Ensemble Kalman filter and related filters

Schedule

- 16 Jan: Gaussian processes (Jo Eidsvik)
- 23 Jan: Hands-on project on Gaussian processes (Team effort, work in groups)
- 30 Jan: Latent Gaussian models and INLA (Jo Eidsvik)
- 6 Feb: Hands-on project on INLA (Team effort, work in groups)
- 12-13 Feb: Template model builder. (Guest lecturer Hans J Skaug)
- **20 Feb: Ensemble Kalman filter and related filters** (Jo Eidsvik)
- 27 Feb: Ensemble Kalman filter and related filters (Henning Omre)
- 6 Mars: Discrete random processes (Håkon Tjelmeland)
- 13 Mars: **No Lecture!**
- 20 Mars: Discrete random processes (Håkon Tjelmeland)
- 10 April: Dim red?..
-

Bayes formula

$$p(\boldsymbol{x}|\boldsymbol{y}) = \frac{p(\boldsymbol{x},\boldsymbol{y})}{p(\boldsymbol{y})} = \frac{p(\boldsymbol{x})p(\boldsymbol{y}|\boldsymbol{x})}{\int p(\boldsymbol{x})p(\boldsymbol{y}|\boldsymbol{x})d\boldsymbol{x}}$$

$p(\boldsymbol{x})$ is prior.
$p(\boldsymbol{y}|\boldsymbol{x})$ is likelihood.

Example of such a model

$$\boldsymbol{x} \sim N(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x)$$

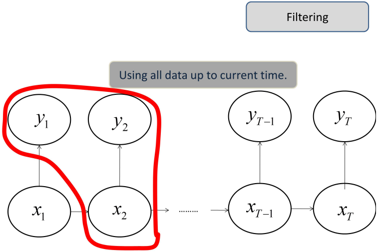
$$\boldsymbol{y}|\boldsymbol{x} \sim N(\boldsymbol{H}\boldsymbol{x}, \boldsymbol{T}), \quad \boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \epsilon, \;\; \epsilon \sim N(\boldsymbol{0}, \boldsymbol{T})$$

$p(\boldsymbol{x}|\boldsymbol{y})$ is Gaussian with

$$\begin{aligned} E(\boldsymbol{x}|\boldsymbol{y}) &= \boldsymbol{\mu}_x + \boldsymbol{\Sigma}_x \boldsymbol{H}'[\boldsymbol{H}\boldsymbol{\Sigma}_x \boldsymbol{H}' + \boldsymbol{T}]^{-1}(\boldsymbol{y} - \boldsymbol{H}\boldsymbol{\mu}_x), \\ \mathrm{Var}(\boldsymbol{x}|\boldsymbol{y}) &= \boldsymbol{\Sigma}_x - \boldsymbol{\Sigma}_x \boldsymbol{H}'[\boldsymbol{H}\boldsymbol{\Sigma}_x \boldsymbol{H}' + \boldsymbol{T}]^{-1} \boldsymbol{H}\boldsymbol{\Sigma}_x. \end{aligned}$$

Mean is linear in conditioning variable.
Variance is not dependent on conditioning variable, only correlations and variances.

Sequential Bayesian assimilation



$$p(\boldsymbol{x}_1), \quad p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}), \quad p(\boldsymbol{y}_t|\boldsymbol{x}_t), \; j = 2, 3, \ldots, T.$$

Dynamic model

Process model is described by:

$$p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1},\ldots,\boldsymbol{x}_1) = p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}),$$

This could be a differential equation, or it could be a simple linear process, or even a static process ($\boldsymbol{x}_t = \boldsymbol{x}_{t-1}$).

The data gathering process is described via the likelihood:

$$p(\boldsymbol{y}_t|\boldsymbol{x}_t,\ldots,\boldsymbol{x}_1,\boldsymbol{y}_{t-1},\ldots,\boldsymbol{y}_1) = p(\boldsymbol{y}_t|\boldsymbol{x}_t)$$

This could also be nonlinear, or it could represent picking a subset of variables (with noise).

General formula

Filtering, solution:

$$
\begin{aligned}
p(\boldsymbol{x}_t|\boldsymbol{y}_1,\ldots,\boldsymbol{y}_{t-1}) &\propto \int p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})p(\boldsymbol{x}_{t-1}|\boldsymbol{y}_1,\ldots,\boldsymbol{y}_{t-1})d\boldsymbol{x}_{t-1}. \\
p(\boldsymbol{x}_t|\boldsymbol{y}_1,\ldots,\boldsymbol{y}_t) &= \frac{p(\boldsymbol{x}_t|\boldsymbol{y}_1,\ldots,,\boldsymbol{y}_{t-1})p(\boldsymbol{y}_t|\boldsymbol{x}_t)}{p(\boldsymbol{y}_t|\boldsymbol{y}_1,\ldots,\boldsymbol{y}_{t-1})}
\end{aligned}
$$

Note Markov assumption in process, and conditionally independent data.

Kalman filter

For a Gaussian prior $\boldsymbol{x}_1 \sim N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$, linear Gaussian dynamic model $\boldsymbol{x}_t|\boldsymbol{x}_{t-1} \sim N(\boldsymbol{G}_t\boldsymbol{x}_{t-1}, \boldsymbol{Q})$, and linear Gaussian likelihood $\boldsymbol{y}_t|\boldsymbol{x}_t \sim N(\boldsymbol{H}_t\boldsymbol{x}_t, \boldsymbol{R})$, there exists an exact recursion for the filtering distribution: $p(\boldsymbol{x}_t|\boldsymbol{y}_1, \ldots, \boldsymbol{y}_t) = N(\boldsymbol{m}_t, \boldsymbol{V}_t)$.

- Initialization:

$$\boldsymbol{\mu}_1 = \boldsymbol{\mu}_1, \quad \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_1,$$

- Recursive updating for $j = 1, \ldots, T$:

$$\begin{aligned}
\boldsymbol{S}_t &= \boldsymbol{H}_t\boldsymbol{\Sigma}_t\boldsymbol{H}_t^t + \boldsymbol{R}, \\
\boldsymbol{K}_t &= \boldsymbol{\Sigma}_t\boldsymbol{H}_t^t\boldsymbol{S}_t^{-1}, \\
\boldsymbol{m}_t &= \boldsymbol{\mu}_t + \boldsymbol{K}_t(\boldsymbol{y}_t - \boldsymbol{H}_t\boldsymbol{\mu}_t), \\
\boldsymbol{V}_t &= \boldsymbol{\Sigma}_t - \boldsymbol{K}_t\boldsymbol{H}_t\boldsymbol{\Sigma}_t. \\
\boldsymbol{\Sigma}_{j+1} &= \boldsymbol{G}_t\boldsymbol{V}_t\boldsymbol{G}_t^t + \boldsymbol{Q} \\
\boldsymbol{\mu}_{j+1} &= \boldsymbol{G}_t\boldsymbol{m}_t
\end{aligned}$$

Non-Gaussian or non-linear

In other situations there is usually no exact solution to the filtering distribution. Approximations:

- Extended Kalman filter (EKF) : linearization
- Unscented Kalman filter (UKF) : design points and 'numerical' integration
- Ensemble Kalman filter (EnKF) : Monte Carlo samples and linear updates
- Particle filter (PF) : Simulation and likelihood weighting.

Algorithms

Summary of some filtering methods ; pros and cons.

Criterion	EKF	UKF	EnKF	PF
Analytic conditioning	V	V	V	
MC based			V	V
Non-linear	w	V	V	V
Scales with dim.	V		V	
Reliable UQ		w		w

Ensemble Kalman filter (EnKF)

- Monte Carlo based data assimilation
- Assimilation based on **linear** update

(Evensen, 1994, Evensen, 2009).

EnKF

- Initial: Independent prior samples $\boldsymbol{x}_0^{b,a} \sim p(\boldsymbol{x}_0)$, $b = 1, \ldots, B$.
- Iterate for samples $b = 1, \ldots, B$ and time steps $t = 1, \ldots, T$:
  Forecast variables (could be non-linear, black-box solver):

  $$\boldsymbol{x}_t^{b,f} = \boldsymbol{g}(\boldsymbol{x}_{t-1}^{b,a}; \boldsymbol{\epsilon}_t^b),$$

  Forecast data (could be non-linear, black-box solver):

  $$\boldsymbol{y}_t^b = \boldsymbol{h}(\boldsymbol{x}_t^{b,f}; \boldsymbol{\delta}_t^b),$$

  Assimilate:

  $$\boldsymbol{x}_t^{b,a} = \boldsymbol{x}_t^{b,f} + \hat{\boldsymbol{K}}_t(\boldsymbol{y}_t - \boldsymbol{y}_t^b).$$

  $$\hat{\boldsymbol{K}}_t = \hat{\boldsymbol{\Sigma}}_{xy,t}\hat{\boldsymbol{\Sigma}}_{y,t}^{-1}.$$

EnKF update

$$\boldsymbol{x}_t^{b,a} = \boldsymbol{x}_t^{b,f} + \hat{\boldsymbol{\Sigma}}_{xy,t}\hat{\boldsymbol{\Sigma}}_{y,t}^{-1}(\boldsymbol{y}_t - \boldsymbol{y}_t^b).$$

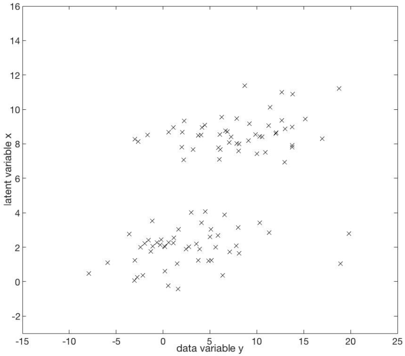
This is a regression problem.

$$\hat{\boldsymbol{\Sigma}}_{y,t} = \frac{1}{B}\sum_{b=1}^{B}(\boldsymbol{y}_t^b - \bar{\boldsymbol{y}}_t)(\boldsymbol{y}_t^b - \bar{\boldsymbol{y}}_t)', \qquad \bar{\boldsymbol{y}}_t = \frac{1}{B}\sum_{b=1}^{B}\boldsymbol{y}_t^b$$

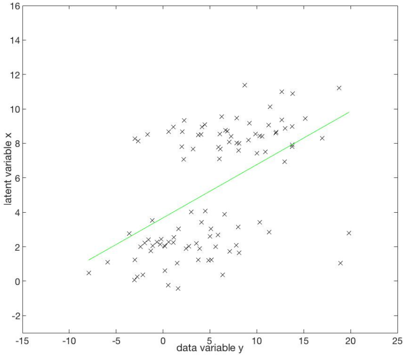
$$\hat{\boldsymbol{\Sigma}}_{xy,t} = \frac{1}{B}\sum_{b=1}^{B}(\boldsymbol{x}_t^{b,f} - \bar{\boldsymbol{x}}_t)(\boldsymbol{y}_t^b - \bar{\boldsymbol{y}}_t)', \quad \bar{\boldsymbol{x}}_t = \frac{1}{B}\sum_{b=1}^{B}\boldsymbol{x}_t^{b,f},$$

Univariate example - samples

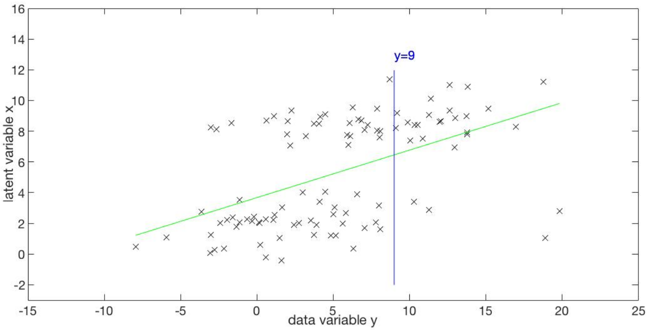
$$x^b \sim p(x), \qquad y^b = x^b + N(0, 5^2)$$



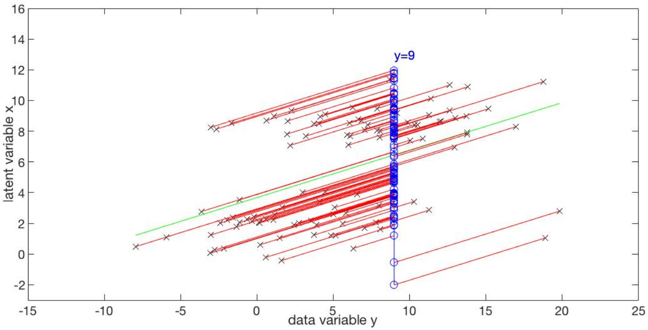
Univariate example - regression fit



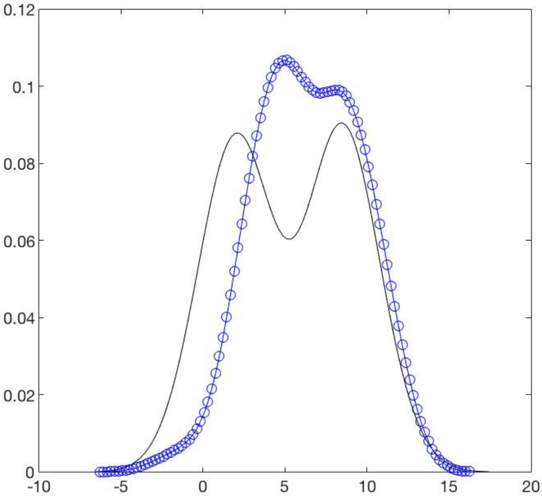
Univariate example - observation



Univariate example - analysis or update step



Univariate example - prior and posterior



Univariate - one-step problem

Implement the Ensemble Kalman filter (one time step only) on the following model, and compare with the exact solution for $p(x|y)$. Plot solutions.

$$x \sim N(0,1), \qquad y = x^{2q+1} + N(0, 0.05^2)$$

Data is $y = 2$.

- Set $q = 0$. Study performance of the EnKF for different ensemble sizes $B = 20, 100$.
- Study performance of the EnKF for different $q = 1, 2, \ldots$.

EnKF, Gauss-linear likelihood

- Initial: Independent samples $\boldsymbol{x}_0^{b,a} \sim p(\boldsymbol{x}_0)$, $b = 1, \ldots, B$.
- Iterate for samples $b = 1, \ldots, B$ and time steps $t = 1, \ldots, T$:
  Forecast variables (could be non-linear, black-box solver):

$$\boldsymbol{x}_t^{b,f} = \boldsymbol{g}(\boldsymbol{x}_{t-1}^{b,a}; \boldsymbol{\epsilon}_t^b),$$

Linear likelihood:

$$\boldsymbol{y}_t^b = \boldsymbol{H}\boldsymbol{x}_t^{b,f} + N(0, \boldsymbol{T})$$

Assimilate :

$$\boldsymbol{x}_t^{b,a} = \boldsymbol{x}_t^{b,f} + \hat{\boldsymbol{\Sigma}}_t \boldsymbol{H}^{'} (\boldsymbol{H}\hat{\boldsymbol{\Sigma}}_t \boldsymbol{H}^{'} + \boldsymbol{T})^{-1} (\boldsymbol{y}_t - \boldsymbol{y}_t^b).$$

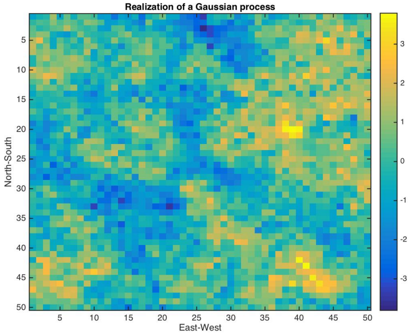
Estimation of $\boldsymbol{\Sigma}_t$

Standard approach:

$$\hat{\boldsymbol{\Sigma}}_t = \frac{1}{B}\sum_{b=1}^{B}(\boldsymbol{x}_t^{b,f} - \bar{\boldsymbol{x}}_t)(\boldsymbol{x}_t^{b,f} - \bar{\boldsymbol{x}}_t)', \qquad \bar{\boldsymbol{x}}_t = \frac{1}{B}\sum_{b=1}^{B}\boldsymbol{x}_t^{b,f}$$

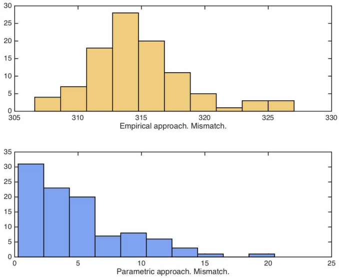
Gives less Monte Carlo error than straightforward estimator for Kalman gain.

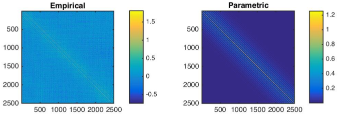
Gaussian random field



Use $B = 100$ ensembles (realizations) to estimate the covariance matrix.

Estimation of covariance matrix





There is lots of Monte Carlo error in the estimated covariance matrix and Kalman gain. Numerous tricks try to resolve this: inflation, localization, etc. Also, since the Kalman gain is estimated from data, there is coupling over many time steps which gives challenges. Resampling from a Gaussian approximation might solve this. (More details next week.).

Spatial AR(1) model

$$\boldsymbol{x}_0 \sim N(0, \boldsymbol{\Sigma}),$$

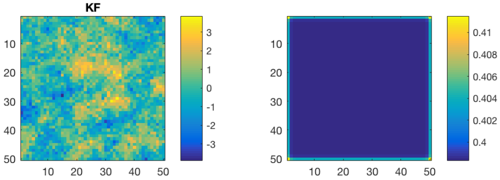
$$\boldsymbol{x}_t = \rho \boldsymbol{x}_{t-1} + N(0, (1-\rho^2)\boldsymbol{\Sigma}), \;\; t = 1, \ldots, T$$

$$\boldsymbol{y}_t = \boldsymbol{x}_t + N(0, \tau^2 \boldsymbol{I}), \;\; t = 1, \ldots, T$$

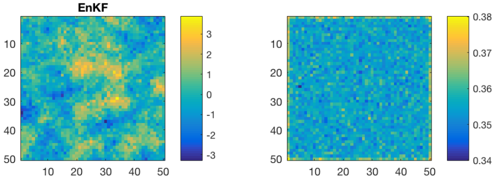
$T = 10$, 50×50 grid.

EnKF is run with $B = 5000$ ensemble members.

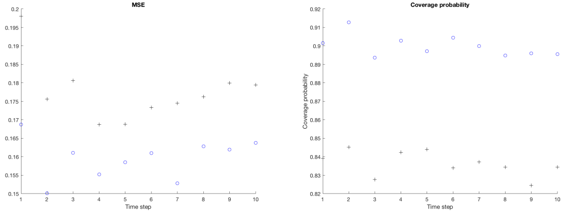
Spatial AR(1): KF pred



Spatial AR(1): EnKF pred

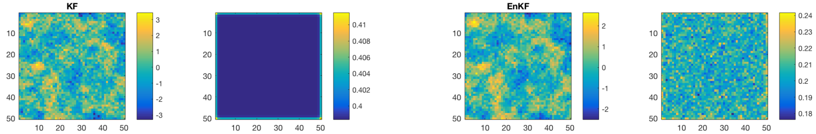


Spatial AR(1): MSE and coverage



KF (o), EnKF (+).

Spatial AR(1): EnKF pred ($B = 500$)



(Average coverage for EnKF is 0.38 for nominal 0.90.)

Univariate - AR(1) time series

Implement the Ensemble Kalman filter on the following model, and compare with the exact Kalman filter solution. Plot solutions (mean with 90 % prediction intervals). Study performance for $B = 20, 100, 1000$ in the EnKF.

$$x_1 \sim N(0,1),$$

$$x_t = 0.9x_{t-1} + N(0, 1-0.9^2), \;\; t = 2, \ldots, 50$$

$$y_t = x_t + N(0, 0.2^2), \;\; t = 1, \ldots, 50$$

Main points

- Filtering methods, sequential data assimilation.
- EnKF is ensemble based, with linear update.
- EnKF has been successful in many applications, but often shows underestimation of variability.
- EnKF is complicated in models with very nonlinear likelihoods.