weakly correlated (DelSole and Yang, 2010)
- Sequential maximum likelihood/Bayes (Dee and da Silva, 1999; Mitchell and Houtekamer, 2000; Stroud and Bengtsson, 2007; Frei and Künsch, 2012; Stroud et al., 2018) — for specific parameters
- Iterative EnKF/EnKS (Gu and Oliver, 2007; Chen and Oliver, 2012; Bocquet and Sakov, 2013) – state augmentation with optimization.
- Also much work on choosing tuning parameters (e.g., Anderson, 2007a,b, 2009; Li et al., 2009; Miyoshi, 2011, ...).

Non-Gaussian Data Assimilation

- Particle Filters (e.g., van Leeuwen, 2010, ...).
- Local Particle Filters (Poterjoy, 2016)
- Hybrid PF/EnKF (Frei and Künsch, 2012, 2013; Slivinski et al., 2015)
- Robust EnKF (Roh et al., 2013)
- Rank Histogram Filters (Anderson, 2010)
- Anamorphism (Lien et al., 2013; Amezcua and van Leeuwen, 2014)
- GIGG-EnKF (Bishop, 2016)
- Moment Matching EnKF (Lei and Bickel, 2011)
- Reviews/Comparisons (Bocquet et al., 2010; Lei et al., 2010)
- ...

Basic idea of extended EnKFs

Conditional on $\mathbf{y}_t$ and $\boldsymbol{\theta}_t$, the HSSM reduces to the "standard" SSM, for which the EnKF is applicable

$\rightarrow$ Take existing techniques for Bayesian inference (e.g., Gibbs sampler, particle filter), but replace the part requiring integrating out or sampling from $\mathbf{x}_t$ by the EnKF

Examples:

- Gibbs-EnKF
- Particle-EnKF
- Gibbs-EnKS
- ...

Basic idea of extended EnKFs

Conditional on $\mathbf{y}_t$ and $\boldsymbol{\theta}_t$, the HSSM reduces to the "standard" SSM, for which the EnKF is applicable

$\rightarrow$ Take existing techniques for Bayesian inference (e.g., Gibbs sampler, particle filter), but replace the part requiring integrating out or sampling from $\mathbf{x}_t$ by the EnKF

Examples:

- Gibbs-EnKF
- Particle-EnKF
- Gibbs-EnKS
- ...

Algorithm 1: Gibbs EnKF (GEnKF)

Assume: forecasts of state and parameters are independent, $\mathcal{M}_t(\mathbf{x}_{t-1})$ does not depend on $\boldsymbol{\theta}_t$, and $\boldsymbol{\theta}_t$ are independent over time. Then:

For $t = 0$: Draw $(\mathbf{x}_0^{(i)}, \boldsymbol{\theta}_0^{(i)})$ from $\mathcal{N}_n(\mathbf{a}_0, \mathbf{P}_0)p(\boldsymbol{\theta}_0)$, $i = 1, \ldots, M$.

For $t \geq 1$:

1. Forecast step: $\widetilde{\mathbf{x}}_t^{(i)} = \mathcal{M}_t(\mathbf{x}_{t-1}^{(i)})$, $i = 1, \ldots, M$.
2. Initialize $\mathbf{y}_t^{(i)}$ and $\boldsymbol{\theta}_t^{(i)}$, for $i = 1, \ldots, M$.
3. For $i = 1, \ldots, M$, iterate between the following steps until convergence:
   (a) Sample $\mathbf{x}_t^{(i)}$ from $\hat{p}(\mathbf{x}_t | \mathbf{y}_t^{(i)}, \boldsymbol{\theta}_t^{(i)}, \widetilde{\mathbf{x}}_t^{(1:M)})$ using EnKF update.
   (b) Sample $\mathbf{y}_t^{(i)}$ from $p(\mathbf{y}_t | \mathbf{x}_t^{(i)}, \boldsymbol{\theta}_t^{(i)}, \mathbf{z}_t)$.
   (c) Sample $\boldsymbol{\theta}_t^{(i)}$ from $p(\boldsymbol{\theta}_t | \mathbf{y}_t^{(i)}, \mathbf{x}_t^{(i)}, \mathbf{z}_t)$.

Then, each $(\mathbf{x}_t^{(i)}, \boldsymbol{\theta}_t^{(i)})$ is a joint sample from $p(\mathbf{x}_t, \boldsymbol{\theta}_t | \mathbf{z}_{1:t})$.

Algorithm 2: Particle EnKF (for low-dim parameters)

Initialize the algorithm with an ensemble of ensembles: $(\boldsymbol{\theta}_0^{(i)}, \mathbf{x}_0^{(i,j)})$ with $w_0^{(i)} = 1/M$, $i = 1, \ldots, M$; $j = 1, \ldots, N$. Then, for each time $t \geq 1$:

1. For $i = 1, \ldots, M$:
   - (a) Sample a particle $\boldsymbol{\theta}_t^{(i)}$ from $p(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1}^{(i)})$
   - (b) Propagate the ensemble: $\widetilde{\mathbf{x}}_t^{(i,j)} = \mathcal{M}_t(\mathbf{x}_{t-1}^{(i,j)}, \boldsymbol{\theta}_t^{(i)}), j = 1, \ldots, N$.
   - (c) Calculate particle weight: $w_t^{(i)} \propto w_{t-1}^{(i)} \mathcal{L}_t^Z(\mathbf{z}_t | \boldsymbol{\theta}_t^{(i)}, \widetilde{\mathbf{x}}_t^{(i,1:N)})$.
   - (d) Generate $\mathbf{x}_t^{(i,j)}$ from $\hat{p}(\mathbf{x}_t | \mathbf{z}_t, \boldsymbol{\theta}_t^{(i)}, \widetilde{\mathbf{x}}_t^{(i,1:N)}), j = 1, \ldots, N$, using EnKF.
2. Filtering distribution: $p(\boldsymbol{\theta}_t, \mathbf{x}_t | \mathbf{z}_{1:t}) \approx \sum_{i=1}^M w_t^{(i)} \frac{1}{N} \sum_{j=1}^N \delta_{(\boldsymbol{\theta}_t^{(i)}, \mathbf{x}_t^{(i,j)})}(\boldsymbol{\theta}_t, \mathbf{x}_t)$
3. If desired, resample the particles $(\boldsymbol{\theta}_t^{(i)}, \mathbf{x}_t^{(i,1:N)})$ with weights $w_t^{(i)}$ to obtain an unweighted sample

Likelihood $\mathcal{L}_t^Z(\mathbf{z}_t | \boldsymbol{\theta}_t, \widetilde{\mathbf{x}}_t^{(i,1:N)})$ is approximated using EnKF

In the case of forecast independence of states and parameters, only need a single ensemble (cf. Frei and Künsch, 2012)

Algorithm 3: Gibbs-EnKS

Want smoothing distribution $p(\mathbf{x}_{1:T}, \boldsymbol{\theta}_{1:T} | \mathbf{z}_{1:T})$ for fixed T.

1. Initialize $\mathbf{y}_{1:T}$ and $\boldsymbol{\theta}_{1:T}$.
2. Iterate between following steps until convergence:
   (a) Draw a sample $\mathbf{x}_{1:T}$ from $p(\mathbf{x}_{1:T} | \boldsymbol{\theta}_{1:T}, \mathbf{y}_{1:T})$ using the EnKS*.
   (b) Draw a sample $\mathbf{y}_{1:T}$ from $p(\mathbf{y}_{1:T} | \mathbf{z}_{1:T}, \mathbf{x}_{1:T}, \boldsymbol{\theta}_{1:T})$.
   (c) Draw a sample $\boldsymbol{\theta}_{1:T}$ from $p(\boldsymbol{\theta}_{1:T} | \mathbf{x}_{1:T}, \mathbf{y}_{1:T}, \mathbf{z}_{1:T})$.

*Ensemble Kalman Smoother (Evensen and van Leeuwen, 2000).

Works even for high-dimensional parameters, if full conditional distribution is available in closed form

Properties of the extended EnKFs

- For linear Gaussian SSMs, algorithms converge to true posterior as $N \to \infty$ and M or the number of MCMC iterations increase
- For small N, algorithms will tend to perform well for HSSMs for which EnKF/EnKS works well for the embedded SSM.
- Computational complexity:
  - EnKF: Have to apply $\mathcal{M}_t$ to N ensemble members; update is $\mathcal{O}(nN^2)$ for most EnKF variants (e.g. Tippett et al., 2003)
  - Extended EnKFs: In general, have to carry out EnKF several times. But: Often only a small number of iterations or particles is necessary, and only the update has to be repeated
  - Thus, increased computational cost of extended EnKFs is minor in some applications

Outline

Data with outliers

Heavy-tailed noise distribution: $v_l \sim t_\nu(0, \sigma^2)$, ν small

Special case of our HSSM:
$\mathbf{z} = \mathbf{y}$ and $\mathbf{R}(\theta) = \sigma^2\, diag(\theta_1, \ldots, \theta_m)$, where $\theta_l \overset{ind.}{\sim} IG(\nu/2, \nu/2)$

GEnKF (Robust EnKF):
Update step: For $j = 1, \ldots, N$, repeat G times (until convergence):
(a) EnKF update of $\mathbf{x}^{(j)}$ from $\mathbf{x}^{f(j)}$ based on $\mathbf{y}$, $\theta^{(j)}$, $\mathbf{x}^{f(1:N)}$.
(b) Sample $\theta_l^{(j)} \overset{ind.}{\sim} IG(\nu/2 + \frac{1}{2}, \nu/2 + \left(\frac{y_l - (\mathbf{H}\mathbf{x}^{(j)})_l}{\sigma_t}\right)^2/2)$, $l = 1, \ldots, m$

Data with outliers

Heavy-tailed noise distribution: $v_l \sim t_\nu(0, \sigma^2)$, ν small

Special case of our HSSM:
$\mathbf{z} = \mathbf{y}$ and $\mathbf{R}(\boldsymbol{\theta}) = \sigma^2\, diag(\theta_1, \ldots, \theta_m)$, where $\theta_l \overset{ind.}{\sim} IG(\nu/2, \nu/2)$

GEnKF (Robust EnKF):
Update step: For $j = 1, \ldots, N$, repeat G times (until convergence):
(a) EnKF update of $\mathbf{x}^{(j)}$ from $\mathbf{x}^{f(j)}$ based on $\mathbf{y}$, $\boldsymbol{\theta}^{(j)}$, $\mathbf{x}^{f(1:N)}$.
(b) Sample $\theta_l^{(j)} \overset{ind.}{\sim} IG(\nu/2 + \frac{1}{2}, \nu/2 + \left(\frac{y_l - (\mathbf{H}\mathbf{x}^{(j)})_l}{\sigma_t}\right)^2/2)$, $l = 1, \ldots, m$

Data with outliers

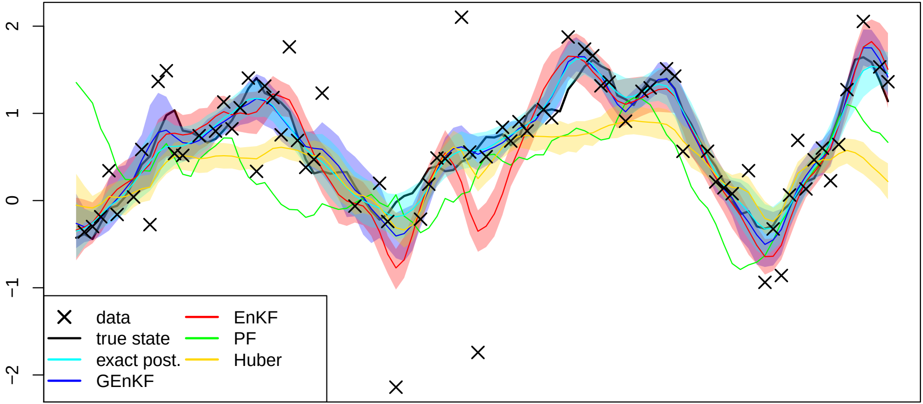
Heavy-tailed noise distribution: $v_l \sim t_\nu(0, \sigma^2)$, ν small

Special case of our HSSM:
$\mathbf{z} = \mathbf{y}$ and $\mathbf{R}(\boldsymbol{\theta}) = \sigma^2\, diag(\theta_1, \ldots, \theta_m)$, where $\theta_l \overset{ind.}{\sim} IG(\nu/2, \nu/2)$

GEnKF (Robust EnKF):
Update step: For $j = 1, \ldots, N$, repeat G times (until convergence):
(a) EnKF update of $\mathbf{x}^{(j)}$ from $\mathbf{x}^{f(j)}$ based on $\mathbf{y}$, $\boldsymbol{\theta}^{(j)}$, $\mathbf{x}^{f(1:N)}$.
(b) Sample $\theta_l^{(j)} \overset{ind.}{\sim} IG(\nu/2 + \frac{1}{2}, \nu/2 + \big(\frac{y_l - (\mathbf{H}\mathbf{x}^{(j)})_l}{\sigma_t}\big)^2/2)$, $l = 1, \ldots, m$

Example: Heavy-tailed data

Simulated heavy-tailed data (with $v_l/\sigma \sim t_2$)



Example: Threshold Models

- Challenge: Observation distributions with point masses (e.g., binary)
- Use transformation equations involving indicator functions
- Example: Rainfall amounts:

$$z_l = g(y_l; \boldsymbol{\theta}) = \begin{cases} y_l^{\kappa}, & y_l > 0 \\ 0, & y_l \leq 0 \end{cases}$$

for some $\kappa > 1$. Assume $\mathbf{R} = \sigma^2 \mathbf{I}_m$.

Gibbs-EnKF for rainfall data:

- Step 3(c): Independently:

$$y_l | z_l, \mathbf{x}, \kappa \begin{cases} = z_l^{1/\kappa}, & z_l > 0 \\ \sim \mathcal{N}^{-}\left((\mathbf{Hx})_l, \sigma^2\right), & z_l = 0 \end{cases}$$

- Step 3(b): If κ is unknown, sample from $p(\kappa | \mathbf{x}^{(j)}, \mathbf{z})$

Example: Threshold Models

- Challenge: Observation distributions with point masses (e.g., binary)
- Use transformation equations involving indicator functions
- Example: Rainfall amounts:

$$z_l = g(y_l; \boldsymbol{\theta}) = \begin{cases} y_l^{\kappa}, & y_l > 0 \\ 0, & y_l \leq 0 \end{cases}$$

for some $\kappa > 1$. Assume $\mathbf{R} = \sigma^2 \mathbf{I}_m$.

Gibbs-EnKF for rainfall data:

- Step 3(c): Independently:

$$y_l | z_l, \mathbf{x}, \kappa \begin{cases} = z_l^{1/\kappa}, & z_l > 0 \\ \sim \mathcal{N}^{-}\left((\mathbf{Hx})_l, \sigma^2\right), & z_l = 0 \end{cases}$$

- Step 3(b): If κ is unknown, sample from $p(\kappa | \mathbf{x}^{(j)}, \mathbf{z})$

Example: Threshold Models

- Challenge: Observation distributions with point masses (e.g., binary)
- Use transformation equations involving indicator functions
- Example: Rainfall amounts:

$$z_l = g(y_l; \boldsymbol{\theta}) = \begin{cases} y_l^{\kappa}, & y_l > 0 \\ 0, & y_l \leq 0 \end{cases}$$

for some $\kappa > 1$. Assume $\mathbf{R} = \sigma^2 \mathbf{I}_m$.

Gibbs-EnKF for rainfall data:

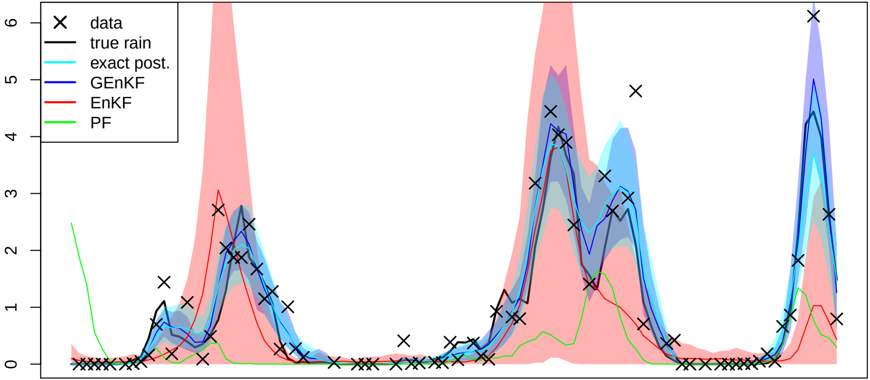
- Step 3(c): Independently:

$$y_l | z_l, \mathbf{x}, \kappa \begin{cases} = z_l^{1/\kappa}, & z_l > 0 \\ \sim \mathcal{N}^{-}\big((\mathbf{Hx})_l, \sigma^2\big), & z_l = 0 \end{cases}$$

- Step 3(b): If κ is unknown, sample from $p(\kappa | \mathbf{x}^{(j)}, \mathbf{z})$

Example: Threshold models

Simulated rainfall data (with $\kappa = 3$)



Simulation study: Non-Gaussian obs. at a single time point

Simulated heavy-tailed and rainfall data. Compared GEnKF update to (matrix-free) EnKF and PF (importance sampler):

	Heavy-tailed		Rainfall		Rain (κ unkn.)	
	MSPE	CRPS	MSPE	CRPS	MSPE	CRPS
exact	0.175	0.096	0.452	0.146	0.450	0.146
GEnKF	0.195	0.107	0.470	0.154	0.895	0.232
EnKF	0.289	0.154	13.907	3.848	>100	>100
PF	0.764	0.582	2.033	1.037	4.289	1.861
Huber (Roh et al., 2013)	0.629	0.405				

Details:

- Simulated 100 true states of size $n = 100$; $m = 75$ randomly chosen observations; $\mathbf{H}$ is subset of identity matrix
- True state distribution: mean 0.2, powered exponential covariance with power 1.8 and scale 10. $\sigma \equiv 0.2$
- EnKF and PF: $N = 100$. GEnKF: $N = 30$; 1 or 3 iterations
- Wendland taper with range 20

Smoothing for Lorenz-96

- $n = 40$ equally-spaced locations on a circle
- Observations at all locations
- Observations every $\delta t = 0.2$, $T = 10$ observation times
- Gaussian observations ($\mathbf{z}_t = \mathbf{y}_t$)
- $\mathcal{M}_t(\mathbf{x}_{t-1}) = \theta \, \text{Lorenz}_{8,0.2}(\mathbf{x}_{t-1})$
- $\mathbf{H}_t = \mathbf{R}_t = \mathbf{I}_n, \quad \mathbf{Q}_t = 0.2\boldsymbol{\Sigma}_L$
- Prior: $\theta \sim \mathcal{N}(0.8, 0.2^2)$.

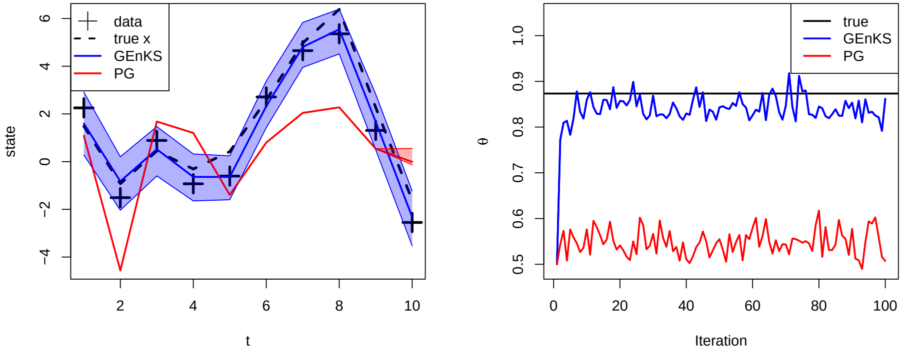
Goal: Find smoothing distribution $p(\theta, \mathbf{x}_{1:T} | \mathbf{y}_{1:T})$

Compared three methods on 100 simulated datasets:

1. EnKS with state augmentation with $N = 1000$, $\theta_t \sim \mathcal{N}(\theta_{t-1}, 0.1^2)$
2. Gibbs-EnKS
3. Particle Gibbs sampler (Andrieu et al., 2010)

For Gibbs-EnKS and particle-Gibbs: $N = 50$; 100 Gibbs iterations; $p(\theta | \mathbf{x}_{1:T}, \mathbf{y}_{1:T})$ is in closed form

Lorenz-96 results for one simulated dataset



(a) State at loc. 1 over time ($\mathbf{x}_{1:T,1}$)

(b) Trace plots for θ

Figure: (a) True state and posterior means and 80% intervals for state variable x_1. (b) True value and trace plots for parameter θ.

Lorenz-96 results averaged over 100 datasets

	Parameter θ		State $\mathbf{x}_{1:T}$	
	MSPE	CRPS	MSPE	CRPS
Gibbs-EnKS	0.002	0.024	0.710	0.478
EnKS+SA	>100	>100	0.914	0.540
Particle Gibbs	0.111	0.262	12.380	2.495
Prior	0.042	0.118		

- EnKS with state augmentation diverges
- Particle Gibbs sampler produces worse inference on θ than simply using the prior distribution (i.e., completely ignoring the data)

Outline

Summary

- EnKF handles high-dimensional, nonlinear SSMs, but is less appropriate under non-Gaussianity or for unknown parameters
- Extended EnKFs handle more general, hierarchical SSMs, including unknown parameters and non-Gaussian observations
- In some cases, computational effort is similar to EnKF
- Results are approximate, but asymptotically correct for linear models

In general, extended EnKFs can only work well if the embedded EnKF for known parameters works well for the problem at hand

References

This talk is largely based on 2 papers:

- Katzfuss, M., Stroud, J.R., and Wikle, C.K. 2016. Understanding the ensemble Kalman filter. *The American Statistician*, 70(4), 350–357.
- Katzfuss, M., Stroud, J.R., and Wikle, C.K. 2017+. Extended ensemble Kalman filters for high-dimensional hierarchical state-space models. *arXiv:1704.06988*.

Funding

- Katzfuss: NSF DMS–1521676 and DMS–1654083
- Wikle: NSF SES-1132031

References I

Amezcua, J. and van Leeuwen, P. J. (2014). Gaussian anamorphosis in the analysis step of the EnKF: A joint state-variable/observation approach. *Tellus A*, 66:23493.

Anderson, J. L. (2001). An ensemble adjustment Kalman filter for data assimilation. *Monthly Weather Review*, 129:2884–2903.

Anderson, J. L. (2007a). An adaptive covariance inflation error correction algorithm for ensemble filters. *Tellus A*, 59:210–224.

Anderson, J. L. (2007b). Exploring the need for localization in ensemble data assimilation using an hierarchical ensemble filter. *Physica D*, 230:99–111.

Anderson, J. L. (2009). Spatially and temporally varying adaptive covariance inflation for ensemble filters. *Tellus*, 61:72–83.

Anderson, J. L. (2010). An non-Gaussian ensemble filter update for data assimilation. *Monthly Weather Review*, 138:4186–4198.

Anderson, J. L. and Anderson, S. L. (1999). A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Monthly Weather Review*, 127:2741–2758.

Andrieu, C., Doucet, A., and Holenstein, R. (2010). Particle Markov chain Monte Carlo. *Journal of the Royal Statistical Society, Series B*, 72:1–33.

Bishop, C. H. (2016). The GIGG-EnKF: Ensemble Kalman filtering for highly skewed non-negative uncertainty distributions. *Quarterly Journal of the Royal Meteorological Society*, 141:1395–1412.

Bocquet, M., Pires, C. A., and Wu, L. (2010). Beyond Gaussian statistical modeling in geophysical data assimilation. *Monthly Weather Review*, 138(8):2997–3023.

Bocquet, M. and Sakov, P. (2013). Joint state and parameter estimation with an iterative ensemble Kalman smoother. *Nonlinear Processes in Geophysics*, 20(5):803–818.

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

Dee, D. P. and da Silva, A. M. (1999). Maximum likelihood estimation of forecast and observation error covariance parameters. Part I: Methodology. *Monthly Weather Review*, 127:1822–1834.

DelSole, T. and Yang, X. (2010). State and parameter estimation in stochastic dynamical models. *Physica D*, 239(18):1781–1788.

References II

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, 99:10143–10162.

Evensen, G. and van Leeuwen, P. J. (2000). An ensemble Kalman smoother for nonlinear dynamics. *Monthly Weather Review*, 128:1852–1867.

Frei, M. and Künsch, H. R. (2012). Sequential state and observation noise covariance estimation using combined ensemble Kalman and particle filters. *Monthly Weather Review*, 140:1476–1495.

Frei, M. and Künsch, H. R. (2013). Bridging the ensemble Kalman and particle filters. *Biometrika*, 100:781–800.

Gordon, N. J., Salmond, D. J., and Smith, A. F. M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. In *IEE Proceedings*, volume F-140, pages 107–113. IEE.

Gu, Y. and Oliver, D. S. (2007). An iterative ensemble Kalman filter for multiphase fluid flow data assimilation. *SPE Journal*, 12(4):438–446.

Hamill, T. M., Whitaker, J., and Snyder, C. (2001). Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Monthly Weather Review*, 129:2776–2790.

Houtekamer, P. L. and Mitchell, H. L. (1998). Data assimilation using an ensemble Kalman filter technique. *Monthly Weather Review*, 126:796–811.

Houtekamer, P. L. and Mitchell, H. L. (2001). A sequential ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*, 129:123–137.

Kalman, R. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35–45.

Katzfuss, M., Stroud, J. R., and Wikle, C. K. (2016). Understanding the ensemble Kalman filter. *The American Statistician*, 70(4):350–357.

Katzfuss, M., Stroud, J. R., and Wikle, C. K. (2017). Extended ensemble Kalman filters for high-dimensional hierarchical state-space models. Technical report, ArXiv:1704.06988.

Lei, J. and Bickel, P. (2011). A moment matching ensemble filter for nonlinear non-Gaussian data assimilation. *Monthly Weather Review*, 139:3964–3973.

References III

Lei, J., Bickel, P., and Snyder, C. (2010). Comparison of ensemble Kalman filters under non-Gaussianity. *Monthly Weather Review*, 138:1293–1306.

Li, H., Kalnay, E., and Miyoshi, T. (2009). Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter. *Quarterly Journal of the Royal Meterological Society*, 135:534–533.

Lien, G.-Y., Kalnay, E., and Miyoshi, T. (2013). Effective assimilation of global precipitation: Simulation experiments. *Tellus A*, 65:19915.

Mitchell, H. L. and Houtekamer, P. L. (2000). An adaptive ensemble Kalman filter. *Monthly Weather Review*, 128:416–433.

Miyoshi, T. (2011). The Gaussian approach to adaptive covariance inflation and its implementation with the local ensemble transform Kalman filter. *Monthly Weather Review*, 139:1519–1535.

Ott, E., Hunt, B. R., Szunyogh, I., Zimin, A. V., Kostelich, E. J., Corazza, M., Kalnay, E., Patil, D. J., and Yorke, J. A. (2004). A local ensemble Kalman filter for atmospheric data assimilation. *Tellus*, 56A:415–428.

Poterjoy, J. (2016). A localized particle filter for high-dimensional nonlinear systems. *Monthly Weather Review*, 144:59–76.

Roh, S., Genton, M. G., Jun, M., Szunyogh, I., and Hoteit, I. (2013). Observation quality control with a robust ensemble Kalman filter. *Monthly Weather Review*, 141(12):4414–4428.

Slivinski, L. and Snyder, C. (2016). Exploring practical estimates of the ensemble size necessary for particle filters. *Monthly Weather Review*, 144:861–875.

Slivinski, L., Spiller, E., Apte, A., and Standstede, B. (2015). A hybrid particle-ensemble Kalman filter for Lagrangian data assimilation. *Monthly Weather Review*, 143:195–211.

Snyder, C., Bengtsson, T., Bickel, P., and Anderson, J. L. (2008). Obstacles to high-dimensional particle filtering. *Monthly Weather Review*, 136:4629–4640.

Stroud, J. R. and Bengtsson, T. (2007). Sequential state and variance estimation within the ensemble Kalman filter. *Monthly Weather Review*, 135:3194–3208.

Stroud, J. R., Katzfuss, M., and Wikle, C. K. (2018). A Bayesian adaptive ensemble Kalman filter for sequential state and parameter estimation. *Monthly Weather Review*, 146(1):373–386.

References IV

Tippett, M. K., Anderson, J. L., Bishop, C. H., Hamill, T. M., and Whitaker, J. S. (2003). Ensemble square-root filters. *Monthly Weather Review*, 131:1485–1490.

van Leeuwen, P. J. (2010). Nonlinear data assimilation in Geosciences: An extremely efficient particle filter. *Quarterly Journal of the Royal Meteorological Society*, 136:1991–1999.

Whitaker, J. S. and Hamill, T. M. (2002). Ensemble data assimilation without perturbed observations. *Monthly Weather Review*, 130:1913–1924.

Zhang, F., Snyder, C., and Sun, J. (2004). Impacts of initial estimate and observation availability on convective scale data assimilation with an ensemble Kalman filter. *Monthly Weather Review*, 132:1238–1253.