Presentation at Crash Course on Data Assimilation

Geir Evensen

NORCE

Available from https://github.com/geirev/Presentations

The Variational Inverse Problem

Geir Evensen

Simple scalar example

Given the model

$$\frac{dx}{dt} = 1$$
$$x(0) = 0$$
$$x(1) = 2$$

- Overdetermined problem.
- No solution.

Allowing for errors

Relax model and conditions

$$\frac{dx}{dt} = 1 + q(t)$$
$$x(0) = 0 + a$$
$$x(1) = 2 + b$$

- Underdetermined problem.
- Infinitively many solutions.

Statistical assumption

Statistical null hypothesis, $\mathcal{H}_0$:

$$
\begin{aligned}
\overline{q(t)} &= 0, & \overline{q(t_1)q(t_2)} &= C_0\delta(t_1 - t_2), & \overline{q(t)a} &= 0,\\
\overline{a} &= 0, & \overline{a^2} &= C_0, & \overline{ab} &= 0,\\
\overline{b} &= 0, & \overline{b^2} &= C_0, & \overline{q(t)b} &= 0.
\end{aligned}
$$

Seeking a solution that:

- is close to the conditions, and
- almost satisfies the model,

by minimizing error terms.

Penalty function

- Define quadratic penalty function

$$\mathcal{J}[x] = W_0 \int_0^1 \left(\frac{dx}{dt} - 1\right)^2 dt + W_0\big(x(0) - 0\big)^2 + W_0\big(x(1) - 2\big)^2$$

  with $W_0 = C_0^{-1}$.

- Then x is an extremum if

$$\delta\mathcal{J}[x] = \mathcal{J}[x + \delta x] - \mathcal{J}[x] = \mathcal{O}\big(\delta x^2\big)$$

  when $\delta x \longrightarrow 0$.

Variation of penalty function

We have

$$
\mathcal{J}[x+\delta x] = W_0 \int_0^1 \left(\frac{dx}{dt} - 1 + \frac{d\delta x}{dt}\right)^2 dt \\
+ W_0\big(x(0) - 0 + \delta x(0)\big)^2 + W_0\big(x(1) - 2 + \delta x(1)\big)^2
$$

and we must have

$$
\int_0^1 \frac{d\delta x}{dt}\left(\frac{dx}{dt} - 1\right) dt + \delta x(0)\big(x(0) - 0\big) + \delta x(1)\big(x(1) - 2\big) = 0,
$$

From integration by part we get

$$
\delta x \left(\frac{dx}{dt} - 1\right)\Bigg|_0^1 - \int_0^1 \delta x \frac{d^2 x}{dt^2} dt + \delta x(0)\big(x(0) - 0\big) + \delta x(1)\big(x(1) - 2\big) = 0.
$$

Minimium of penalty function

This gives the following system of equations

$$\delta x(0)\left(-\frac{dx}{dt}+1+x\right)\bigg|_{t=0}=0,$$
$$\delta x(1)\left(\frac{dx}{dt}-1+x-2\right)\bigg|_{t=1}=0,$$
$$\int_0^1 \delta x\left(\frac{d^2x}{dt^2}\right)dt=0,$$

or since δx is arbitrary....

Euler-Lagrange equation

The Euler–Lagrange equation

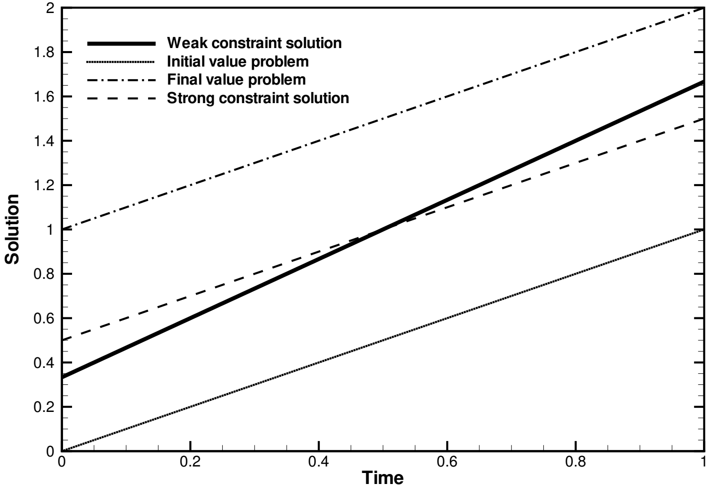
$$\begin{aligned} \frac{dx}{dt} - x &= 1 \quad \text{for } t = 0, \\ \frac{dx}{dt} + x &= 3 \quad \text{for } t = 1, \\ \frac{d^2x}{dt^2} &= 0. \end{aligned}$$

- Elliptic boundary value problem in time.
- It has a unique solution.

$$x = c_1 t + c_2,$$

with $c_1 = 4/3$ and $c_2 = 1/3$.

Results



Weak constraint solution
Initial value problem
Final value problem
Strong constraint solution
Solution
Time
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2

Summary

- Model with conditions has a unique solution.
- Additional data makes problem over determined.
- Allowing for errors gives infinitely many solutions.
- Specify mean and covariance for error terms.
- Define variational inverse problem for least-squares solution.
- Weigths are the inverses of the error covariances.
- Least-squares solution is defined by Euler-Lagrange eqs.
- Boundary value problem in time.
- Weak-constraint solution: almost satisfies dynamics and data.
- Strong-constraint solution: satisfies dynamics, and close to data.

Bayes' and the data assimilation problem

Geir Evensen

Why Bayes Theorem?

- Provides a fundamental *framework* for data assimilation.
- All data-assimilation methods can be derived from Bayes'.

Properties of a probability denisty function

- The graph of the density function is continuous, since it is defined over a continuous range over a continuous variable.
- The total probability

$$P(x) = \int_{-\infty}^{\infty} f(x)dx = 1$$

- The probability of $x \in [a, b]$ is

$$P(x \in [a, b]) = \int_a^b f(x)dx$$

- And two special cases

$$P(x = c) = \int_c^c f(x)dx = 0 \qquad \text{and} \qquad \int_{-\infty}^{\infty} \delta(x)dx = 1$$

Also, we have

- The joint probability
$$f(x,y) = f(x)f(y|x) = f(y)f(x|y)$$
- Solving for $f(x|y)$ gives Bayes' theorem
$$f(x|y) = \frac{f(x)f(y|x)}{f(y)}$$
- Bayes states that "the probability of x given y, is equal to the probability of x, times the likelihood of y given x, divided by the probability of y."
- Here $f(y)$ is a normalization constant so that the integral of $f(x|y)$ becomes one.

Bayes' theorem

Given:

- A state variable $\boldsymbol{x}$ and its prior pdf: $f(\boldsymbol{x})$
- A vector of observations $\boldsymbol{d}$ and their likelihood: $f(\boldsymbol{d}|\boldsymbol{x})$
- Bayes' theorem defines the posterior pdf, $f(\boldsymbol{x}|\boldsymbol{d})$:

Bayes' theorem

$$f(\boldsymbol{x}|\boldsymbol{d}) = \frac{f(\boldsymbol{x})f(\boldsymbol{d}|\boldsymbol{x})}{f(\boldsymbol{d})}$$

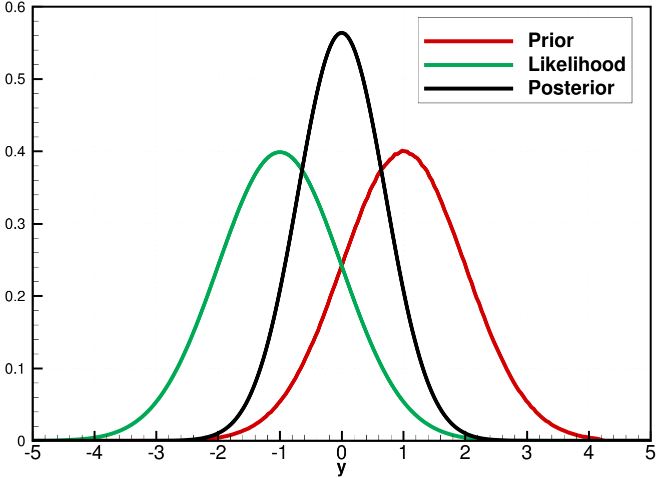
What is the likelihood function: $f(d|x)$

- The likelihood function $f(d|x)$ is the probability of the observed data d for various values of the unknown parameters x.
- The likelihood is used after data are available to describe a plausibility of a parameter value x.
- The likelihood does not have to integrate to one.

Likelihood is the plausability of a particular distribution explaining the given data. The higher the likelihood of a distribution, the more likely it is to explain the observed data.

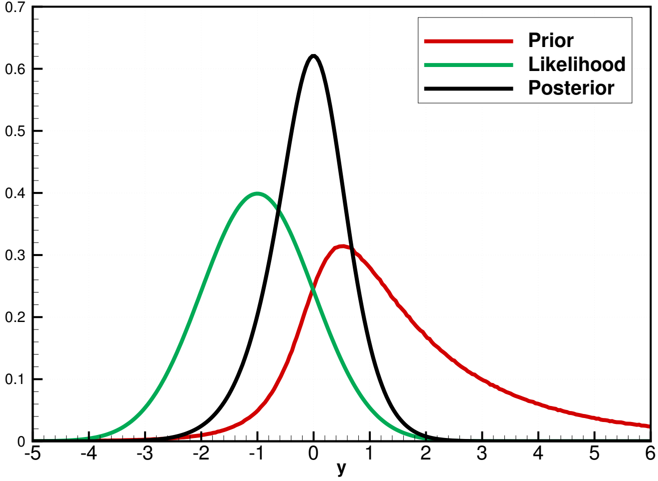
Probability is how likely are the chances of a certain data to occur if the model parameters are fixed and Likelihood is the chances of a particular model parameter explaining the given observed data.

Example of using Bayes' theorem



Available from: `https://github.com/geirev/Presentations`

Example of using Bayes' theorem



Revert to inverse problem again

With Gaussian errors

$$\frac{dx}{dt} = 1 + q(t) \qquad f(q(t)) \propto \exp\left\{-\frac{1}{2}\frac{q^2(t)}{C_{qq}}\right\}$$

$$x(0) = 0 + a \qquad f(a) \propto \exp\left\{-\frac{1}{2}\frac{a^2}{C_{aa}}\right\}$$

$$x(1) = 2 + b \qquad f(b) \propto \exp\left\{-\frac{1}{2}\frac{b^2}{C_{bb}}\right\}$$

$$f(a) f\left(\int_0^1 q(t)dt\right) f(b) = \exp\left(-\frac{1}{2}J[x]\right)$$

$$\mathcal{J}[x] = C_{qq}^{-1}\int_0^1 \left(\frac{dx}{dt} - 1\right)^2 dt + C_{aa}^{-1}\left(x(0) - 0\right)^2 + C_{bb}^{-1}\left(x(1) - 2\right)^2$$

Thus, Bayes' theorem leads to least-squares variational inverse problem for Gaussian error distributions.

Summary

- Bayes' theorem defines the "ultimate" data-assimilation problem.
- Impossible to solve in high dimensions.
- Gaussian approximation is key and leads to least-squares inverse problem.

Linear estimation theory and update equations

Geir Evensen

Best Linear Unbiased Estimator

We have given a prediction x^{f} and a measurement d:

$$
\begin{aligned}
x^{\mathrm{f}} &= x^{\mathrm{t}} + p^{\mathrm{f}}, & \overline{p^{\mathrm{f}}} &= 0 & \overline{(p^{\mathrm{f}})^2} &= C^{\mathrm{f}}_{xx} & & \\
d &= x^{\mathrm{t}} + \epsilon, & \overline{\epsilon} &= 0 & \overline{(\epsilon)^2} &= C_{dd} & \overline{(\epsilon p^{\mathrm{f}})} &= 0
\end{aligned}
$$

What is the Best Linear Unbiased Estimator (BLUE) of x?

- x^{f} could be a Forecast or a First-guess solution.
- d is a measurement.

Original BLUE derivation from system control theory

Given

$$
\begin{aligned}
x^{\mathrm{f}} &= x^{\mathrm{t}} + p^{\mathrm{f}}, & \overline{p^{\mathrm{f}}} &= 0 & \overline{(p^{\mathrm{f}})^2} &= C_{xx}^{\mathrm{f}} \\
x^{\mathrm{a}} &= x^{\mathrm{t}} + p^{\mathrm{a}}, & \overline{p^{\mathrm{a}}} &= 0 & \overline{(p^{\mathrm{a}})^2} &= C_{xx}^{\mathrm{a}} \\
d &= x^{\mathrm{t}} + \epsilon, & \overline{\epsilon} &= 0 & \overline{(\epsilon)^2} &= C_{dd}
\end{aligned}
$$

A linear unbiased estimator is

$$
\begin{aligned}
x^{\mathrm{a}} &= (1-\alpha)x^{\mathrm{f}} + \alpha d \\
&= x^{\mathrm{f}} + \alpha(d - x^{\mathrm{f}})
\end{aligned}
$$

Inserting gives

$$
\begin{aligned}
x^{\mathrm{a}} = x^{\mathrm{t}} + p^{\mathrm{a}} &= x^{\mathrm{t}} + p^{\mathrm{f}} + \alpha(x^{\mathrm{t}} + \epsilon - x^{\mathrm{t}} - p^{\mathrm{f}}) \\
p^{\mathrm{a}} &= p^{\mathrm{f}} + \alpha(\epsilon - p^{\mathrm{f}})
\end{aligned}
$$

Derivation

$$\begin{aligned}
\overline{(p^{\mathrm{a}})^2} = C_{xx}^{\mathrm{a}} &= \overline{(p^{\mathrm{f}} + \alpha(\epsilon - p^{\mathrm{f}}))^2} \\
&= \overline{(p^{\mathrm{f}})^2} + 2\alpha\overline{p^{\mathrm{f}}(\epsilon - p^{\mathrm{f}})} + \alpha^2\overline{\epsilon^2 - 2\epsilon p^{\mathrm{f}} + (p^{\mathrm{f}})^2} \\
&= C_{xx}^{\mathrm{f}} - 2\alpha C_{xx}^{\mathrm{f}} + \alpha^2(C_{dd} + C_{xx}^{\mathrm{f}}),
\end{aligned}$$

Set derivative equal to zero

$$\frac{\partial C_{xx}^{\mathrm{a}}}{\partial \alpha} = -2C_{xx}^{\mathrm{f}} + 2\alpha(C_{dd} + C_{xx}^{\mathrm{f}}) = 0.$$

to get

$$\alpha = \frac{C_{xx}^{\mathrm{f}}}{C_{xx}^{\mathrm{f}} + C_{dd}}$$

Derivation

The optimal (BLUE) solution is

$$x^{\mathrm{a}} = x^{\mathrm{f}} + \frac{C_{xx}^{\mathrm{f}}}{C_{xx}^{\mathrm{f}} + C_{dd}}(d - x^{\mathrm{f}})$$

Error estimate when inserting for α

$$C_{xx}^{\mathrm{a}} = C_{xx}^{\mathrm{f}} \left(1 - \frac{C_{xx}^{\mathrm{f}}}{C_{xx}^{\mathrm{f}} + C_{dd}}\right)$$

Derivation from Bayes' Theorem

Assuming a Gaussian prior and likelihood

$$f(x) = \exp\left\{-\frac{1}{2}(x - x^{\mathrm{f}})C_{xx}^{-1}(x - x^{\mathrm{f}})\right\}$$
$$f(d|x) = \exp\left\{-\frac{1}{2}(d - x)C_{dd}^{-1}(d - x)\right\}$$

From Bayes

$$f(x|d) \propto f(x)f(d|x)$$

By taking the logarithm we get the cost function

$$\mathcal{J}(x) = (x - x^{\mathrm{f}})C_{xx}^{-1}(x - x^{\mathrm{f}}) + (d - x)C_{dd}^{-1}(d - x)$$

Derivation from Bayes' Theorem

Derivative of cost function set to zero

$$
\begin{aligned}
\frac{1}{2}\frac{\partial \mathcal{J}(x)}{\partial x} &= (x - x^{\mathrm{f}})C_{xx}^{-1} - (d - x)C_{dd}^{-1} \\
&= x(C_{xx}^{-1} + C_{dd}^{-1}) - x^{\mathrm{f}} C_{xx}^{-1} - dC_{dd}^{-1} = 0
\end{aligned}
$$

Solve for x

$$
\begin{aligned}
x &= x^{\mathrm{f}} \frac{C_{xx}^{-1}}{C_{xx}^{-1} + C_{dd}^{-1}} + d\frac{C_{dd}^{-1}}{C_{xx}^{-1} + C_{dd}^{-1}} & \times\frac{C_{dd}C_{xx}}{C_{dd}C_{xx}} \\
&= x^{\mathrm{f}} \frac{C_{dd}}{C_{dd} + C_{xx}} + d\frac{C_{xx}}{C_{dd} + C_{xx}} & +x^{\mathrm{f}} - \frac{C_{dd} + C_{xx}}{C_{dd} + C_{xx}}x^{\mathrm{f}} \\
&= x^{\mathrm{f}} + \frac{C_{xx}}{C_{xx} + C_{dd}}(d - x^{\mathrm{f}})
\end{aligned}
$$

Summary

- The BLUE is the optimal way of combining two linear estimates of a parameter.
- We can derive it from Bayes' formula when assuming Gaussian error statistics.

$$f(x|d) \propto f(x)f(d|x)$$

$$\mathcal{J}(x) = (x - x^{\mathrm{f}})C_{xx}^{-1}(x - x^{\mathrm{f}}) + (d - x)C_{dd}^{-1}(d - x)$$

$$x^{\mathrm{a}} = x^{\mathrm{f}} + \frac{C_{xx}}{C_{xx} + C_{dd}}(d - x^{\mathrm{f}})$$

$$C_{xx}^{\mathrm{a}} = C_{xx}^{\mathrm{f}}\left(1 - \frac{C_{xx}^{\mathrm{f}}}{C_{xx}^{\mathrm{f}} + C_{dd}}\right)$$

BLUE in vector form

Geir Evensen

Vector state and data

Given a state vector $\boldsymbol{x} \in \Re^n$ and a data vector $\boldsymbol{d} \in \Re^m$.

$$\begin{aligned}
\boldsymbol{x}^{\mathrm{f}} &= \boldsymbol{x}^{\mathrm{t}} + \boldsymbol{p} & \overline{\boldsymbol{p}} &= 0 & \overline{\boldsymbol{p}\boldsymbol{p}^{\mathrm{T}}} &= \boldsymbol{C}_{xx} & & \\
\boldsymbol{d} &= \boldsymbol{H}\boldsymbol{x}^{\mathrm{t}} + \boldsymbol{\epsilon} & \overline{\boldsymbol{\epsilon}} &= 0 & \overline{\boldsymbol{\epsilon}\boldsymbol{\epsilon}^{\mathrm{T}}} &= \boldsymbol{C}_{dd} & \overline{\boldsymbol{p}\boldsymbol{\epsilon}^{\mathrm{T}}} &= 0
\end{aligned}$$

where we define the measurement operator $\boldsymbol{H} \in \Re^{m \times n}$.

As an example consider the case with $m = 2$ and $n = 10$.

$$\begin{aligned}
\boldsymbol{x}^{\mathrm{f}} &= [x_1, x_2, x_3, \ldots, x_{10}]^{\mathrm{T}} \\
\boldsymbol{d} &= [d_1, d_2]^{\mathrm{T}}
\end{aligned}$$

with

$$\boldsymbol{H} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

implying that $d_1 = x_4^{\mathrm{t}} + \epsilon_1$ and $d_2 = x_7^{\mathrm{t}} + \epsilon_2$.

Start from Bayes' Theroem

Bayes for the state vector $\boldsymbol{x}$ given the measurement vector $\boldsymbol{d}$

$$f(\boldsymbol{x}|\boldsymbol{d}) \propto f(\boldsymbol{x})f(\boldsymbol{d}|\boldsymbol{x})$$

Assume Gaussian prior and measurement errors

$$f(\boldsymbol{x}|\boldsymbol{d}) \propto \exp -\frac{1}{2}\left\{(\boldsymbol{x}-\boldsymbol{x}^{\mathrm{f}})^{\mathrm{T}}\boldsymbol{C}_{xx}^{-1}(\boldsymbol{x}-\boldsymbol{x}^{\mathrm{f}}) + (\boldsymbol{d}-\boldsymbol{H}\boldsymbol{x})^{\mathrm{T}}\boldsymbol{C}_{dd}^{-1}(\boldsymbol{d}-\boldsymbol{H}\boldsymbol{x})\right\}$$

Maximizing $f(\boldsymbol{x}|\boldsymbol{d})$ identical to minimizing

$$\mathcal{J}(\boldsymbol{x}) = (\boldsymbol{x}-\boldsymbol{x}^{\mathrm{f}})^{\mathrm{T}}\boldsymbol{C}_{xx}^{-1}(\boldsymbol{x}-\boldsymbol{x}^{\mathrm{f}}) + (\boldsymbol{d}-\boldsymbol{H}\boldsymbol{x})^{\mathrm{T}}\boldsymbol{C}_{dd}^{-1}(\boldsymbol{d}-\boldsymbol{H}\boldsymbol{x})$$

Minimum is defined by zero-gradient of cost function

$$
\begin{aligned}
\frac{1}{2}\nabla\mathcal{J}(\boldsymbol{x}) &= \boldsymbol{C}_{xx}^{-1}(\boldsymbol{x}-\boldsymbol{x}^{\mathrm{f}}) - \boldsymbol{H}^{\mathrm{T}}\boldsymbol{C}_{dd}^{-1}(\boldsymbol{d}-\boldsymbol{H}\boldsymbol{x}) \\
&= \boldsymbol{C}_{xx}^{-1}\boldsymbol{x} - \boldsymbol{C}_{xx}^{-1}\boldsymbol{x}^{\mathrm{f}} - \boldsymbol{H}^{\mathrm{T}}\boldsymbol{C}_{dd}^{-1}\boldsymbol{d} + \boldsymbol{H}^{\mathrm{T}}\boldsymbol{C}_{dd}^{-1}\boldsymbol{H}\boldsymbol{x} = 0
\end{aligned}
$$

$$
\left(\boldsymbol{C}_{xx}^{-1} + \boldsymbol{H}^{\mathrm{T}}\boldsymbol{C}_{dd}^{-1}\boldsymbol{H}\right)\boldsymbol{x} = \boldsymbol{C}_{xx}^{-1}\boldsymbol{x}^{\mathrm{f}} + \boldsymbol{H}^{\mathrm{T}}\boldsymbol{C}_{dd}^{-1}\boldsymbol{d}
$$

$$
\boldsymbol{x} = \left(\boldsymbol{C}_{xx}^{-1} + \boldsymbol{H}^{\mathrm{T}}\boldsymbol{C}_{dd}^{-1}\boldsymbol{H}\right)^{-1}\left(\boldsymbol{C}_{xx}^{-1}\boldsymbol{x}^{\mathrm{f}} + \boldsymbol{H}^{\mathrm{T}}\boldsymbol{C}_{dd}^{-1}\boldsymbol{d}\right)
$$

Minimizing solution

State space formulation

$$\boldsymbol{x} = \left(\boldsymbol{C}_{xx}^{-1} + \boldsymbol{H}^{\mathrm{T}}\boldsymbol{C}_{dd}^{-1}\boldsymbol{H}\right)^{-1}\boldsymbol{C}_{xx}^{-1}\boldsymbol{x}^{\mathrm{f}} + \left(\boldsymbol{C}_{xx}^{-1} + \boldsymbol{H}^{\mathrm{T}}\boldsymbol{C}_{dd}^{-1}\boldsymbol{H}\right)^{-1}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{C}_{dd}^{-1}\boldsymbol{d}$$

Using the following two lemmas (from the Woodbury matrix identity):

$$\boxed{\begin{aligned}\left(\boldsymbol{C}_{xx}^{-1} + \boldsymbol{H}^{\mathrm{T}}\boldsymbol{C}_{dd}^{-1}\boldsymbol{H}\right)^{-1}\boldsymbol{C}_{xx}^{-1} &= \boldsymbol{I} - \boldsymbol{C}_{xx}\boldsymbol{H}^{\mathrm{T}}(\boldsymbol{H}\boldsymbol{C}_{xx}\boldsymbol{H}^{\mathrm{T}} + \boldsymbol{C}_{dd})^{-1}\boldsymbol{H} \\ \left(\boldsymbol{C}_{xx}^{-1} + \boldsymbol{H}^{\mathrm{T}}\boldsymbol{C}_{dd}^{-1}\boldsymbol{H}\right)^{-1}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{C}_{dd}^{-1} &= \boldsymbol{C}_{xx}\boldsymbol{H}^{\mathrm{T}}(\boldsymbol{H}\boldsymbol{C}_{xx}\boldsymbol{H}^{\mathrm{T}} + \boldsymbol{C}_{dd})^{-1}\end{aligned}}$$

we obtain the observation space formulation

$$\boldsymbol{x} = \boldsymbol{x}^{\mathrm{f}} + \boldsymbol{C}_{xx}\boldsymbol{H}^{\mathrm{T}}(\boldsymbol{H}\boldsymbol{C}_{xx}\boldsymbol{H}^{\mathrm{T}} + \boldsymbol{C}_{dd})^{-1}(\boldsymbol{d} - \boldsymbol{H}\boldsymbol{x}^{\mathrm{f}})$$

* What is the update error covariance

The error variance for the update is defined as

$$\boldsymbol{C}_{xx}^{\mathrm{a}} = \overline{(\boldsymbol{x}^{\mathrm{t}} - \boldsymbol{x}^{\mathrm{a}})(\boldsymbol{x}^{\mathrm{t}} - \boldsymbol{x}^{\mathrm{a}})^{\mathrm{T}}}$$

Let us define for simplicity

$$\begin{aligned}
\boldsymbol{C} &= \boldsymbol{H}\boldsymbol{C}_{xx}^{\mathrm{f}}\boldsymbol{H}^{\mathrm{T}} + \boldsymbol{C}_{dd} \\
\boldsymbol{R} &= \boldsymbol{H}\boldsymbol{C}_{xx}^{\mathrm{f}} \\
\boldsymbol{h} &= \boldsymbol{d} - \boldsymbol{H}\boldsymbol{x}^{\mathrm{f}} \\
&= \boldsymbol{H}\boldsymbol{x}^{\mathrm{t}} + \boldsymbol{\epsilon} - \boldsymbol{H}\boldsymbol{x}^{\mathrm{f}} = \boldsymbol{H}(\boldsymbol{x}^{\mathrm{t}} - \boldsymbol{x}^{\mathrm{f}}) + \boldsymbol{\epsilon}
\end{aligned}$$

And we can write

$$\boldsymbol{x}^{\mathrm{a}} = \boldsymbol{x}^{\mathrm{f}} + \boldsymbol{R}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{h}$$

* Update error covariance

$$
\begin{aligned}
\boldsymbol{C}^{\mathrm{a}}_{xx} &= \overline{(\boldsymbol{x}^{\mathrm{t}} - \boldsymbol{x}^{\mathrm{a}})(\boldsymbol{x}^{\mathrm{t}} - \boldsymbol{x}^{\mathrm{a}})^{\mathrm{T}}} \\
&= \overline{(\boldsymbol{x}^{\mathrm{t}} - \boldsymbol{x}^{\mathrm{f}} - \boldsymbol{R}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{h})(\boldsymbol{x}^{\mathrm{t}} - \boldsymbol{x}^{\mathrm{f}} - \boldsymbol{R}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{h})^{\mathrm{T}}} \\
&= \overline{(\boldsymbol{x}^{\mathrm{t}} - \boldsymbol{x}^{\mathrm{f}})(\boldsymbol{x}^{\mathrm{t}} - \boldsymbol{x}^{\mathrm{f}})^{\mathrm{T}}} && (\to \boldsymbol{C}^{\mathrm{f}}_{xx}) \\
&- 2\overline{(\boldsymbol{x}^{\mathrm{t}} - \boldsymbol{x}^{\mathrm{f}})(\boldsymbol{R}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{h})^{\mathrm{T}}} \\
&+ \overline{(\boldsymbol{R}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{h})(\boldsymbol{R}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{h})^{\mathrm{T}}}
\end{aligned}
$$

* Second term

$$
\begin{aligned}
& - 2\overline{(\boldsymbol{x}^{\mathrm{t}} - \boldsymbol{x}^{\mathrm{f}})(\boldsymbol{R}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{h})^{\mathrm{T}}} \\
&= -2\overline{(\boldsymbol{x}^{\mathrm{t}} - \boldsymbol{x}^{\mathrm{f}})\boldsymbol{h}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{R}} \\
&= -2\overline{(\boldsymbol{x}^{\mathrm{t}} - \boldsymbol{x}^{\mathrm{f}})(\boldsymbol{H}(\boldsymbol{x}^{\mathrm{t}} - \boldsymbol{x}^{\mathrm{f}}) + \boldsymbol{\epsilon})^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{R}} \\
&= -2\overline{(\boldsymbol{x}^{\mathrm{t}} - \boldsymbol{x}^{\mathrm{f}})(\boldsymbol{x}^{\mathrm{t}} - \boldsymbol{x}^{\mathrm{f}})^{\mathrm{T}}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{R}} \\
&= -2\boldsymbol{C}^{\mathrm{f}}_{xx}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{R} \\
&= -2\boldsymbol{R}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{R}
\end{aligned}
$$

* Third term

$$
\begin{aligned}
&\overline{(\boldsymbol{R}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{h})(\boldsymbol{R}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{h})^{\mathrm{T}}} \\
&= \overline{\boldsymbol{R}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{h}\boldsymbol{h}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{R}} \\
&= \overline{\boldsymbol{R}^{\mathrm{T}}\boldsymbol{C}^{-1}(\boldsymbol{H}(\boldsymbol{x}^{\mathrm{t}}-\boldsymbol{x}^{\mathrm{f}})+\boldsymbol{\epsilon})(\boldsymbol{H}(\boldsymbol{x}^{\mathrm{t}}-\boldsymbol{x}^{\mathrm{f}})+\boldsymbol{\epsilon})^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{R}} \\
&= \boldsymbol{R}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{C}\boldsymbol{C}^{-1}\boldsymbol{R} \\
&= \boldsymbol{R}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{R}
\end{aligned}
$$

* Update error covariance

$$
\begin{aligned}
\boldsymbol{C}_{xx}^{\mathrm{a}} &= \boldsymbol{C}_{xx}^{\mathrm{f}} - 2\boldsymbol{R}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{R} + \boldsymbol{R}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{R} \\
&= \boldsymbol{C}_{xx}^{\mathrm{f}} - \boldsymbol{R}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{R} \\
&= \boldsymbol{C}_{xx}^{\mathrm{f}} - \boldsymbol{C}_{xx}^{\mathrm{f}}\boldsymbol{H}^{\mathrm{T}}\left(\boldsymbol{H}\boldsymbol{C}_{xx}^{\mathrm{f}}\boldsymbol{H}^{\mathrm{T}} + \boldsymbol{C}_{dd}\right)^{-1}\boldsymbol{H}\boldsymbol{C}_{xx}^{\mathrm{f}}
\end{aligned}
$$

Minimizing solution in the Gaussian case

Kalman filter update equations

$$\boldsymbol{x}^{\mathrm{a}} = \boldsymbol{x}^{\mathrm{f}} + \boldsymbol{C}_{xx}^{\mathrm{f}} \boldsymbol{H}^{\mathrm{T}} \left(\boldsymbol{H} \boldsymbol{C}_{xx}^{\mathrm{f}} \boldsymbol{H}^{\mathrm{T}} + \boldsymbol{C}_{dd} \right)^{-1} \left(\boldsymbol{d} - \boldsymbol{H} \boldsymbol{x}^{\mathrm{f}} \right)$$

$$\boldsymbol{C}_{xx}^{\mathrm{a}} = \boldsymbol{C}_{xx}^{\mathrm{f}} - \boldsymbol{C}_{xx}^{\mathrm{f}} \boldsymbol{H}^{\mathrm{T}} \left(\boldsymbol{H} \boldsymbol{C}_{xx}^{\mathrm{f}} \boldsymbol{H}^{\mathrm{T}} + \boldsymbol{C}_{dd} \right)^{-1} \boldsymbol{H} \boldsymbol{C}_{xx}^{\mathrm{f}}.$$

Kalman gain matrix

The control theory community defines the Kalman Gain Matrix

$$\boldsymbol{K} = \boldsymbol{C}_{xx}^{\mathrm{f}} \boldsymbol{H}^{\mathrm{T}} \left(\boldsymbol{H} \boldsymbol{C}_{xx}^{\mathrm{f}} \boldsymbol{H}^{\mathrm{T}} + \boldsymbol{C}_{dd} \right)^{-1}$$

to obtain a simpler expression of the update:

$$\boldsymbol{x}^{\mathrm{a}} = \boldsymbol{x}^{\mathrm{f}} + \boldsymbol{K}\left(\boldsymbol{d} - \boldsymbol{H}\boldsymbol{x}^{\mathrm{f}} \right)$$
$$\boldsymbol{C}_{xx}^{\mathrm{a}} = \boldsymbol{C}_{xx}^{\mathrm{f}} - \boldsymbol{K}\boldsymbol{H}\boldsymbol{C}_{xx}^{\mathrm{f}}$$

"Representer" formulation

$$\begin{aligned} \boldsymbol{x}^{\mathrm{a}} &= \boldsymbol{x}^{\mathrm{f}} + \boldsymbol{C}_{xx}^{\mathrm{f}} \boldsymbol{H}^{\mathrm{T}} \left(\boldsymbol{H} \boldsymbol{C}_{xx}^{\mathrm{f}} \boldsymbol{H}^{\mathrm{T}} + \boldsymbol{C}_{dd} \right)^{-1} \left(\boldsymbol{d} - \boldsymbol{H} \boldsymbol{x}^{\mathrm{f}} \right) \\ &= \boldsymbol{x}^{\mathrm{f}} + \boldsymbol{R}^{\mathrm{T}} \boldsymbol{C}^{-1} \boldsymbol{h} \end{aligned}$$

Define $\boldsymbol{b} = \boldsymbol{C}^{-1}\boldsymbol{h}$ as the solution of the linear system

$$\left(\boldsymbol{H} \boldsymbol{C}_{xx}^{\mathrm{f}} \boldsymbol{H}^{\mathrm{T}} + \boldsymbol{C}_{dd} \right) \boldsymbol{b} = \left(\boldsymbol{d} - \boldsymbol{H} \boldsymbol{x}^{\mathrm{f}} \right)$$

So we can write the update as

$$\boldsymbol{x}^{\mathrm{a}} = \boldsymbol{x}^{\mathrm{f}} + \boldsymbol{R}^{\mathrm{T}} \boldsymbol{b}$$

$$\boldsymbol{C}_{xx}^{\mathrm{a}} = \boldsymbol{C}_{xx}^{\mathrm{f}} - \boldsymbol{R}^{\mathrm{T}} \boldsymbol{C}^{-1} \boldsymbol{R}$$

Note that the covariance update is independent of the actual measurement values.

Summary

- Introduced the concept of a measurement operator.
- Derived Kalman filter update equations in the vector case.
- We update both the state estimate and its error covariance.
- We defined the Kalman Gain.
- We heard about the Representer formulation.
- Original minimization problem is of dimension n.
- The KF update reduces the dimension to $m \ll n$.

Sequential and Smoother solutions from Bayes

Geir Evensen

Starting from Bayes (again)

$$f(\boldsymbol{x}|\boldsymbol{d}) \propto f(\boldsymbol{x})f(\boldsymbol{d}|\boldsymbol{x})$$

- $f(\boldsymbol{x})$ is the density for the model state in space and time.
- $f(\boldsymbol{d}|\boldsymbol{x})$ is the measurement likelihood.

Assume that the model is 1st order Markov process

$$x_i = g(x_{i-1}) + q_i, \qquad q_i \leftarrow f(x_i|x_{i-1})$$

- Since the solution x_i only depends on x_{i-1} we can write

$$f(\boldsymbol{x}) = f(x_0, x_1, \ldots, x_k) = f(x_0) \prod_{i=1}^{k} f(x_i|x_{i-1}).$$

- Valid for most numerical prediction models.

Assume independent data in time

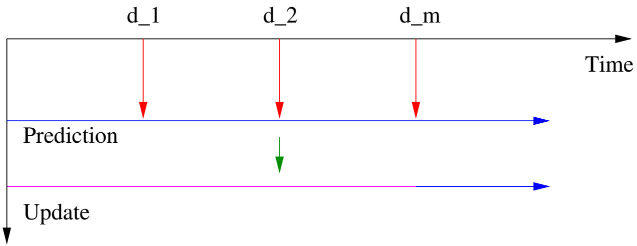
- When measurement errors are uncorrelated in time

$$f(\boldsymbol{d}|\boldsymbol{x}) = \prod_{j=1}^{m} f(d_j|x_j).$$

- Allows for recursive processing of measurements in time

Bayes' becomes

$$f(x_0, x_1, \ldots, x_k, |\boldsymbol{d}) \propto f(x_0) \prod_{i=1}^{k} f(x_i | x_{i-1}) \prod_{j=1}^{m} f(d_j | x_j)$$



Rewrite as

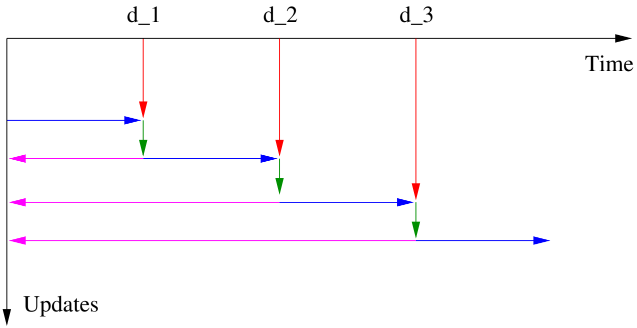
$$
\begin{aligned}
f(\boldsymbol{x}|\boldsymbol{d}) \propto f(\boldsymbol{x}) f(\boldsymbol{d}|\boldsymbol{x}) = \\
& f(x_0) \\
& f(x_1|x_0) f(d_1|x_1) \\
& f(x_2|x_1) f(d_2|x_2) \\
& \vdots \\
& f(x_k|x_{k-1}) f(d_m|x_m) \\
& f(x_{k+1}|x_k)
\end{aligned}
$$

Recursive “smoother” updates

The recursive idea: "Today’s posterior is tomorrow’s prior"

$$f(x_0, x_1 | d_1) = f(x_0)\, f(x_1 | x_0)\, f(d_1 | x_1)$$

$$f(x_0, x_1, x_2 | d_1, d_2) = f(x_0, x_1 | d_1)\, f(x_2 | x_1)\, f(d_2 | x_2)$$

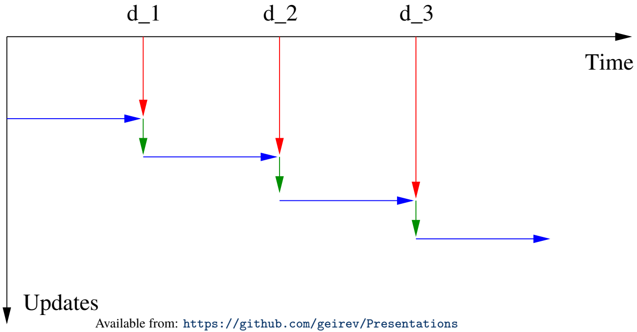


Recursive "filter" updates

Marginal pdfs

$$f(x_1|d_1) = \int_{x_0} f(x_0)f(x_1|x_0)dx_0\, f(d_1|x_1) \qquad = f(x_1)f(d_1|x_1)$$

$$f(x_2|d_1,d_2) = \int_{x_1} f(x_1|d_1)f(x_2|x_1)dx_1\, f(d_2|x_2) \qquad = f(x_2|d_1)f(d_2|x_2)$$



Summary

- Assume model is Marov process.
- Assume measurements are independent in time. (Not generally true.)
- We can process independent measurements sequentially in time.
- The solution of one sub-problem is prior for the next one.

Kalman Filter and Extended Kalman Filter

Geir Evensen

Kalman Filter

- Recursively updates model state and uncertainty.
- Variance minimizing update step.
- Estimate improves and uncertainty reduces at each update.

Error propagation

Derivation for linear scalar model

- Evolution of true state

$$x_k^{\mathrm{t}} = G x_{k-1}^{\mathrm{t}} + q_{k-1}$$

- The model state evolves according to

$$x_k^{\mathrm{f}} = G x_{k-1}^{\mathrm{a}}$$

- Difference is

$$x_k^{\mathrm{t}} - x_k^{\mathrm{f}} = G(x_{k-1}^{\mathrm{t}} - x_{k-1}^{\mathrm{a}}) + q_{k-1}$$

Predicting the errors

Square difference and take the expectation

$$\overline{(x_k^{\mathrm{t}} - x_k^{\mathrm{f}})^2} = G\overline{(x_{k-1}^{\mathrm{t}} - x_{k-1}^{\mathrm{a}})^2}G + \overline{q_{k-1}^2} + 2G\overline{(x_{k-1}^{\mathrm{t}} - x_{k-1}^{\mathrm{a}})q_{k-1}}$$

Error covariance evolution equation

$$C_{xx}^{\mathrm{f}}(t_k) = GC_{xx}^{\mathrm{a}}(t_{k-1})G + C_{qq}(t_{k-1}).$$

- Model errors uncorrelated with state error.

The full Kalman Filter (vector form)

Linear model prediction

$$\boldsymbol{x}_k^{\mathrm{f}} = \boldsymbol{G}\boldsymbol{x}_{k-1}^{\mathrm{a}}$$

Error covariance prediction

$$\boldsymbol{C}_{xx}^{\mathrm{f}}(t_k) = \boldsymbol{G}\boldsymbol{C}_{xx}^{\mathrm{a}}(t_{k-1})\boldsymbol{G}^{\mathrm{T}} + \boldsymbol{C}_{qq}(t_{k-1}).$$

Analysis update (skipped t_k index)

$$\boldsymbol{x}_k^{\mathrm{a}} = \boldsymbol{x}_k^{\mathrm{f}} + \boldsymbol{C}_{xx}^{\mathrm{f}}\boldsymbol{H}^{\mathrm{T}}\left(\boldsymbol{H}\boldsymbol{C}_{xx}^{\mathrm{f}}\boldsymbol{H}^{\mathrm{T}} + \boldsymbol{C}_{dd}\right)^{-1}\left(\boldsymbol{d}_k - \boldsymbol{H}\boldsymbol{x}_k^{\mathrm{f}}\right)$$

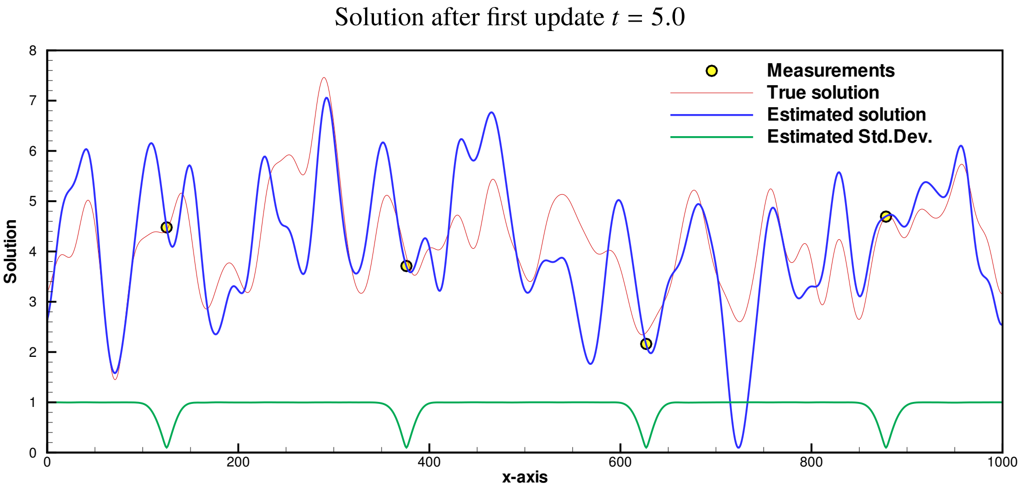
Error covariance update (for each t_k)

$$\boldsymbol{C}_{xx}^{\mathrm{a}} = \boldsymbol{C}_{xx}^{\mathrm{f}} - \boldsymbol{C}_{xx}^{\mathrm{f}}\boldsymbol{H}^{\mathrm{T}}\left(\boldsymbol{H}\boldsymbol{C}_{xx}^{\mathrm{f}}\boldsymbol{H}^{\mathrm{T}} + \boldsymbol{C}_{dd}\right)^{-1}\boldsymbol{H}\boldsymbol{C}_{xx}^{\mathrm{f}}.$$

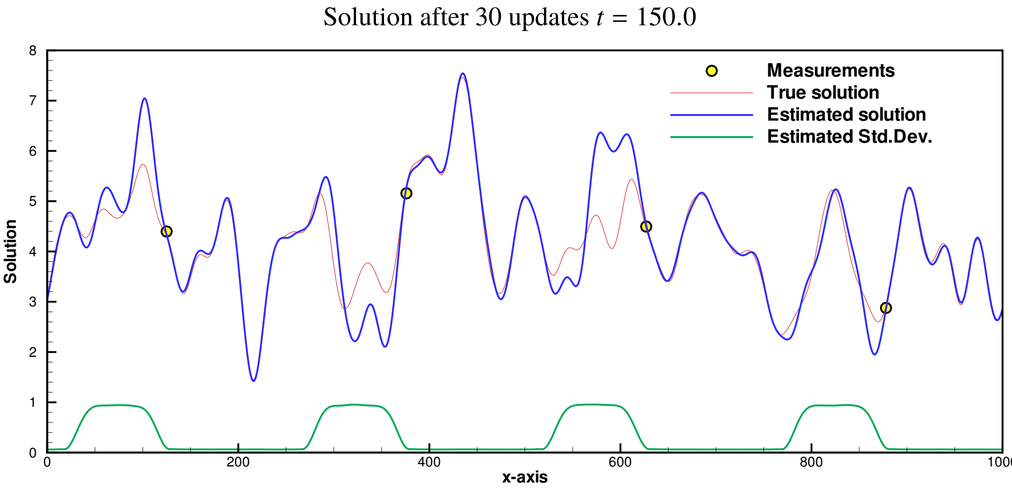
Kalman Filter Example

- Linear advection equation
- Periodic domain
- Random reference solution (truth).
- First guess is reference plus random perturbation.
- Initial variance is 1.0 m^2
- Four measurements every 5 time units.
- Measurement variance is 0.01 m^2.
- Cases without and including system noise of 0.0004 m^2.

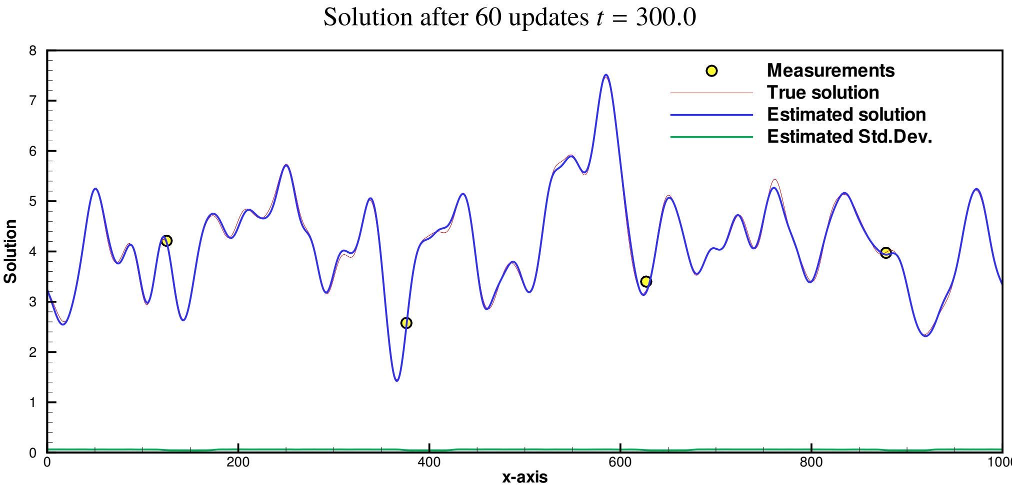
Kalman Filter example: Perfect model



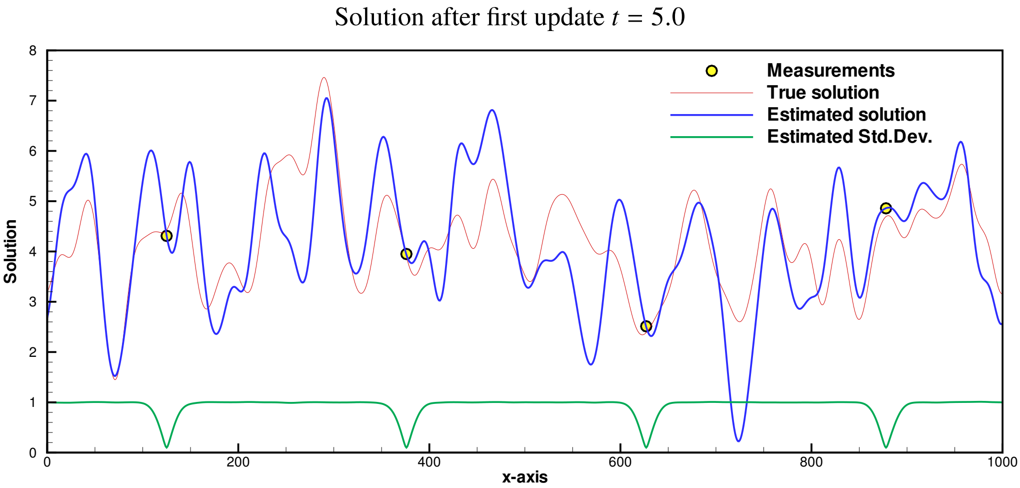
Kalman Filter example: Perfect model



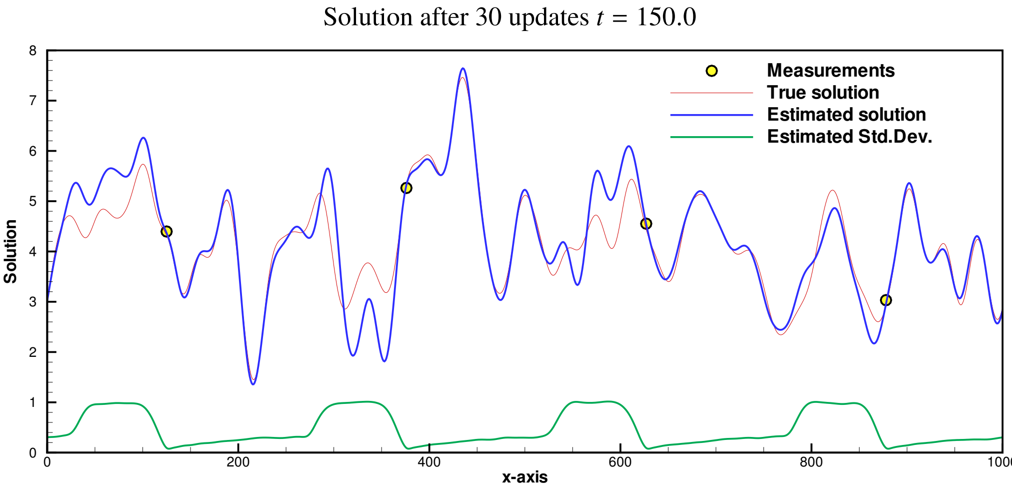
Kalman Filter example: Perfect model



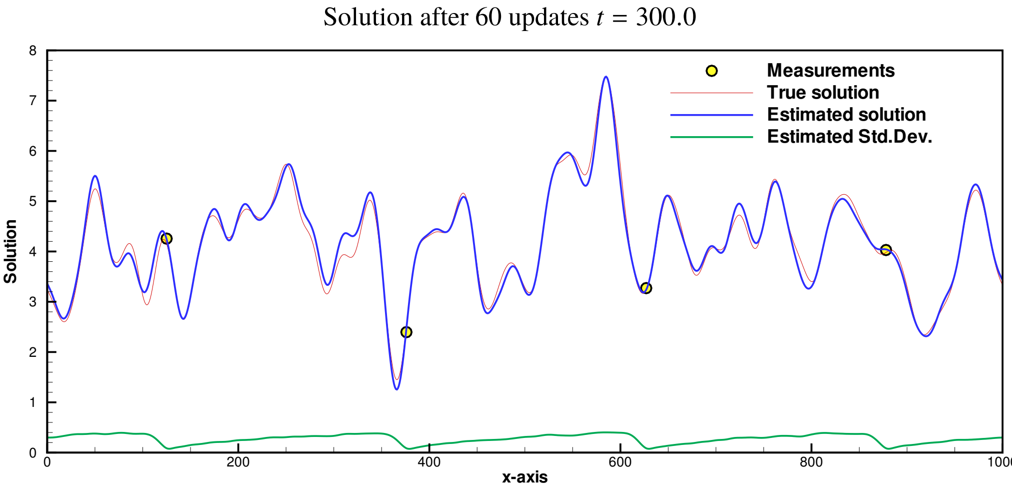
Kalman Filter example: with model error



Kalman Filter example: with model error



Kalman Filter example: with model error



Available from: `https://github.com/geirev/Presentations`

Kalman Filter example: with model error

Animation

Inverse problem revisited

What is the KF solution for the linear inverse problem?

$$
\begin{aligned}
\frac{dx}{dt} &= 1 \\
x(0) &= 0 \\
d = x(1) &= 2
\end{aligned}
$$

KF solution

Solve initial value problem

$$\frac{dx}{dt} = 1$$
$$x(0) = 0$$

$$\Longrightarrow x^{\mathrm{f}}(t) = t$$

Predicted error variance

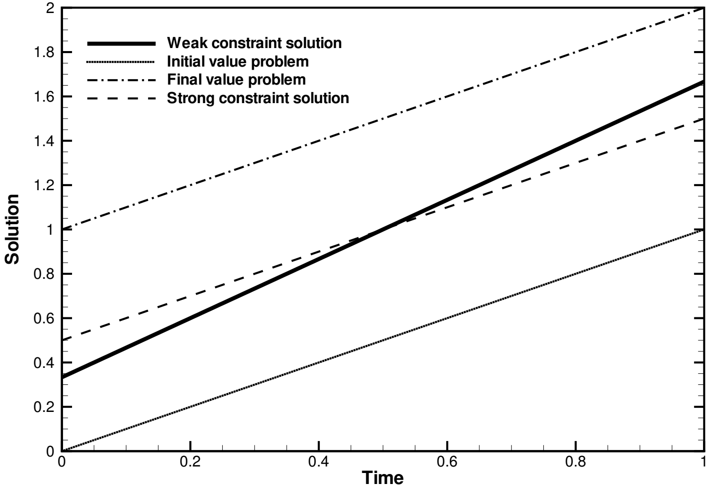
$$C^{\mathrm{f}}_{xx}(1) = C^{\mathrm{a}}_{xx}(0) + C_{qq} = 2C_0$$

Update at $t = 1$ is

$$x^{\mathrm{a}} = x^{\mathrm{f}} + \frac{C^{\mathrm{f}}_{xx}}{C_{dd} + C^{\mathrm{f}}_{xx}}(d - x^{\mathrm{f}}) = 1 + \frac{2C_0}{C_0 + 2C_0}(2 - 1) = 5/3$$

KF solution at final time equals the weak-constraint variational solution

Inverse problem revisited



Nonlinear dynamics

Derivation of Extended Kalman Filter (EKF)

$$
\begin{aligned}
\boldsymbol{x}_k^{\mathrm{t}} &= \boldsymbol{g}(\boldsymbol{x}_{k-1}^{\mathrm{t}}) + \boldsymbol{q}_{k-1}, \\
\boldsymbol{x}_k^{\mathrm{f}} &= \boldsymbol{g}(\boldsymbol{x}_{k-1}^{\mathrm{a}}), \\
\boldsymbol{x}_k^{\mathrm{t}} - \boldsymbol{x}_k^{\mathrm{f}} &= \boldsymbol{g}(\boldsymbol{x}_{k-1}^{\mathrm{t}}) - \boldsymbol{g}(\boldsymbol{x}_{k-1}^{\mathrm{a}}) + \boldsymbol{q}_{k-1}.
\end{aligned}
$$

Use Taylor expansion

$$
\boldsymbol{g}(\boldsymbol{x}_{k-1}^{\mathrm{t}}) = \boldsymbol{g}(\boldsymbol{x}_{k-1}^{\mathrm{a}}) + \boldsymbol{G}(\boldsymbol{x}_{k-1}^{\mathrm{a}})(\boldsymbol{x}_{k-1}^{\mathrm{t}} - \boldsymbol{x}_{k-1}^{\mathrm{a}}) + \frac{1}{2}\boldsymbol{\mathcal{H}}(\boldsymbol{x}_{k-1}^{\mathrm{a}})(\boldsymbol{x}_{k-1}^{\mathrm{t}} - \boldsymbol{x}_{k-1}^{\mathrm{a}})^2 + \cdots .
$$

EKF: Derivation

Difference becomes

$$\boldsymbol{x}_k^{\mathrm{t}} - \boldsymbol{x}_k^{\mathrm{f}} = \boldsymbol{G}(\boldsymbol{x}_{k-1}^{\mathrm{a}})(\boldsymbol{x}_{k-1}^{\mathrm{t}} - \boldsymbol{x}_{k-1}^{\mathrm{a}}) + \frac{1}{2}\boldsymbol{\mathcal{H}}(\boldsymbol{x}_{k-1}^{\mathrm{a}})(\boldsymbol{x}_{k-1}^{\mathrm{t}} - \boldsymbol{x}_{k-1}^{\mathrm{a}})^2 + \cdots + \boldsymbol{q}_{k-1}.$$

By squaring and taking the expectation we get

$$\begin{aligned}
\boldsymbol{C}_{xx}^{\mathrm{f}}(t_k) &= \overline{(\boldsymbol{x}_k^{\mathrm{t}} - \boldsymbol{x}_k^{\mathrm{f}})(\boldsymbol{x}_k^{\mathrm{t}} - \boldsymbol{x}_k^{\mathrm{f}})^{\mathrm{T}}} \\
&= \boldsymbol{G}(\boldsymbol{x}_{k-1}^{\mathrm{a}})\overline{(\boldsymbol{x}_{k-1}^{\mathrm{t}} - \boldsymbol{x}_{k-1}^{\mathrm{a}})(\boldsymbol{x}_{k-1}^{\mathrm{t}} - \boldsymbol{x}_{k-1}^{\mathrm{a}})^{\mathrm{T}}}(\boldsymbol{G}(\boldsymbol{x}_{k-1}^{\mathrm{a}}))^{\mathrm{T}} + \cdots + \boldsymbol{C}_{qq}(t_{k-1}) \\
&= \boldsymbol{G}(\boldsymbol{x}_{k-1}^{\mathrm{a}})\boldsymbol{C}_{xx}^{\mathrm{a}}(t_{k-1})(\boldsymbol{G}(\boldsymbol{x}_{k-1}^{\mathrm{a}}))^{\mathrm{T}} + \cdots + \boldsymbol{C}_{qq}(t_{k-1}).
\end{aligned}$$

EKF: Error evolution

Close by discarding higher order moments to get

$$
\begin{aligned}
\boldsymbol{x}_k^{\mathrm{f}} &= \boldsymbol{g}(\boldsymbol{x}_{k-1}^{\mathrm{a}}), \\
\boldsymbol{C}_{xx}^{\mathrm{f}}(t_k) &\simeq \boldsymbol{G}_{k-1} \boldsymbol{C}_{xx}^{\mathrm{a}}(t_{k-1}) \boldsymbol{G}_{k-1}^{\mathrm{T}} + \boldsymbol{C}_{qq}(t_{k-1}),
\end{aligned}
$$

together with standard analysis equations.

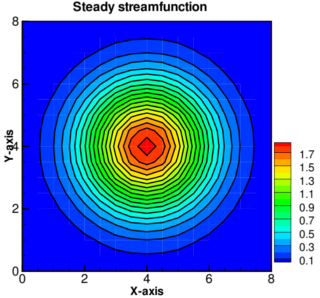
Example of Extended KF

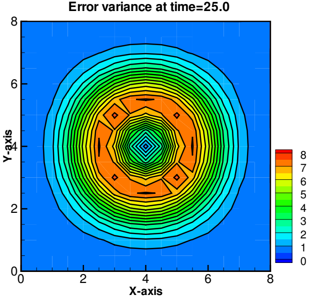
Nonlinear quasi-geostrophic model

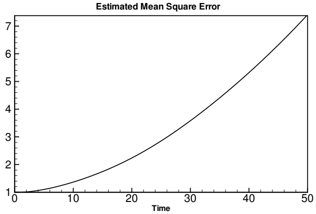
- Steady stream function solution.
- Curved and sheared flow.
- Supports instability.
- Initial variance is 1.0.

Example of Extended KF

Results (from Evensen 1992):







- Linear closure approximation not valid!
- Leads to linear instability and exponential error growth.

EKF: Summary

KF is optimal linear filter method!

- Storage of $\mathcal{O}(n^2)$ elements.
- Integration of $2n$ models.

EKF applies closure approximation in error covariance equation.

- Requires implementation of tangent linear model.
- Too simple closure may lead to linear instabilities.
- Does not work for strongly nonlinear models.

We need a better alternative!

Ensemble Kalman Filter

A Monte Carlo alternative to KF and EKF

Geir Evensen

The error covariance matrix

Define ensemble covariances around the ensemble mean

$$\begin{aligned}
\boldsymbol{C}^{\mathrm{f}}_{xx} &\simeq \overline{\boldsymbol{C}}^{\mathrm{f}}_{xx} = \overline{(\boldsymbol{x}^{\mathrm{f}} - \overline{\boldsymbol{x}^{\mathrm{f}}})(\boldsymbol{x}^{\mathrm{f}} - \overline{\boldsymbol{x}^{\mathrm{f}}})^{\mathrm{T}}} \\
\boldsymbol{C}^{\mathrm{a}}_{xx} &\simeq \overline{\boldsymbol{C}}^{\mathrm{a}}_{xx} = \overline{(\boldsymbol{x}^{\mathrm{a}} - \overline{\boldsymbol{x}^{\mathrm{a}}})(\boldsymbol{x}^{\mathrm{a}} - \overline{\boldsymbol{x}^{\mathrm{a}}})^{\mathrm{T}}}
\end{aligned}$$

- The ensemble mean $\overline{\boldsymbol{x}}$ is the best-guess.
- The ensemble spread defines the error variance.
- The ensemble smoothness defines the error covariance.

Dynamical evolution of error statistics

- Ensemble of models (particles) defines probability $f(\boldsymbol{x})$.
- Ensemble members evolve according to the model dynamics.

$$d\boldsymbol{x} = \boldsymbol{g}(\boldsymbol{x})dt + d\boldsymbol{q}.$$

- Probability density evolve according to Kolmogorov's equation.

$$\frac{\partial f}{\partial t} + \sum_i \frac{\partial (g_i f)}{\partial x_i} = \frac{1}{2} \sum_{i,j} \boldsymbol{C}_{qq} \frac{\partial^2 f}{\partial x_i \partial x_j}.$$

- Fundamental equation for evolution of error statistics.
- Integrating a large ensemble of stochastic models is a MC method for solving Kolmogorov's equation

Analysis scheme (1)

Define the model-forcast error-covariance matrix

$$\boldsymbol{C}_{xx}^{\mathrm{f}} \simeq \overline{\boldsymbol{C}}_{xx}^{\mathrm{f}} = \overline{(\boldsymbol{x}^{\mathrm{f}} - \overline{\boldsymbol{x}^{\mathrm{f}}})(\boldsymbol{x}^{\mathrm{f}} - \overline{\boldsymbol{x}^{\mathrm{f}}})^{\mathrm{T}}}.$$

and the measurement error-covariance matrix

$$\boldsymbol{C}_{dd} \simeq \overline{\boldsymbol{C}}_{dd} = \overline{\boldsymbol{\epsilon}\boldsymbol{\epsilon}^{\mathrm{T}}}.$$

Create an ensemble of perturbed observations

$$\boldsymbol{d}_j = \boldsymbol{d} + \boldsymbol{\epsilon}_j,$$

where $\boldsymbol{d}$ is the vector of observed values and $\boldsymbol{\epsilon}_j$, is a vector of observation noise.

Analysis scheme (2)

Update each ensemble member according to

$$\boxed{\begin{aligned} \boldsymbol{x}_j^{\mathrm{a}} &= \boldsymbol{x}_j^{\mathrm{f}} + \overline{\boldsymbol{C}}_{xx}^{\mathrm{f}} \boldsymbol{H}^{\mathrm{T}} \left(\boldsymbol{H} \overline{\boldsymbol{C}}_{xx}^{\mathrm{f}} \boldsymbol{H}^{\mathrm{T}} + \overline{\boldsymbol{C}}_{dd} \right)^{-1} \left(\boldsymbol{d}_j - \boldsymbol{H} \boldsymbol{x}_j^{\mathrm{f}} \right) \\ &= \boldsymbol{x}_j^{\mathrm{f}} + \overline{\boldsymbol{K}} \left(\boldsymbol{d}_j - \boldsymbol{H} \boldsymbol{x}_j^{\mathrm{f}} \right) \end{aligned}}$$

Thus, the update of the mean becomes

$$\begin{aligned} \overline{\boldsymbol{x}^{\mathrm{a}}} &= \overline{\boldsymbol{x}^{\mathrm{f}}} + \overline{\boldsymbol{C}}_{xx}^{\mathrm{f}} \boldsymbol{H}^{\mathrm{T}} \left(\boldsymbol{H} \overline{\boldsymbol{C}}_{xx}^{\mathrm{f}} \boldsymbol{H}^{\mathrm{T}} + \overline{\boldsymbol{C}}_{dd} \right)^{-1} \left(\boldsymbol{d} - \boldsymbol{H} \overline{\boldsymbol{x}^{\mathrm{f}}} \right) \\ &= \overline{\boldsymbol{x}^{\mathrm{f}}} + \overline{\boldsymbol{K}} \left(\boldsymbol{d} - \boldsymbol{H} \overline{\boldsymbol{x}^{\mathrm{f}}} \right) \end{aligned}$$

Analysis scheme (3)

The error covariance update then becomes

$$
\begin{aligned}
\overline{\boldsymbol{C}}^{\mathrm{a}}_{xx} &= \overline{\left(\boldsymbol{x}^{\mathrm{a}} - \overline{\boldsymbol{x}^{\mathrm{a}}}\right)\left(\boldsymbol{x}^{\mathrm{a}} - \overline{\boldsymbol{x}^{\mathrm{a}}}\right)^{\mathrm{T}}} \\
&= \overline{\Big(\big(\boldsymbol{I} - \overline{\boldsymbol{K}}\boldsymbol{H}\big)\big(\boldsymbol{x}^{\mathrm{f}} - \overline{\boldsymbol{x}^{\mathrm{f}}}\big) + \overline{\boldsymbol{K}}\big(\boldsymbol{d} - \overline{\boldsymbol{d}}\big)\Big)\Big(\cdots\Big)^{\mathrm{T}}} \\
&= (\boldsymbol{I} - \overline{\boldsymbol{K}}\boldsymbol{H})\overline{\boldsymbol{C}}^{\mathrm{f}}_{xx}(\boldsymbol{I} - \boldsymbol{H}^{\mathrm{T}}\overline{\boldsymbol{K}}^{\mathrm{T}}) + \overline{\boldsymbol{K}}\,\overline{\boldsymbol{C}}_{dd}\overline{\boldsymbol{K}}^{\mathrm{T}} \\
&= \overline{\boldsymbol{C}}^{\mathrm{f}}_{xx} - \overline{\boldsymbol{K}}\boldsymbol{H}\overline{\boldsymbol{C}}^{\mathrm{f}}_{xx} - \overline{\boldsymbol{C}}^{\mathrm{f}}_{xx}\boldsymbol{H}^{\mathrm{T}}\overline{\boldsymbol{K}}^{\mathrm{T}} + \overline{\boldsymbol{K}}\big(\boldsymbol{H}\overline{\boldsymbol{C}}^{\mathrm{f}}_{xx}\boldsymbol{H}^{\mathrm{T}} + \overline{\boldsymbol{C}}_{dd}\big)\overline{\boldsymbol{K}}^{\mathrm{T}} \\
&= (\boldsymbol{I} - \overline{\boldsymbol{K}}\boldsymbol{H})\overline{\boldsymbol{C}}^{\mathrm{f}}_{xx} \\
&= \overline{\boldsymbol{C}}^{\mathrm{f}}_{xx} - \overline{\boldsymbol{C}}^{\mathrm{f}}_{xx}\boldsymbol{H}^{\mathrm{T}}\left(\boldsymbol{H}\overline{\boldsymbol{C}}^{\mathrm{f}}_{xx}\boldsymbol{H}^{\mathrm{T}} + \overline{\boldsymbol{C}}_{dd}\right)^{-1}\boldsymbol{H}\overline{\boldsymbol{C}}^{\mathrm{f}}_{xx}
\end{aligned}
$$

Note that we need to perturb observations to have $\overline{\boldsymbol{C}}_{dd} = \overline{(\boldsymbol{d} - \overline{\boldsymbol{d}})(\boldsymbol{d} - \overline{\boldsymbol{d}})^{\mathrm{T}}}$ (Burgers et al., 1998)

Ensemble Kalman Filter (EnKF)

- Represents error statistics using an ensemble of model states.
- Evolves error statistics by ensemble integrations.
- “Variance minimizing” analysis scheme operating on the ensemble.

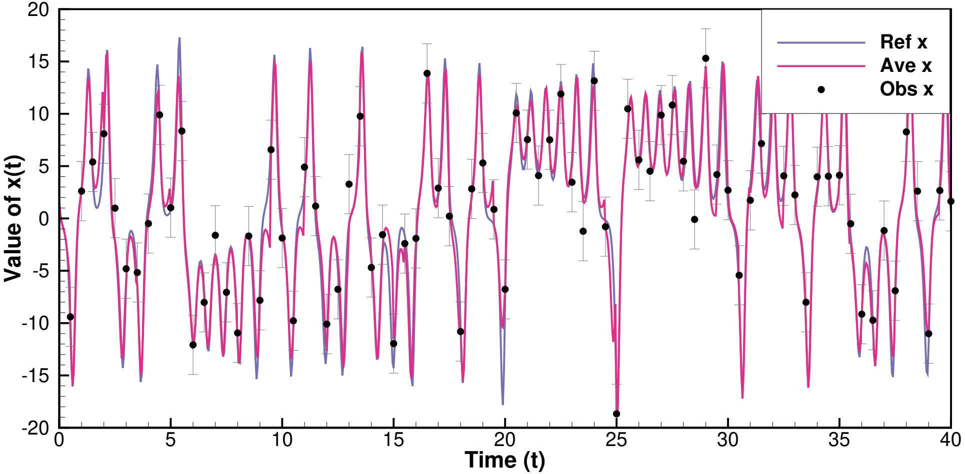
$$\Downarrow$$

- Monte Carlo, low rank, ensemble subspace method.
- Linear model: EnKF converges to the KF with large ensemble size.
- Fully nonlinear error evolution, contrary to EKF.
- Assumption of Gaussian statistics in analysis scheme.

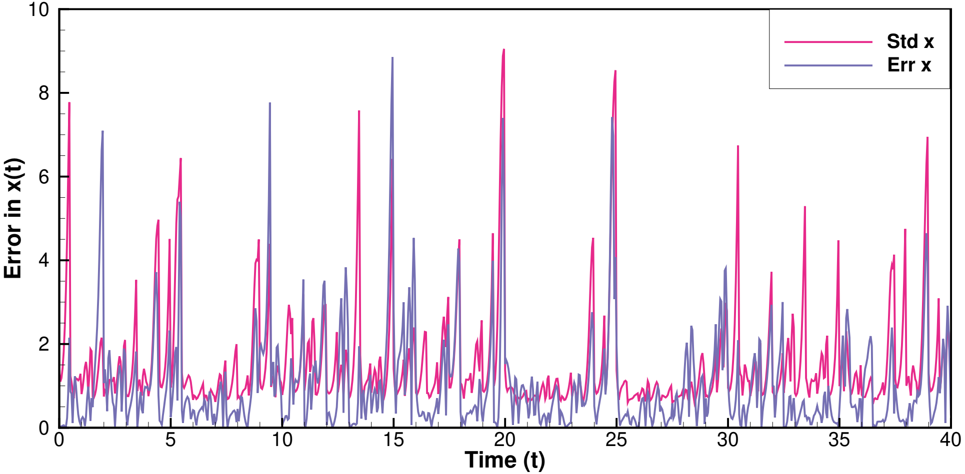
Example: Lorenz model

- Application with the chaotic Lorenz model.
- Illustrates properties with higly nonlinear dynamical models.
- From Evensen (1997), MWR.

EnKF solution



EnKF error variance



Analysis equation (1)

- Define the ensemble matrix
$$\boldsymbol{A} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N) \in \Re^{n \times N}.$$
- The ensemble mean is (defining $\mathbf{1}_N \in \Re^{N \times N} \equiv 1/N$)
$$\overline{\boldsymbol{A}} = \boldsymbol{A}\mathbf{1}_N.$$
- The ensemble perturbations become
$$\boldsymbol{A}' = \boldsymbol{A} - \overline{\boldsymbol{A}} = \boldsymbol{A}(\boldsymbol{I} - \mathbf{1}_N).$$
- The ensemble covariance matrix $\overline{\boldsymbol{C}}_{xx} \in \Re^{n \times n}$ becomes
$$\overline{\boldsymbol{C}}_{xx} = \frac{\boldsymbol{A}'(\boldsymbol{A}')^{\mathrm{T}}}{N-1}.$$

Analysis equation (2)

- Given a vector of measurements $\boldsymbol{d} \in \Re^m$, define

$$\boldsymbol{d}_j = \boldsymbol{d} + \boldsymbol{\epsilon}_j, \qquad j = 1, \ldots, N,$$

  stored in

$$\boldsymbol{D} = (\boldsymbol{d}_1, \boldsymbol{d}_2, \ldots, \boldsymbol{d}_N) \in \Re^{m \times N}.$$

- The ensemble perturbations are stored in

$$\boldsymbol{E} = (\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2, \ldots, \boldsymbol{\epsilon}_N) \in \Re^{m \times N},$$

  thus, the measurement error covariance matrix becomes

$$\overline{\boldsymbol{C}}_{dd} = \frac{\boldsymbol{E}\boldsymbol{E}^{\mathrm{T}}}{N-1}.$$

Analysis equation (3)

The analysis equation

$$\boldsymbol{A}^{\mathrm{a}} = \boldsymbol{A} + \overline{\boldsymbol{C}}_{xx}\boldsymbol{H}^{\mathrm{T}}\left(\boldsymbol{H}\overline{\boldsymbol{C}}_{xx}\boldsymbol{H}^{\mathrm{T}} + \overline{\boldsymbol{C}}_{dd}\right)^{-1}\left(\boldsymbol{D} - \boldsymbol{H}\boldsymbol{A}\right).$$

can now be written

$$\boldsymbol{A}^{\mathrm{a}} = \boldsymbol{A} + \boldsymbol{A}'(\boldsymbol{H}\boldsymbol{A}')^{\mathrm{T}}\left((\boldsymbol{H}\boldsymbol{A}')(\boldsymbol{H}\boldsymbol{A}')^{\mathrm{T}} + \boldsymbol{E}\boldsymbol{E}^{\mathrm{T}}\right)^{-1}\left(\boldsymbol{D} - \boldsymbol{H}\boldsymbol{A}\right).$$

The update is expressed entirely in terms of the ensemble

Define $\boldsymbol{S} = \boldsymbol{H}\boldsymbol{A}'$

$$\boldsymbol{A}^{\mathrm{a}} = \boldsymbol{A} + \boldsymbol{A}'\boldsymbol{S}^{\mathrm{T}}\left(\boldsymbol{S}\boldsymbol{S}^{\mathrm{T}} + \boldsymbol{E}\boldsymbol{E}^{\mathrm{T}}\right)^{-1}\left(\boldsymbol{D} - \boldsymbol{H}\boldsymbol{A}\right).$$

Analysis equation (4)

Define $\boldsymbol{C} = \boldsymbol{S}\boldsymbol{S}^{\mathrm{T}} + \boldsymbol{E}\boldsymbol{E}^{\mathrm{T}}$ and the innovations $\boldsymbol{D}' = \boldsymbol{D} - \boldsymbol{H}\boldsymbol{A}$.

$$
\begin{aligned}
\boldsymbol{A}^{\mathrm{a}} &= \boldsymbol{A} + \boldsymbol{A}'\boldsymbol{S}^{\mathrm{T}}\left(\boldsymbol{S}\boldsymbol{S}^{\mathrm{T}} + \boldsymbol{E}\boldsymbol{E}^{\mathrm{T}}\right)^{-1}\left(\boldsymbol{D} - \boldsymbol{H}\boldsymbol{A}\right). \\
&= \boldsymbol{A} + \boldsymbol{A}'\boldsymbol{S}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{D}' \\
&= \boldsymbol{A} + \boldsymbol{A}(\boldsymbol{I} - \boldsymbol{1}_N)\boldsymbol{S}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{D}' \\
&= \boldsymbol{A}\left(\boldsymbol{I} + \boldsymbol{S}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{D}'\right) \\
&= \boldsymbol{A}\boldsymbol{X}
\end{aligned}
$$

where we have used

- $\boldsymbol{A}' = \boldsymbol{A}(\boldsymbol{I} - \boldsymbol{1}_N)$.
- $\boldsymbol{1}_N\boldsymbol{S}^{\mathrm{T}} \equiv \boldsymbol{0}$.

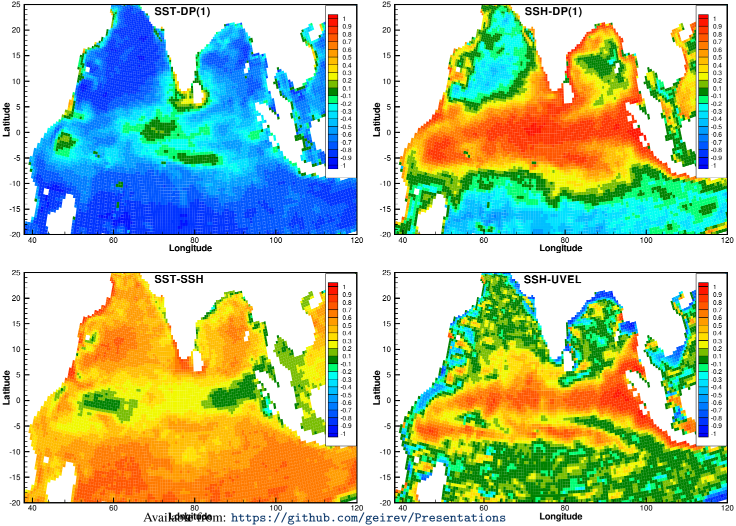
Remarks

- $\overline{\boldsymbol{C}}_{xx}$ is never computed.
- Even $\boldsymbol{H}\overline{\boldsymbol{C}}_{xx}\boldsymbol{H}^{\mathrm{T}} = \boldsymbol{S}\boldsymbol{S}^{\mathrm{T}}$ need not be computed.
- Analysis may be interpreted as:
  - combination of forecast ensemble members, or,
  - forecast plus combination of covariance functions.
- Accuracy of analysis is determined by:
  - the accuracy of $\boldsymbol{X}$,
  - the properties of the ensemble space.
- For a linear model, any choice of $\boldsymbol{X}$ will result in an analysis which is also a solution of the model.

Examples of ensemble statistics

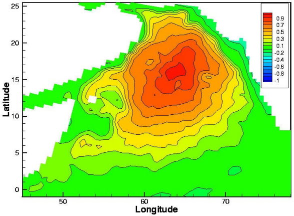
- Taken from Haugen et al. (2002), Ocean Dynamics.
- OGCM (MICOM) for the Indian Ocean.
- Assimilation of SST and SLA data.

Spatial correlations

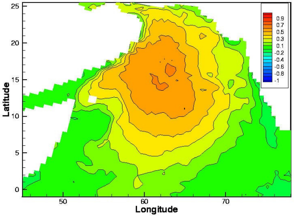


Correlation functions

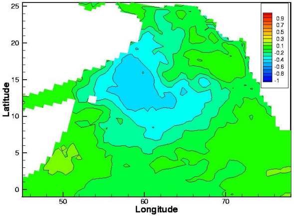
SSH–SSH



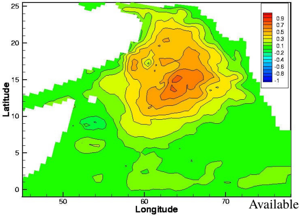
SSH–SST



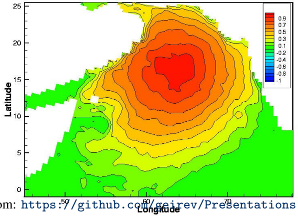
SSH–DP



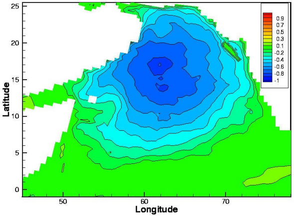
SST–SSH



SST–SST

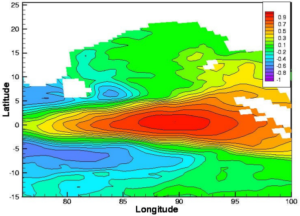


SST–DP

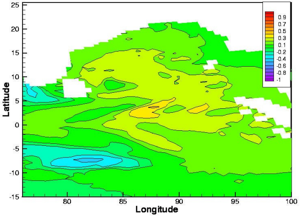


Correlation functions

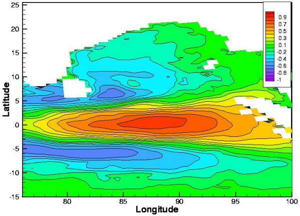
SSH–SSH



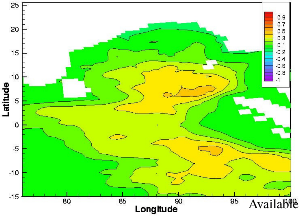
SSH–SST



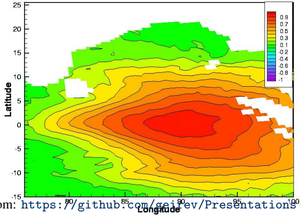
SSH–DP



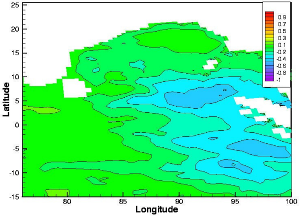
SST–SSH



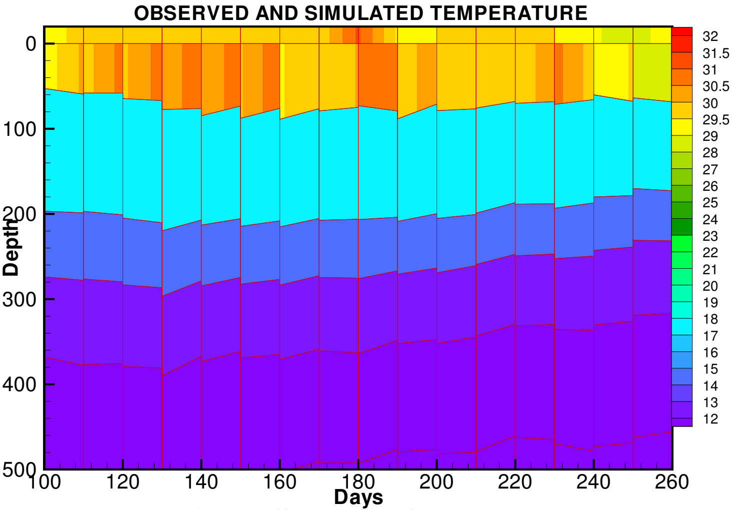
SST–SST



SST–DP



Time–Depth: Temperature



Available from: `https://github.com/geirev/Presentations`

Computational aspects (1)

Analysis scheme

$$\boldsymbol{A}^{\mathrm{a}} = \boldsymbol{A}\boldsymbol{X} = \boldsymbol{A}\left(\boldsymbol{I} + \boldsymbol{S}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{D}'\right)$$

How to compute the inverse

$$\boldsymbol{C}^{-1} = \left(\boldsymbol{S}\boldsymbol{S}^{\mathrm{T}} + \boldsymbol{E}\boldsymbol{E}^{\mathrm{T}}\right)^{-1} \qquad (\approx \boldsymbol{Z}\boldsymbol{\Lambda}^{+}\boldsymbol{Z}^{\mathrm{T}})$$

- Low rank $(N-1)$.
- Dimension is number of measurents m.
- Direct inversion requires $\mathcal{O}(m^3)$ computations.

Subspace inversion using $\boldsymbol{C}_{dd} \approx \boldsymbol{E}\boldsymbol{E}^{\mathrm{T}}$

Measurement errors are more often than not highly correlated.

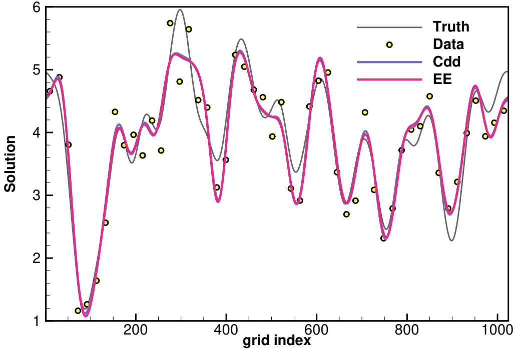
- Algorithm by Evensen (2004) works directly with $\boldsymbol{E}$.

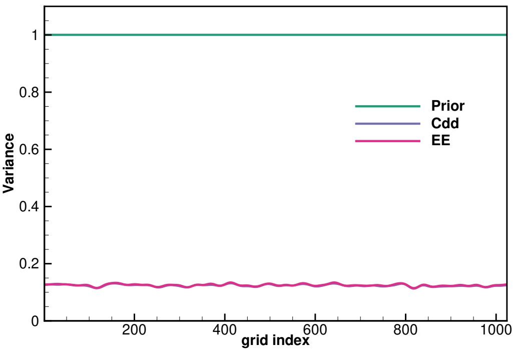
$$
\begin{aligned}
&\left(\boldsymbol{S}\boldsymbol{S}^{\mathrm{T}} + \boldsymbol{E}\boldsymbol{E}^{\mathrm{T}}\right) \\
&\qquad \approx \boldsymbol{S}\boldsymbol{S}^{\mathrm{T}} + (\boldsymbol{S}\boldsymbol{S}^{+})\boldsymbol{E}\boldsymbol{E}^{\mathrm{T}}(\boldsymbol{S}\boldsymbol{S}^{+})^{\mathrm{T}} \\
&\qquad = \boldsymbol{U}\boldsymbol{\Sigma}\left(\boldsymbol{I}_N + \boldsymbol{\Sigma}^{+}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{E}\boldsymbol{E}^{\mathrm{T}}\boldsymbol{U}(\boldsymbol{\Sigma}^{+})^{\mathrm{T}}\right)\boldsymbol{\Sigma}^{\mathrm{T}}\boldsymbol{U}^{\mathrm{T}} \\
&\qquad = \boldsymbol{U}\boldsymbol{\Sigma}\left(\boldsymbol{I}_N + \boldsymbol{Z}\boldsymbol{\Lambda}\boldsymbol{Z}^{\mathrm{T}}\right)\boldsymbol{\Sigma}^{\mathrm{T}}\boldsymbol{U}^{\mathrm{T}} \\
&\qquad = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{Z}(\boldsymbol{I}_N + \boldsymbol{\Lambda})\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\Sigma}^{\mathrm{T}}\boldsymbol{U}^{\mathrm{T}}.
\end{aligned}
$$

$$
\left(\boldsymbol{S}\boldsymbol{S}^{\mathrm{T}} + \boldsymbol{E}\boldsymbol{E}^{\mathrm{T}}\right)^{-1} \approx \boldsymbol{U}(\boldsymbol{\Sigma}^{+})^{\mathrm{T}}\boldsymbol{Z}\left(\boldsymbol{I}_N + \boldsymbol{\Lambda}\right)^{-1}\left(\boldsymbol{U}(\boldsymbol{\Sigma}^{+})^{\mathrm{T}}\boldsymbol{Z}\right)^{\mathrm{T}}
$$

Computational cost is $\mathcal{O}(mN^2)$.

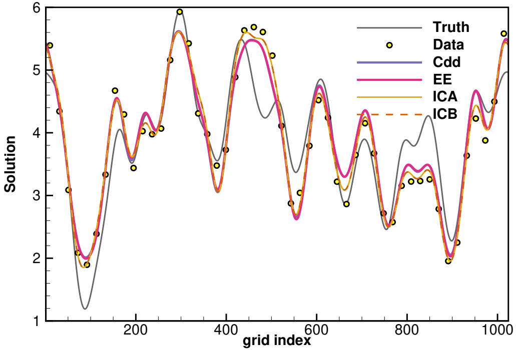
EnKF analysis with uncorrelated measurement errors

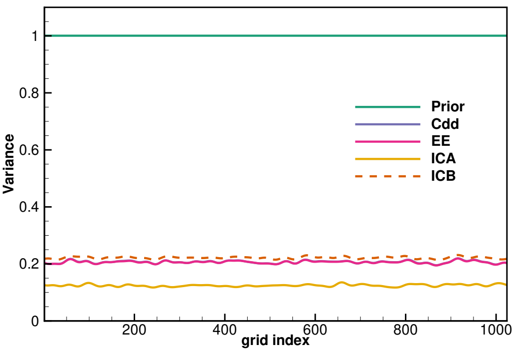




- Ensemble size $N = 2000$.
- Cdd is the solution inverting $\boldsymbol{C} = \boldsymbol{S}\boldsymbol{S}^{\mathrm{T}} + \boldsymbol{C}_{dd}$.
- EE is the subspace inversion using the measurement perturbations $\boldsymbol{E}$.
- Measurement error variance is 0.5.

EnKF analysis with correlated measurement errors

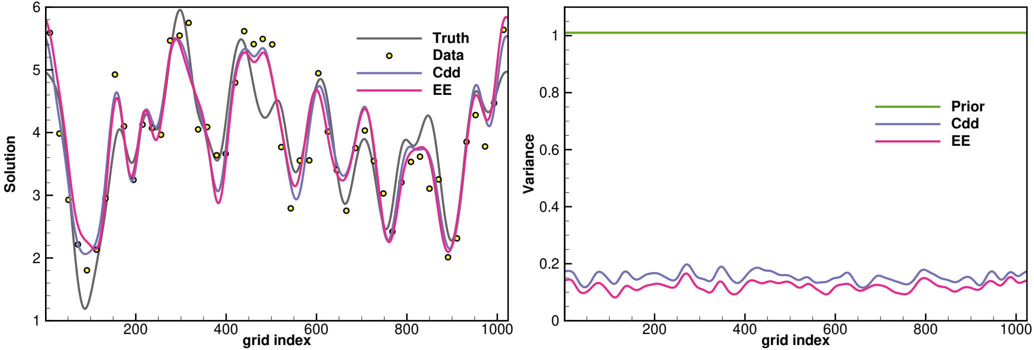




- Ensemble size $N = 2000$.
- Cdd is the solution inverting $\boldsymbol{C} = \boldsymbol{S}\boldsymbol{S}^{\mathrm{T}} + \boldsymbol{C}_{dd}$.
- EE is the subspace inversion using the measurement perturbations $\boldsymbol{E}$.
- ICA is inconsistent update erroneously assuming uncorrelated measurement errors.

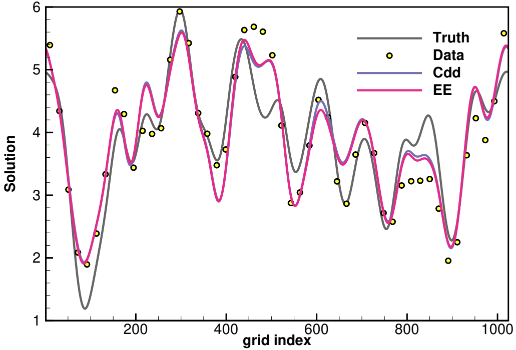
NORCE

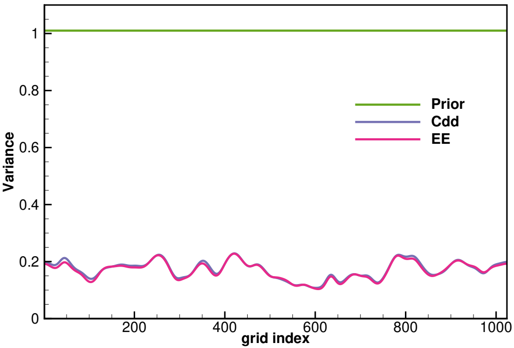
EnKF analysis with correlated measurement errors



- Ensemble size $N = 100$, $\boldsymbol{E} \in \Re^{m \times N}$.
- Measurement error $r_d = 40$ and ensemble $r_d = 40$.
- Using $\boldsymbol{E}$ under-estimates posterior variance.

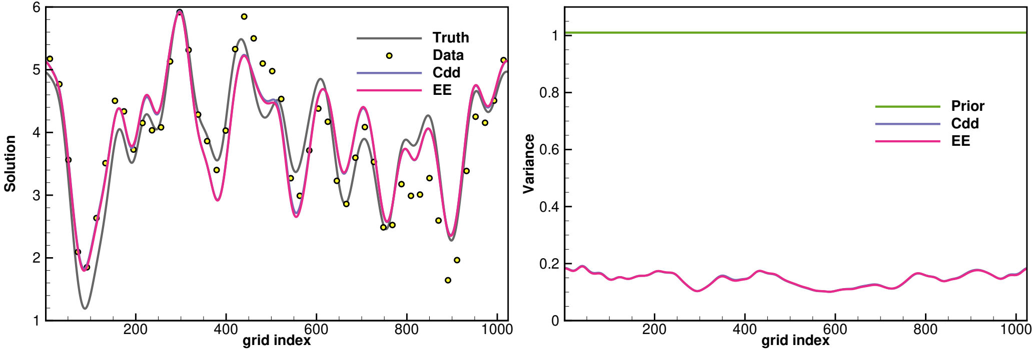
EnKF analysis with correlated smooth measurement errors





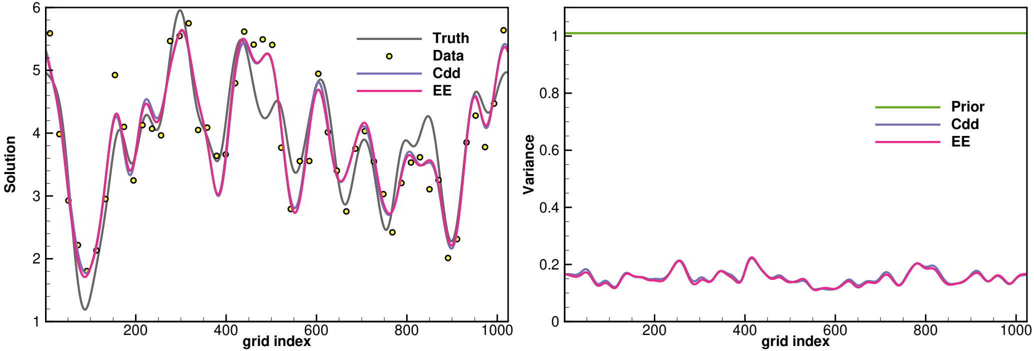
- Ensemble size $N = 100$, $\boldsymbol{E} \in \Re^{m \times 10N}$.
- Measurement error $r_d = 40$ and ensemble $r_d = 40$.
- Using $\boldsymbol{E}$ works perfectly.

EnKF analysis with correlated smooth measurement errors



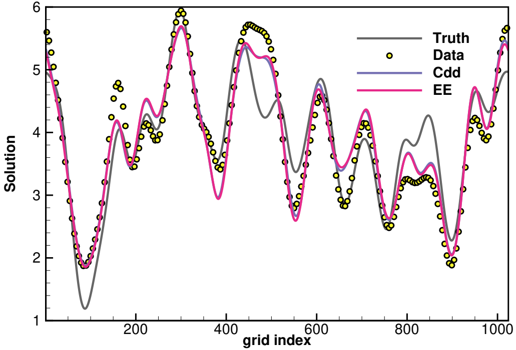
- Ensemble size $N = 100$, $\boldsymbol{E} \in \Re^{m \times 10N}$.
- Measurement error $r_d = 80$ while ensemble $r_d = 40$.
- Using $\boldsymbol{E}$ works perfectly.

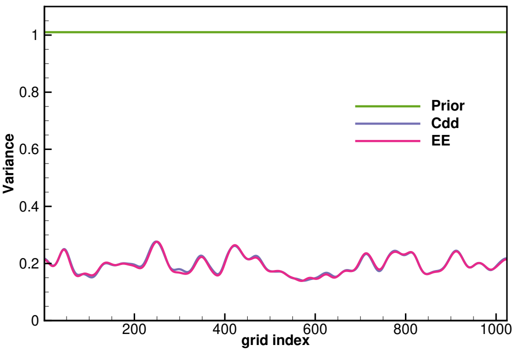
EnKF analysis with correlated measurement errors



- Ensemble size $N = 100$, $\boldsymbol{E} \in \Re^{m \times 10N}$.
- Measurement error $r_d = 20$ while ensemble $r_d = 40$.
- Cannot represent scales in $\boldsymbol{E}$ shorter than $r_d = 40$.

EnKF analysis with many measurements





- Ensemble size $N = 100$, $\boldsymbol{E} \in \Re^{m \times 10N}$.
- Number of measurements 200.
- Correlated measurement errors.

Square-Root Schemes

Mean updated using

$$\overline{\boldsymbol{x}^{\mathrm{a}}} = \overline{\boldsymbol{x}^{\mathrm{f}}} + \overline{\boldsymbol{C}}^{\mathrm{f}}_{xx}\boldsymbol{H}^{\mathrm{T}}\left(\boldsymbol{H}\overline{\boldsymbol{C}}^{\mathrm{f}}_{xx}\boldsymbol{H}^{\mathrm{T}} + \overline{\boldsymbol{C}}_{dd}\right)^{-1}\left(\boldsymbol{d} - \boldsymbol{H}\overline{\boldsymbol{x}^{\mathrm{f}}}\right).$$

Perturbations updated using factorization of posterior variance

$$\overline{\boldsymbol{C}}^{\mathrm{a}}_{xx} = \overline{\boldsymbol{C}}^{\mathrm{f}}_{xx} - \overline{\boldsymbol{C}}^{\mathrm{f}}_{xx}\boldsymbol{H}^{\mathrm{T}}\left(\boldsymbol{H}\overline{\boldsymbol{C}}^{\mathrm{f}}_{xx}\boldsymbol{H}^{\mathrm{T}} + \overline{\boldsymbol{C}}_{dd}\right)^{-1}\boldsymbol{H}\overline{\boldsymbol{C}}^{\mathrm{f}}_{xx}.$$

Ensemble notation and simple illustration

$$\boldsymbol{A}^{\mathrm{a}\prime}\boldsymbol{A}^{\mathrm{a}\prime\mathrm{T}} = \boldsymbol{A}'\left(\boldsymbol{I} - \boldsymbol{S}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{S}\right)\boldsymbol{A}'^{\mathrm{T}}$$

Square-Root Schemes

Ensemble notation and simple illustration

$$
\begin{aligned}
\boldsymbol{A}^{\mathrm{a}\prime}\boldsymbol{A}^{\mathrm{a}\prime\mathrm{T}} &= \boldsymbol{A}'\left(\boldsymbol{I} - \boldsymbol{S}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{S}\right)\boldsymbol{A}'^{\mathrm{T}} \\
&= \boldsymbol{A}'\left(\mathbf{Z}\boldsymbol{\Lambda}\mathbf{Z}^{\mathrm{T}}\right)\boldsymbol{A}'^{\mathrm{T}} \\
&= \boldsymbol{A}'\mathbf{Z}\boldsymbol{\Lambda}^{\frac{1}{2}}\left(\boldsymbol{A}'\mathbf{Z}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^{\mathrm{T}} && \text{Non-symmetrical square root} \\
&= \boldsymbol{A}'\mathbf{Z}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{Z}\left(\boldsymbol{A}'\mathbf{Z}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{Z}\right)^{\mathrm{T}} && \text{Symmetrical square root}
\end{aligned}
$$

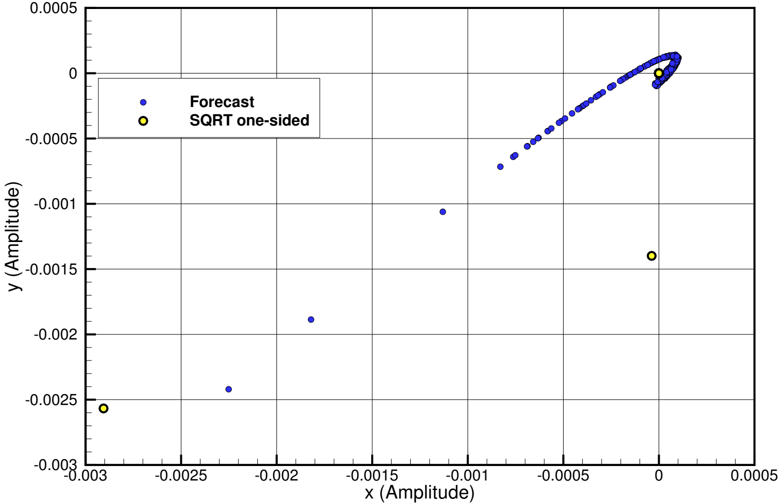
Update becomes

$$\boldsymbol{A}^{\mathrm{a}\prime} = \boldsymbol{A}'\mathbf{Z}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{Z}$$

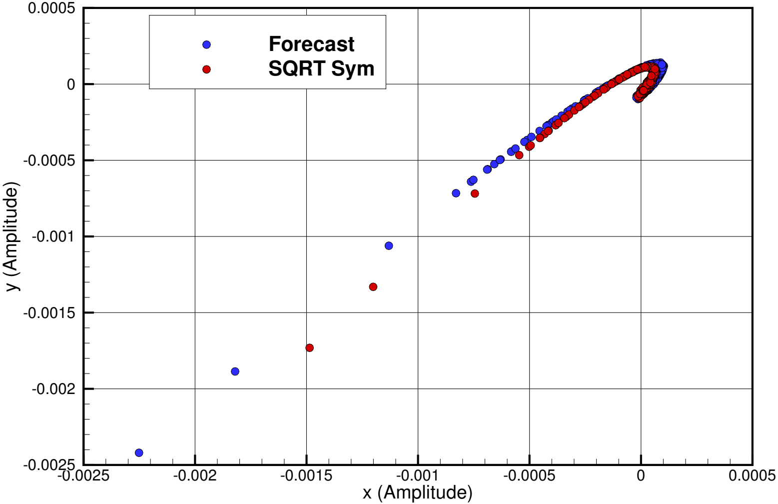
or when including mean preserving random rotation $\boldsymbol{\Phi}\boldsymbol{\Phi}^{\mathrm{T}} = \boldsymbol{I}$

$$\boldsymbol{A}^{\mathrm{a}\prime} = \boldsymbol{A}'\mathbf{Z}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{Z}\boldsymbol{\Phi}$$

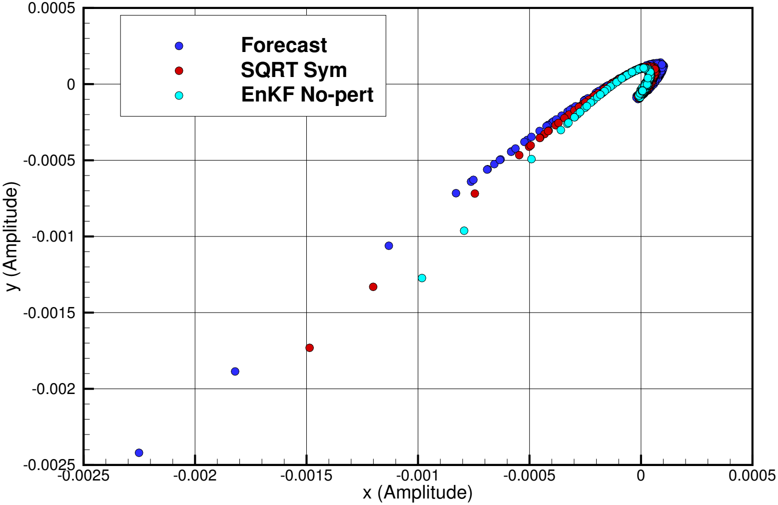
Square-root schemes: Non symmetrical



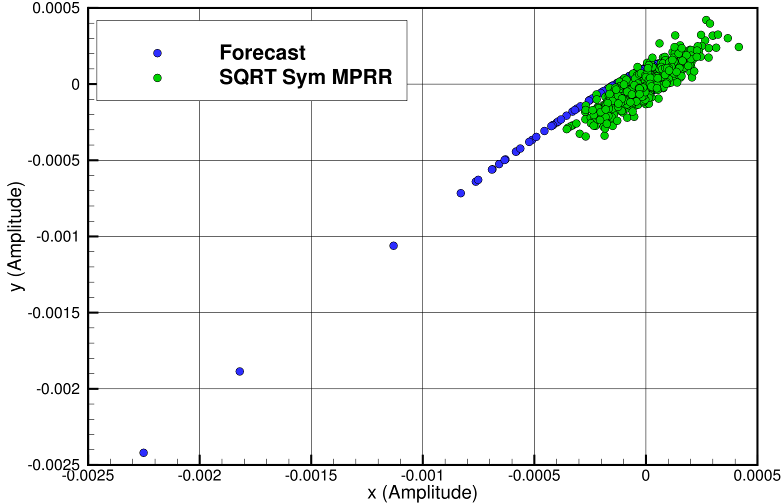
Square-root schemes: Symmetrical



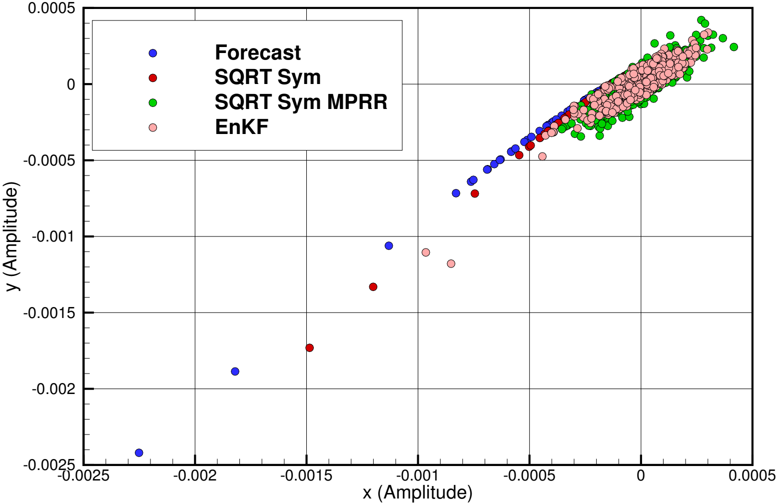
Square-root schemes vs no-pert EnKF



Square root schemes: Symmetrical with MPRR



Square root schemes vs EnKF



Available from: `https://github.com/geirev/Presentations`

Summary

- The EnKF has worked well with highly nonlinear dynamics.
- The EnKF has worked well with high-dimensional models.
- There is no linearization in the evolution of error statistics.
- Major approximation is Gaussian assumption in update step.
- Another approximation is limited ensemble size.

More on ensemble methods

Geir Evensen

Ensemble methods

EnKF: Ensemble Kalman Filter

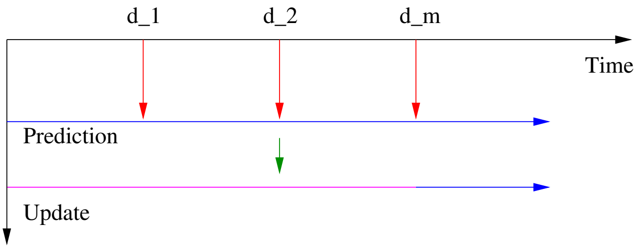
ES: Ensemble Smoother

EnKS: Ensemble Kalman Smoother

- Ensemble representation for pdfs.
- Ensemble prediction for time evolution of pdfs.
- Linear ensemble analysis scheme.

ES: The Ensemble Smoother

Smoother solution processing all data in one go.

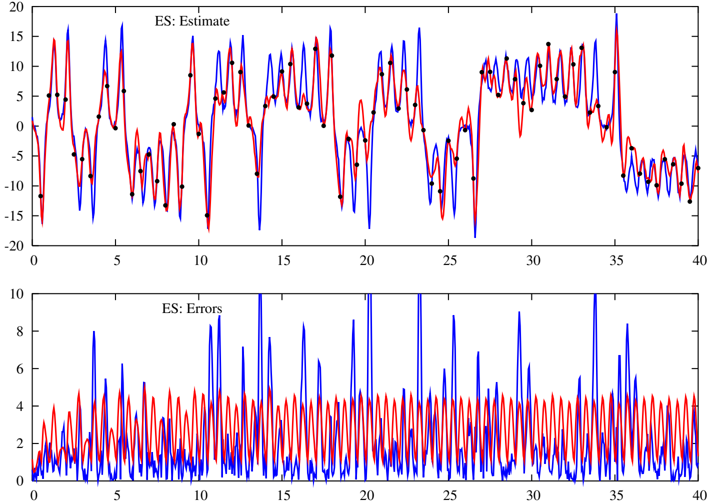


ES: summary

Gauss–Markov interpolation in space and time.

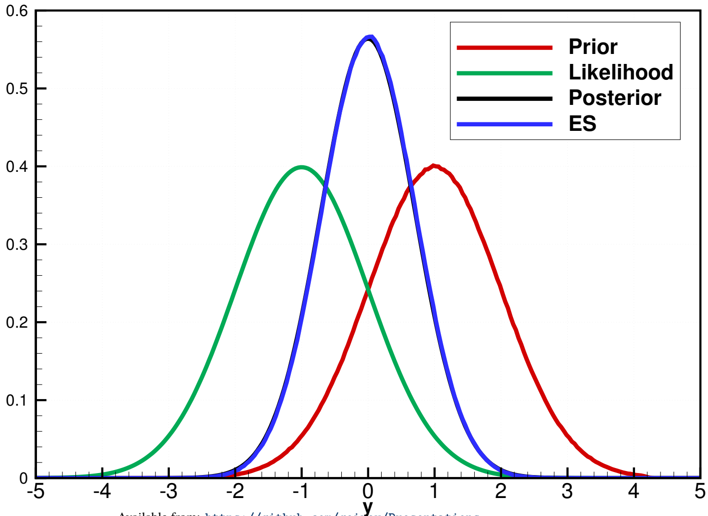
- Creates an ensemble for the model prediction.
- Assumes Gaussian pdf for model prediction.
- Computes variance minimizing ensemble analysis.
- Exact solution for linear problems.

ES: Example with Lorenz equations



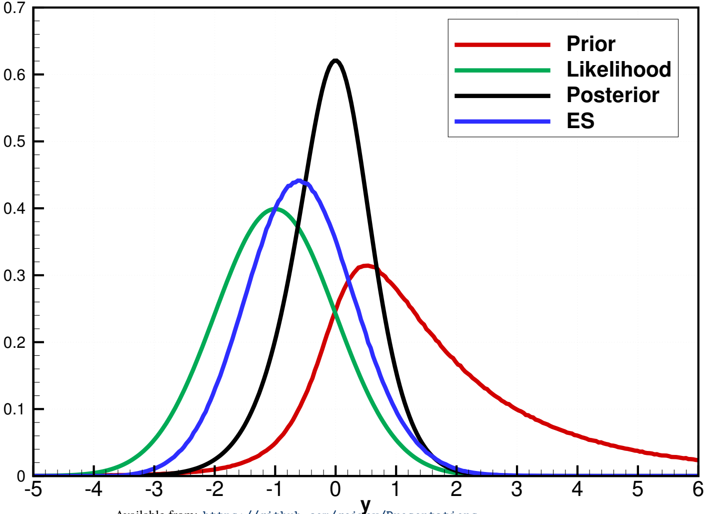
Available from: `https://github.com/geirev/Presentations`

ES vs Bayes' (Gaussian prior)



Available from: https://github.com/geirev/Presentations

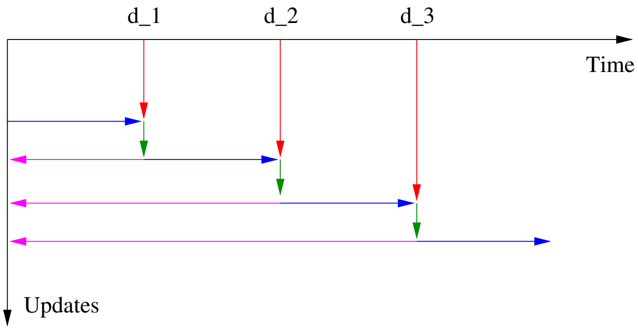
ES vs Bayes' (non-Gaussian prior)



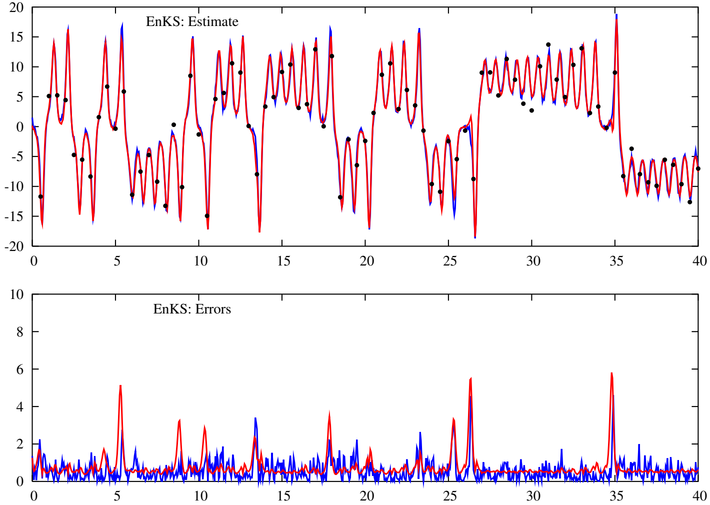
Available from: https://github.com/geirev/Presentations

EnKS: The ensemble Kalman smoother

- Smoother solution with sequential processing of data



EnKS solution



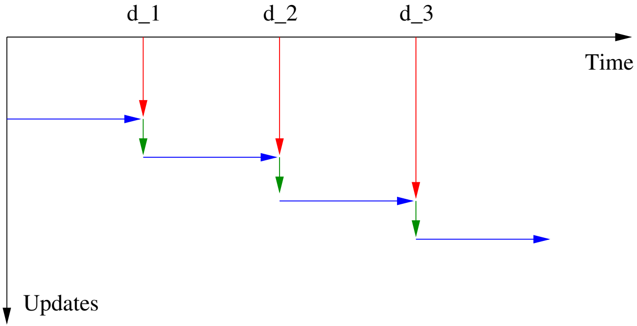
Available from: `https://github.com/geirev/Presentations`

EnKS summary

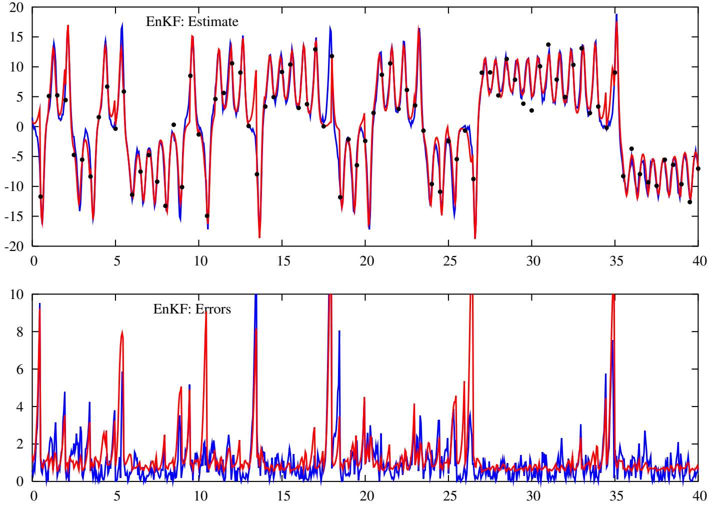
- ES and EnKS give identical results for linear models.
- EnKS is superior to the ES with nonlinear models.
  - Sequential processing of measurements introduces "Gaussianity".
  - Ensemble is kept close to the true state.

EnKF: Ensemble Kalman Filter

- Filtering solution



EnKF solution



Available from: `https://github.com/geirev/Presentations`

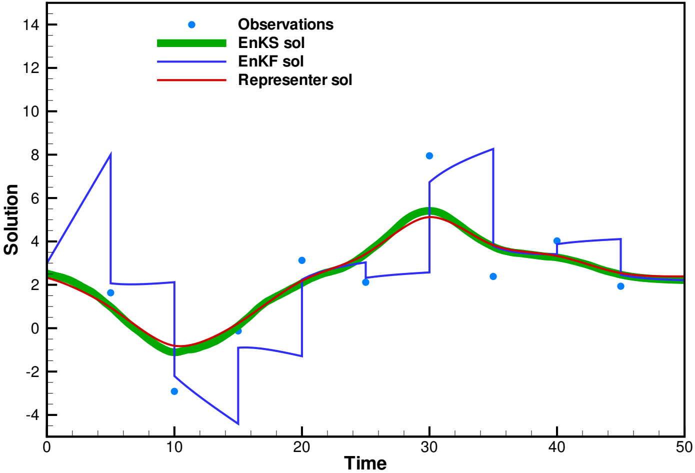
State and parameter estimation

- Scalar model for x with parameter α.

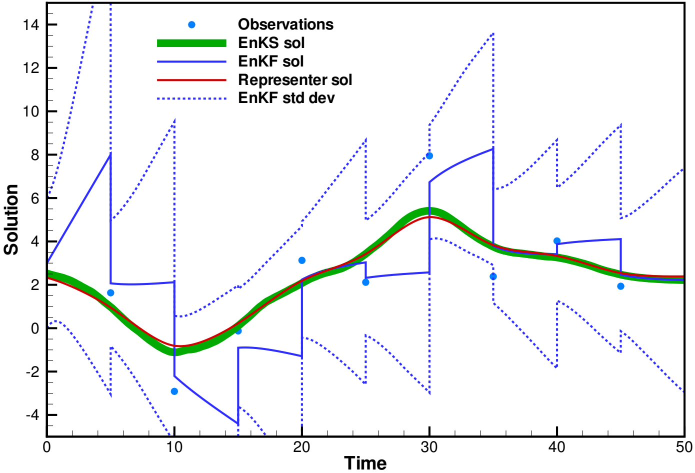
$$\begin{aligned} \frac{\partial x}{\partial t} &= 1 - \alpha + q, \\ x(t=0) &= 3 + a, \\ \alpha &= 0 + \alpha', \\ \mathcal{M}(x) &= \boldsymbol{d} + \boldsymbol{\epsilon}. \end{aligned}$$

- True parameter value is $\alpha = 1$.
- Truly linear model.
- Solved using EnKF, EnKS and Representer methods.
- Exponential time correlation for model errors.

State and parameter estimation

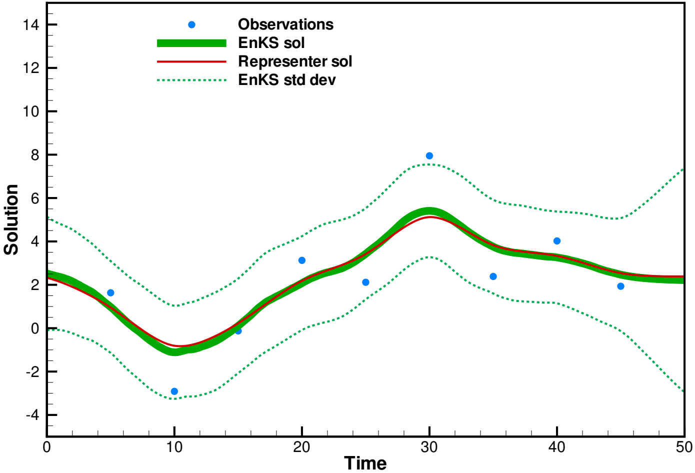


State and parameter estimation



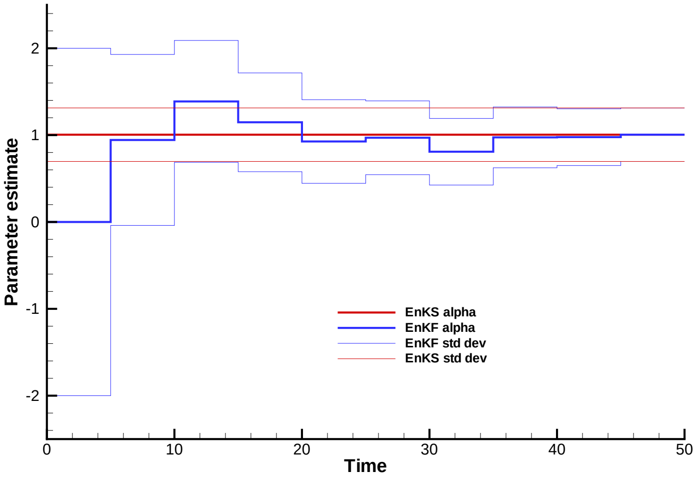
Available from: `https://github.com/geirev/Presentations`

State and parameter estimation



Available from: `https://github.com/geirev/Presentations`

Estimate of parameter



Available from: `https://github.com/geirev/Presentations`

Spurious correlations, inflation and localization

Chapter 15 in book

Geir Evensen

Sampling errors

- EnKF uses a finite ensemble size.
- The update can be written as (slide d_14)

$$\boldsymbol{x}_j^{\mathrm{a}} = \boldsymbol{x}_j^{\mathrm{f}} + \boldsymbol{R}_{\mathrm{e}}^{\mathrm{T}} \boldsymbol{b}_j,$$

  with $\boldsymbol{R}_{\mathrm{e}} = \boldsymbol{H}\boldsymbol{P}_{\mathrm{e}}$ and $\boldsymbol{b}_j = \boldsymbol{C}_{\mathrm{e}}^{-1} \boldsymbol{d}_j$.
- Thus, we add ensemble covariance functions to first guess.

- Long-range spurious correlations introduce sampling errors.

Spurious correlations

- Given an ensemble matrix $\boldsymbol{A} \in \Re^{n\times N}$ and update $\boldsymbol{A}^{\mathrm{a}} = \boldsymbol{A}^{\mathrm{f}}\boldsymbol{X}$.
- Define another matrix $\boldsymbol{B} \in \Re^{\hat{n}\times N}$
  1. Elements are independent random normal-distributed numbers.
  2. Each row has zero mean and unit variance.
- Compute analysis from

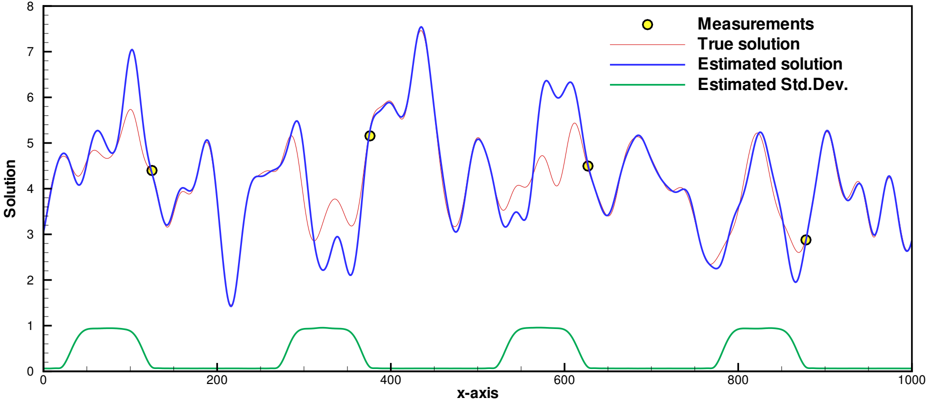
$$\begin{pmatrix}\boldsymbol{A}^{\mathrm{a}}\\ \boldsymbol{B}^{\mathrm{a}}\end{pmatrix} = \begin{pmatrix}\boldsymbol{A}^{\mathrm{f}}\\ \boldsymbol{B}^{\mathrm{f}}\end{pmatrix}\boldsymbol{X}.$$

- Small updates are expected in $\boldsymbol{B}$ even though

$$\lim_{N\to\infty}\frac{\boldsymbol{B}\boldsymbol{S}^{\mathrm{T}}}{N-1} = \boldsymbol{0}.$$

with $\boldsymbol{S} = \boldsymbol{H}\boldsymbol{A}'$.

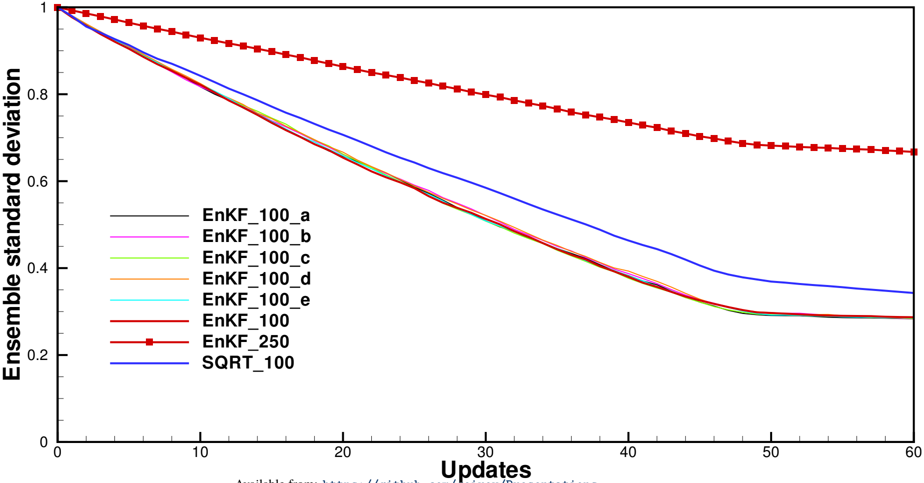
Linear advection example



- Exact wave propagation.
- No model errors.

Impact of spurious correlations

Variance reduction in $\boldsymbol{B}$ ensemble.



Remedies

- Increase ensemble size.
- Inflation: Inflate ensemble variance after update to counteract impact of spurious correlations.
- Localization: localize update to variables that are located close to observations or are strongly correlated to the predicted observations.
- Not always a problem for nonlinear and unstable dynamics, or with large model errors.

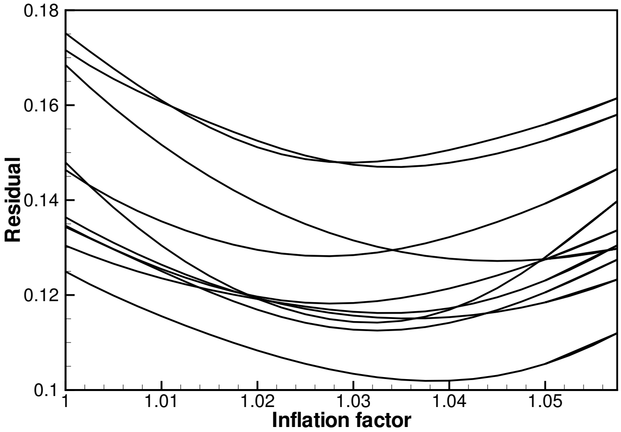
Inflation

$$\boldsymbol{x}_j = \rho(\boldsymbol{x}_j - \overline{\boldsymbol{x}}) + \overline{\boldsymbol{x}}$$

- Accounts for possible under-representation of variance:
  1. Spurious correlations.
  2. Unrepresented model errors.
  3. Too small ensemble space.
- Initial works by Anderson and Anderson (1999), Pham (2001), Hamill et al. (2001)
- Wang and Bishop (2003), estimation based on innovation statistics.
- Anderson (2007a), inflation augmented to state vector
- Sacher and Bartello (2008) derive an analytical expression.
- Li et al. (2009), online estimation of inflation and obs errors.
- Anderson (2009), estimation using a Bayesian algorithm.

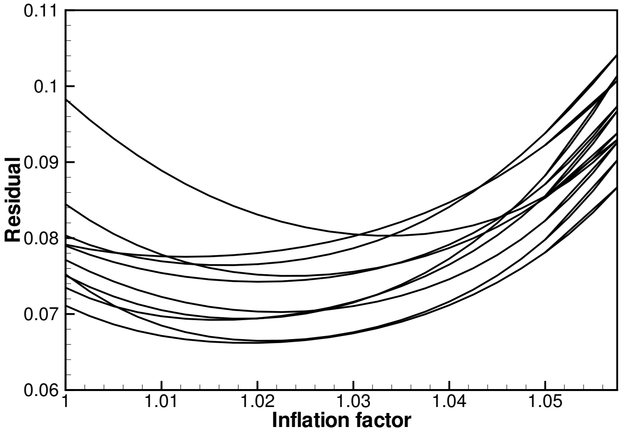
Inflation example (EnKF)

- 10 experiments with different random seeds

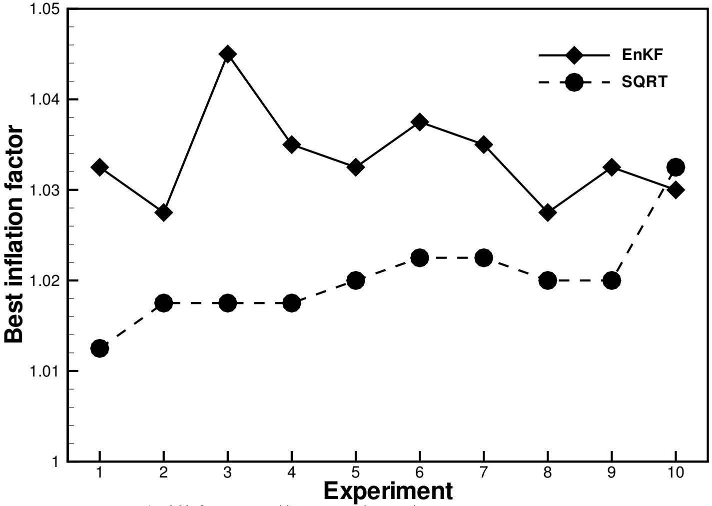


Inflation example (SQRT)

- 10 experiments with different random seeds



Best constant inflation



Available from: `https://github.com/geirev/Presentations`

An adaptive inflation

- Correctly accounts for impact of spurious correlations.

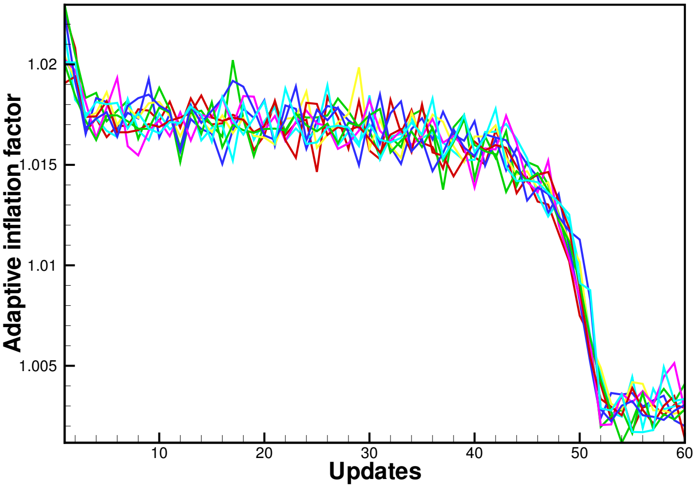
$$\boldsymbol{x}_j = \rho(\boldsymbol{x}_j - \overline{\boldsymbol{x}}) + \overline{\boldsymbol{x}}$$

- At each analysis time, generate (sample) a $\boldsymbol{B}^{\mathrm{f}}$.
- Compute analysis according to

$$\begin{pmatrix} \boldsymbol{A}^{\mathrm{a}} \\ \boldsymbol{B}^{\mathrm{a}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{A}^{\mathrm{f}} \\ \boldsymbol{B}^{\mathrm{f}} \end{pmatrix} \boldsymbol{X}.$$

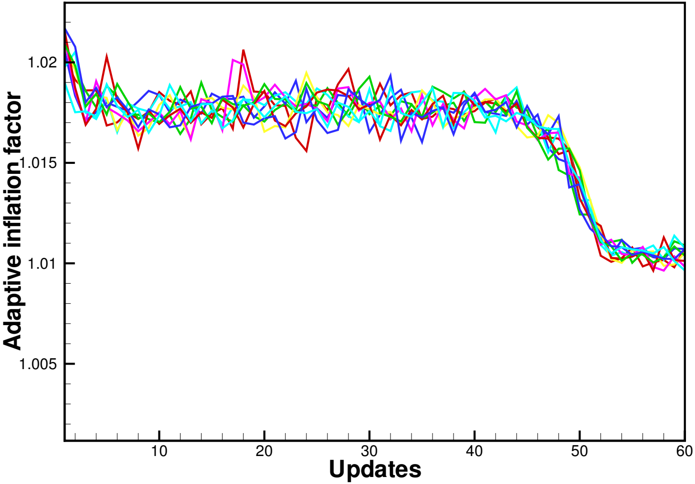
- Define ρ as one over the average std. dev. from the rows in $\boldsymbol{B}^{\mathrm{a}}$.
- Restores the average variance of $\boldsymbol{B}^{\mathrm{a}}$ to one.
- Published in Evensen (2009a,b).

Adaptive inflation (EnKF)



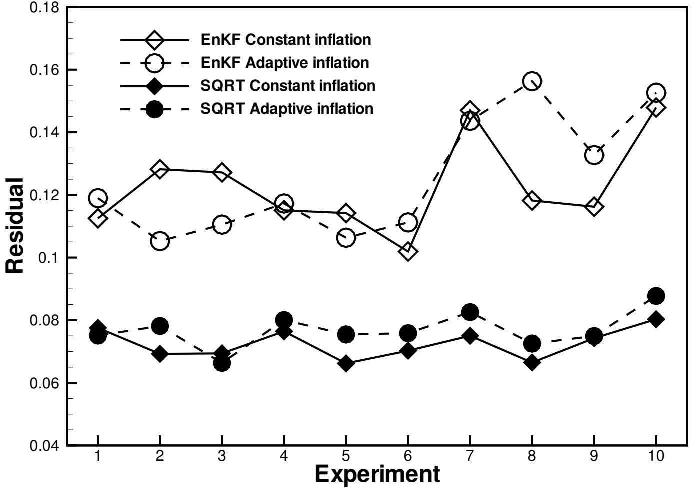
Available from: `https://github.com/geirev/Presentations`

Adaptive inflation (SQRT)



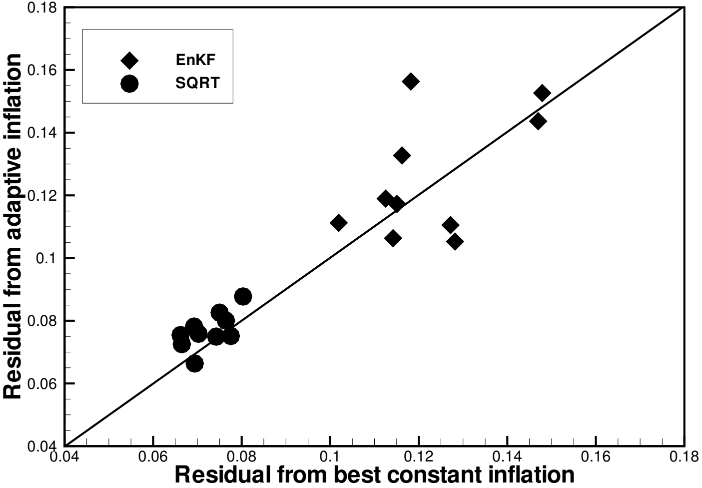
Available from: `https://github.com/geirev/Presentations`

Adaptive vs best constant inflation



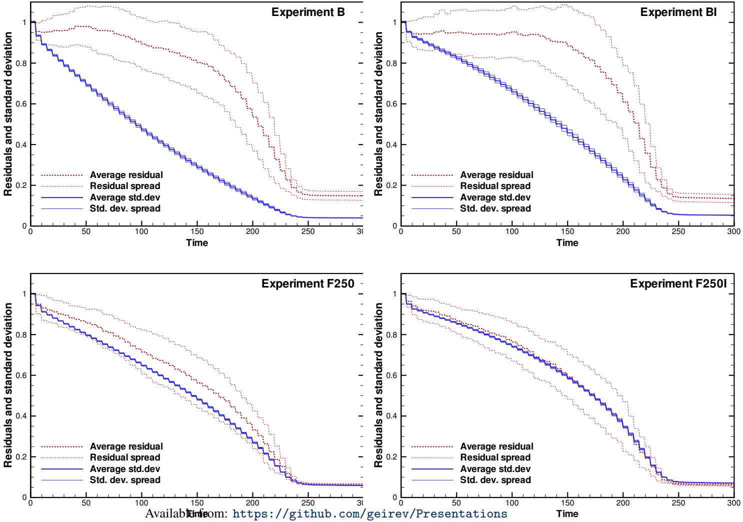
Available from: `https://github.com/geirev/Presentations`

Adaptive vs best constant inflation



Available from: `https://github.com/geirev/Presentations`

Inflation experiments



Inflation Summary

- Inflation can counteract the variance reduction caused by spurious correlations (and possibly other deficiencies).
- Used to tune most large-scale operational systems.
- It is possible to compute a best adaptive inflation to counteract spurious correlations.
  - based on ensemble size, measurement configuration, innovation, and predicted error statistics.

Localization

- Smoothing of Kalman gain using tempering function.
  - Used by Bishop et al. (2001), Hamill et al. (2001), Houtekamer and Mitchell (2001), Whitaker and Hamill (2002) Anderson (2003).

- Updating gridpoint by gridpoint using selected measurements.
  - Used by Brusdal et al. (2003), Evensen (2003), Haugen and Evensen (2002), and Ott et al. (2004).
  - Update equation

$$\boldsymbol{A}^{\mathrm{a}}_{(i,j)} = \boldsymbol{A}_{(i,j)}\boldsymbol{X}_{(i,j)}$$

  - Distance based or adaptive measurement selection?

- Sakov and Bertino (2011) compares the two approaches.

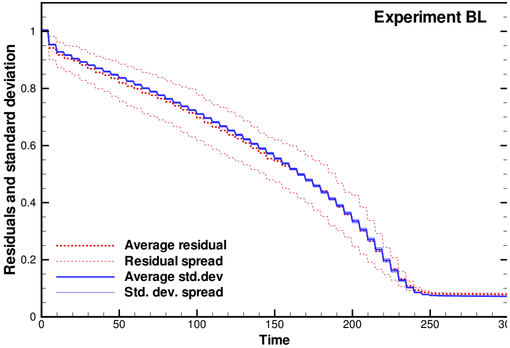
- Localization introduces unbalanced modes and discontinuities.
  - Use large influence radius or filter/smooth updated realizations.

Adaptive localization

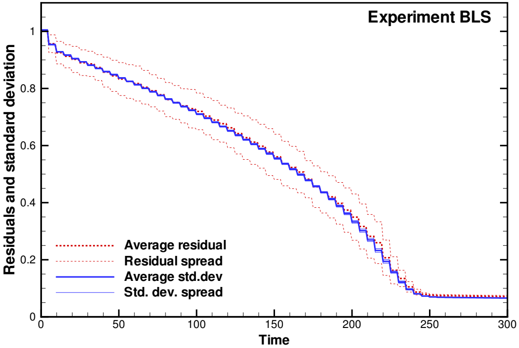
- Determines which measurements shuld be used to update a particular gridpoint.
  - Anderson (2007b) uses many small ensembles to check if correlations are significant.
  - Bishop and Hodyss (2007) use correlations function to derive a tempering function.
  - Fertig et al. (2007) truncate all small correlations.
- We will use the approach by Fertig et al. (2007) below.

Distance based localization

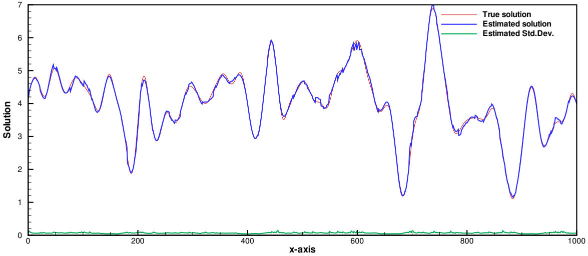
No smoothing



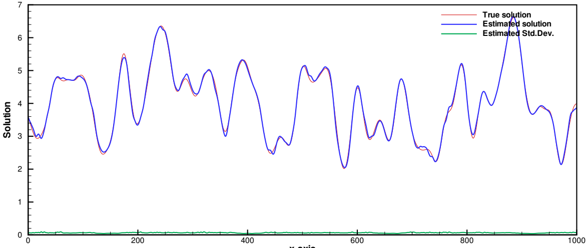
Smoothing



Distance based localization



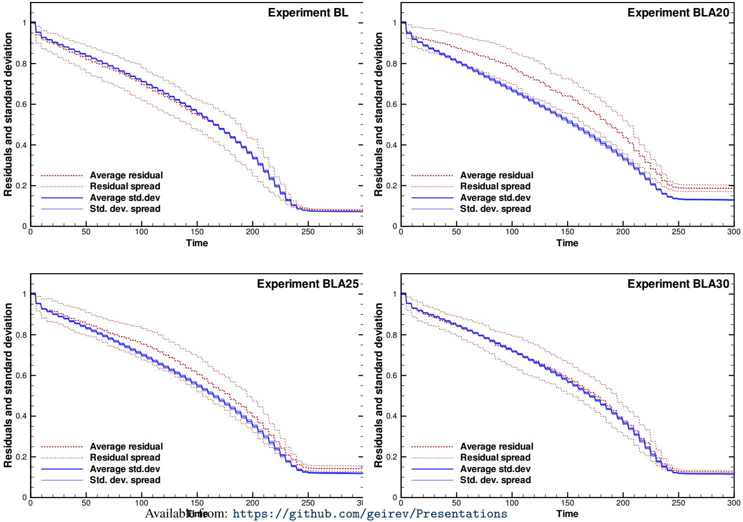
No smoothing



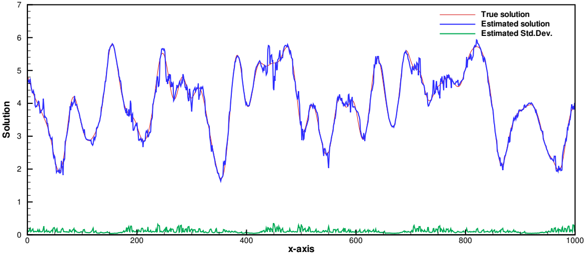
Smoothing

Available from: `https://github.com/geirev/Presentations`

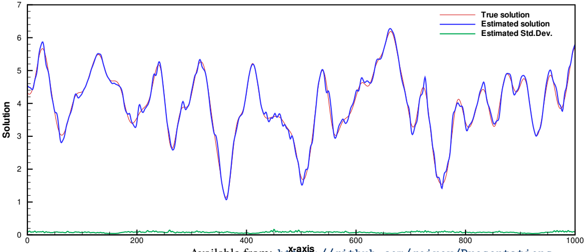
Adaptive localization



Adaptive localization



No smoothing BLA25



Smoothing BLA25S

Available from: `https://github.com/geirev/Presentations`

Summary Localization

- Distance-based localization works when the distance is known.
- Adaptive localization may work in cases with unknown distance.
- Localization eliminates impact of spurious correlations.
- Localization allows to fit small ensembles to large data sets.
- Maintaining a balanced estimate is a challenge.
- Additional smoothing may be needed.
- Localization is used in most operational systems.

Inverse problem

From Evensen (2018)
Link to pdf

Inverse problem

Find $\boldsymbol{x}$ given:

- "Perfect" model

  $$\boldsymbol{y} = \boldsymbol{g}(\boldsymbol{x})$$

  that predicts $\boldsymbol{y} \in \Re^m$ given inputs $\boldsymbol{x} \in \Re^n$.

- Observations $\boldsymbol{d} \in \Re^m$ of $\boldsymbol{y}$.

- The number of measurements m is typically $m \ll n$.
- Standard History-Matching problem in oil-reservoir models.

Inverse problem formulation

Bayes with a perfect model: $\boldsymbol{y} = \boldsymbol{g}(\boldsymbol{x})$

Joint conditional pdf

$$f(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{d}) \propto f(\boldsymbol{x},\boldsymbol{y})f(\boldsymbol{d}|\boldsymbol{y}) = f(\boldsymbol{x})f(\boldsymbol{y}|\boldsymbol{x})f(\boldsymbol{d}|\boldsymbol{y}) = f(\boldsymbol{x})\delta(\boldsymbol{y}-\boldsymbol{g}(\boldsymbol{x}))f(\boldsymbol{d}|\boldsymbol{y}).$$

Marginal pdf

$$f(\boldsymbol{x}|\boldsymbol{d}) = \int f(\boldsymbol{x})\delta(\boldsymbol{y}-\boldsymbol{g}(\boldsymbol{x}))f(\boldsymbol{d}|\boldsymbol{y})d\boldsymbol{y} = f(\boldsymbol{x})f(\boldsymbol{d}|\boldsymbol{g}(\boldsymbol{x})).$$

Gaussian priors

$$f(\boldsymbol{x}|\boldsymbol{d}) = \exp -\frac{1}{2}\left\{(\boldsymbol{x}-\boldsymbol{x}^{\mathrm{f}})^{\mathrm{T}}\boldsymbol{C}_{xx}^{-1}(\boldsymbol{x}-\boldsymbol{x}^{\mathrm{f}}) + (\boldsymbol{g}(\boldsymbol{x})-\boldsymbol{d})^{\mathrm{T}}\boldsymbol{C}_{dd}^{-1}(\boldsymbol{g}(\boldsymbol{x})-\boldsymbol{d})\right\}.$$

Solution by ensemble methods

Maximizing $f(\boldsymbol{x}|\boldsymbol{d})$ is equivalent to minimizing

$$\mathcal{J}(\boldsymbol{x}) = (\boldsymbol{x} - \boldsymbol{x}^{\mathrm{f}})^{\mathrm{T}} \boldsymbol{C}_{xx}^{-1} (\boldsymbol{x} - \boldsymbol{x}^{\mathrm{f}}) + (\boldsymbol{g}(\boldsymbol{x}) - \boldsymbol{d})^{\mathrm{T}} \boldsymbol{C}_{dd}^{-1} (\boldsymbol{g}(\boldsymbol{x}) - \boldsymbol{d}).$$

Exact direct solution in the case when $\boldsymbol{g}(\boldsymbol{x})$ is linear (KF update)

$$\boldsymbol{x}^{\mathrm{a}} = \boldsymbol{x}^{\mathrm{f}} + \boldsymbol{K}(\boldsymbol{d} - \boldsymbol{g}(\boldsymbol{x})), \qquad \boldsymbol{C}_{xx}^{\mathrm{a}} = (\boldsymbol{I} - \boldsymbol{K}\boldsymbol{H})\boldsymbol{C}_{xx}^{\mathrm{f}}.$$

Ensemble representation (EnKF)

$$\boldsymbol{x}_j^{\mathrm{a}} = \boldsymbol{x}_j^{\mathrm{f}} + \boldsymbol{K}_{\mathrm{e}}(\boldsymbol{d}_j - \boldsymbol{g}(\boldsymbol{x}_j)), \qquad \left(\boldsymbol{C}_{xx}^{\mathrm{a}}\right)_{\mathrm{e}} = (\boldsymbol{I} - \boldsymbol{K}_e\boldsymbol{H})\left(\boldsymbol{C}_{xx}^{\mathrm{f}}\right)_{\mathrm{e}}.$$

Equivalent to minimizing

$$\mathcal{J}(\boldsymbol{x}_j) = (\boldsymbol{x}_j - \boldsymbol{x}_j^{\mathrm{f}})^{\mathrm{T}} \boldsymbol{C}_{xx}^{-1} (\boldsymbol{x}_j - \boldsymbol{x}_j^{\mathrm{f}}) + (\boldsymbol{g}(\boldsymbol{x}_j) - \boldsymbol{d}_j)^{\mathrm{T}} \boldsymbol{C}_{dd}^{-1} (\boldsymbol{g}(\boldsymbol{x}_j) - \boldsymbol{d}_j).$$

Solution methods

We will consider three methods

- Ensemble Smoother (ES), Evensen and van Leeuwen (2000), Skjervheim et al. (2011), van Leeuwen and Evensen (1996).
- Iterative Ensemble Smoother (IES), Chen and Oliver (2013).
- Ensemble Smoother with Multiple Data Assimilation (ESMDA), Emerick and Reynolds (2013).

ES/IES/MDA derivation for scalar case

Cost function for each realization j

$$J(x_j) = (x_j - x_j^{\mathrm{f}})C_{xx}^{-1}(x_j - x_j^{\mathrm{f}}) + (g(x_j) - d_j)C_{dd}^{-1}(g(x_j) - d_j)$$

with gradient

$$\frac{\partial J(x_j)}{\partial x_j} = C_{xx}^{-1}\left(x_j^{\mathrm{a}} - x_j^{\mathrm{f}}\right) + g'\left(x_j^{\mathrm{a}}\right)C_{dd}^{-1}\left(g\left(x_j^{\mathrm{a}}\right) - d_j\right) = 0$$

and Hessian

$$\frac{\partial^2 J(x_j)}{\partial {x_j}^2} = 2C_{xx}^{-1} + 2g'(x_j)C_{dd}^{-1}g'(x_j) + 2g''(x_j)C_{dd}^{-1}(g(x_j) - d_j)$$

ES

Insert linearization around x_j^f

$$g\left(x_j^\mathrm{a}\right) \approx g\left(x_j^\mathrm{f}\right) + g'\left(x_j^\mathrm{f}\right)\left(x_j^\mathrm{a} - x_j^\mathrm{f}\right)$$
$$g'\left(x_j^\mathrm{a}\right) \approx g'\left(x_j^\mathrm{f}\right) + g''\left(x_j^\mathrm{f}\right)\left(x_j^\mathrm{a} - x_j^\mathrm{f}\right)$$

in gradient to get

$$C_{dd}\left(x_j^\mathrm{a} - x_j^\mathrm{f}\right) + C_{xx}\left(g\left(x_j^\mathrm{f}\right) + g'\left(x_j^\mathrm{f}\right)\left(x_j^\mathrm{a} - x_j^\mathrm{f}\right) - d_j\right)g'\left(x_j^\mathrm{f}\right) = 0$$

and solve for x_j:

ES with analytic gradient

$$x_j^\mathrm{a} = x_j^\mathrm{f} + g'\left(x_j^\mathrm{f}\right)C_{xx}\left(g'\left(x_j^\mathrm{f}\right)C_{xx}g'\left(x_j^\mathrm{f}\right) + C_{dd}\right)^{-1}\left(d_j - g\left(x_j^\mathrm{f}\right)\right),$$
$$y_j^\mathrm{a} = g\left(x_j^\mathrm{a}\right).$$

Ensemble gradient

Expansion around ensemble mean

$$g\left(x_j^{\mathrm{f}}\right) \approx g\left(x^{\mathrm{f}}\right) + g'\left(x^{\mathrm{f}}\right)\left(x_j^{\mathrm{f}} - x^{\mathrm{f}}\right)$$

$$\begin{aligned}
C_{xx}^{\mathrm{e}} &= \overline{\left(x_j^{\mathrm{f}} - x^{\mathrm{f}}\right)^2}, \\
C_{xy}^{\mathrm{e}} &= \overline{\left(x_j^{\mathrm{f}} - x^{\mathrm{f}}\right)\left(y_j^{\mathrm{f}} - y^{\mathrm{f}}\right)} = \overline{\left(x_j^{\mathrm{f}} - x^{\mathrm{f}}\right)\left(g\left(x_j^{\mathrm{f}}\right) - \overline{g\left(x_j^{\mathrm{f}}\right)}\right)} \\
&\approx \overline{\left(x_j^{\mathrm{f}} - x^{\mathrm{f}}\right)\left(g\left(x^{\mathrm{f}}\right) + g'\left(x^{\mathrm{f}}\right)\left(x_j^{\mathrm{f}} - x^{\mathrm{f}}\right)\right) - \left(\overline{g\left(x^{\mathrm{f}}\right) + g'\left(x^{\mathrm{f}}\right)\left(x_j^{\mathrm{f}} - x^{\mathrm{f}}\right)}\right)} \\
&= g'\left(x^{\mathrm{f}}\right)\overline{\left(x_j^{\mathrm{f}} - x^{\mathrm{f}}\right)^2} = g'\left(x^{\mathrm{f}}\right)C_{xx}^{\mathrm{e}}, \\
C_{yy}^{\mathrm{e}} &= \overline{\left(y_j^{\mathrm{f}} - y^{\mathrm{f}}\right)^2} = g'\left(x^{\mathrm{f}}\right)C_{xx}^{\mathrm{e}}g'\left(x^{\mathrm{f}}\right).
\end{aligned}$$

Ensemble Smoother

ES with ensemble gradient

$$x_j^{\mathrm{a}} = x_j^{\mathrm{f}} + C_{xy}^{\mathrm{e}}\left(C_{yy}^{\mathrm{e}} + C_{dd}^{\mathrm{e}}\right)^{-1}\left(d_j - g\left(x_j^{\mathrm{f}}\right)\right),$$
$$y_j^{\mathrm{a}} = g\left(x_j^{\mathrm{a}}\right).$$

- Make an ensemble prediction to represent covariances.
- For each realization we solve for x_j^{a}, and
- then integrate model to find updated prediction y_j^{a}.
- Approximation is linearization of the nonlinear model:
  - Linearization used to derive update equation, and
  - to replace derivatives with ensemble covariances.

IES

Gauss-Newton iteration (or often Levenberg Marquardt is used)

$$x_{i+1} = x_i - \gamma \frac{\left.\frac{\partial J(x)}{\partial x}\right|_{x=x_i}}{\left.\frac{\partial^2 J(x)}{\partial x^2}\right|_{x=x_i}}$$

IES with analytic gradient

$$x_{j,i+1} = x_{j,i} - \gamma \frac{C^{\mathrm{e}}_{dd}\left(x_{j,i} - x^{\mathrm{f}}_j\right) + g'(x_{j,i}) C^{\mathrm{e}}_{xx}\left(g(x_{j,i}) - d_j\right)}{g'(x_{j,i}) C^{\mathrm{e}}_{xx} g'(x_{j,i}) + C^{\mathrm{e}}_{dd}},$$

$$y_{j,i+1} = g(x_{j,i+1}).$$

IES

Gradient

$$
\begin{aligned}
g'(x_{j,i})C^{\mathrm{e}}_{xx} &= g'(x_{j,i})C^{\mathrm{e},i}_{xx}\left(C^{\mathrm{e},i}_{xx}\right)^{-1}C^{\mathrm{e}}_{xx}\\
&\approx g'(\overline{x_i})C^{\mathrm{e},i}_{xx}\left(C^{\mathrm{e},i}_{xx}\right)^{-1}C^{\mathrm{e}}_{xx}\\
&\approx C^{\mathrm{e},i}_{xy}\left(C^{\mathrm{e},i}_{xx}\right)^{-1}C^{\mathrm{e}}_{xx},
\end{aligned}
$$

IES with ensemble gradient

$$
\begin{aligned}
x_{j,i+1} &= x_{j,i} - \gamma\frac{C^{\mathrm{e}}_{dd}\left(x_{j,i}-x^{\mathrm{f}}_j\right) - C^{\mathrm{e},i}_{xy}\left(C^{\mathrm{e},i}_{xx}\right)^{-1}C^{\mathrm{e}}_{xx}\left(g(x_{j,i})-d_j\right)}{C^{\mathrm{e},i}_{yy}+C^{\mathrm{e}}_{dd}}\\
y_{j,i+1} &= g(x_{j,i+1})
\end{aligned}
$$

ESMDA likelihood

Tempering of likelihood

$$f(d|y) = f(d|y)^{\left(\sum_{i=1}^{N} \frac{1}{\alpha_i}\right)} = \prod_{i=1}^{N} f(d|y)^{\frac{1}{\alpha_i}}, \qquad \left(\sum_{i=1}^{N} \frac{1}{\alpha_i} = 1\right)$$

Gaussian likelihood

$$\begin{aligned} f(d|y) &\propto \exp\left\{-\frac{1}{2}(y-d)C_{dd}^{-1}(y-d)\right\} \\ &= \exp\left\{-\frac{1}{2}\left(\sum_{i=1}^{N} \frac{1}{\alpha_i}\right)(y-d)C_{dd}^{-1}(y-d)\right\} \\ &= \prod_{i=1}^{N} \exp\left\{-\frac{1}{2\alpha_i}(y-d)C_{dd}^{-1}(y-d)\right\} \end{aligned}$$

ESMDA recursion

Bayes'

$$f(x|d) \propto f(x) \prod_{i=1}^{N} f(d|g(x_i))^{\frac{1}{\alpha_i}}$$

Recursive Bayes'

$$\begin{aligned}
f(x_1|d) &\propto f(x)f(d|g(x_1))^{\frac{1}{\alpha_1}} \\
f(x_2|d) &\propto f(x_1|d)f(d|g(x_2))^{\frac{1}{\alpha_2}} \\
&\vdots \\
f(x_N|d) &\propto f(x_{N-1}|d)f(d|g(x_N))^{\frac{1}{\alpha_N}}
\end{aligned}$$

ESMDA

ESMDA solves for each realization j

$$J(x_{j,i+1}) = (x_{j,i+1} - x_{j,i}) \left(C_{xx}^{\mathrm{e},i}\right)^{-1} (x_{j,i+1} - x_{j,i}) \\ + \left(g(x_{j,i+1}) - d - \sqrt{\alpha_i}\epsilon\right) \left(\alpha_i C_{dd}^{\mathrm{e}}\right)^{-1} \left(g(x_{j,i+1}) - d - \sqrt{\alpha_i}\epsilon\right)$$

ESMDA equations

ESMDA with ensemble gradient

$$x_{j,i+1} = x_{j,i} + C_{xy}^{\mathrm{e},i}\left(C_{yy}^{\mathrm{e},i} + \alpha_i C_{dd}^{\mathrm{e}}\right)^{-1}\left(d + \sqrt{\alpha_i}\epsilon_j - g(x_{j,i})\right),$$

$$y_{j,i+1} = g(x_{j,i+1}).$$

Some questions

When using ESMDA or IES:

- Why does it help to iterate?
- What are we iterating?
- What is the difference between IES and ESMDA?
- What does IES and ESMDA converge to?

Ensemble smoothers

Approximately minimizes:

$$J(x_j) = (x_j - x_j^{\mathrm{f}})^{\mathrm{T}} C_{xx}^{-1} (x_j - x_j^{\mathrm{f}}) + (g(x_j) - d_j)^{\mathrm{T}} C_{dd}^{-1} (g(x_j) - d_j)$$

Linear model:

1. ES, ESMDA and IES exactly sample the posterior.

Nonlinear model:

1. ES computes one step based on linearization around x^{f}.
2. ESMDA applies a sequence of local linearizations around $\overline{x_i}$.
3. IES applies an approximate ensemble-based gradient in minimization.
4. Minimization of the cost functions does not exactly sample Bayes.

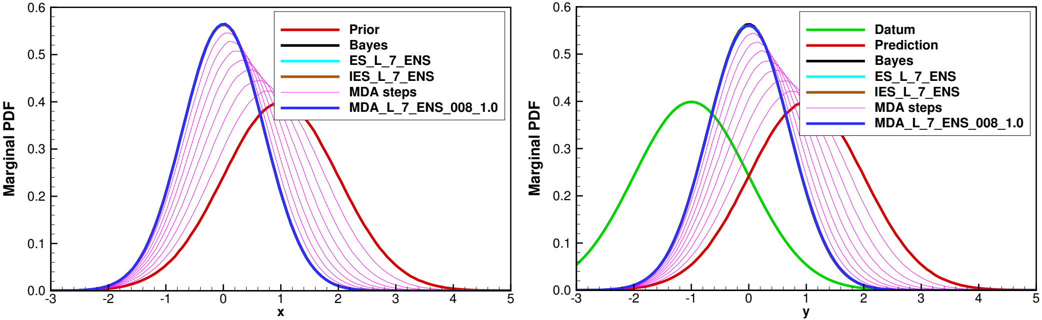
Scalar example

Model

$$y = x(1 + \beta x^2)$$

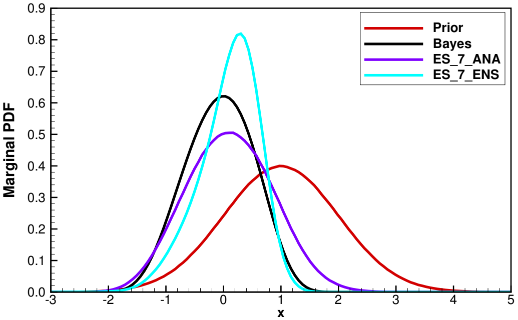
- Linear case: $\beta = 0$.
- Nonlinear case: $\beta = 0.2$.
- Prior ensemble for x: $N(1, 1)$.
- Likelihood for measurement of y: $N(-1, 1)$.

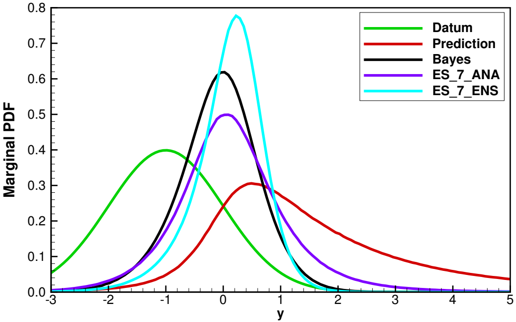
Linear problem



All smoothers converge exactly to the Bayesian posterior.

Nonlinear problem ES



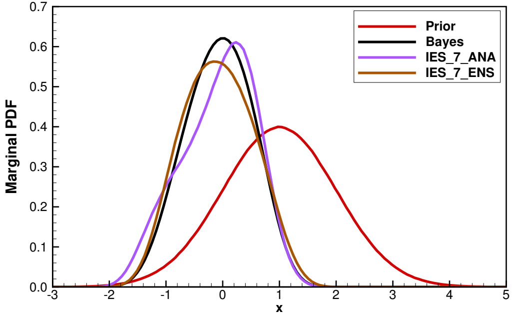


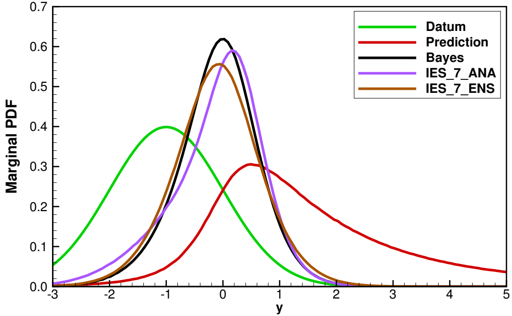
$$x_j^{\mathrm{a}} = x_j^{\mathrm{f}} + g'(x^{\mathrm{f}})C_{xx}\left(g'(x^{\mathrm{f}})C_{xx}g'(x^{\mathrm{f}}) + C_{dd}\right)^{-1}\left(d_j - g(x_j^{\mathrm{f}})\right) \qquad \text{ES\_7\_ANA}$$

$$x_j^{\mathrm{a}} = x_j^{\mathrm{f}} + C_{xy}^{\mathrm{e}}(C_{yy}^{\mathrm{e}} + C_{dd}^{\mathrm{e}})^{-1}(d_j - g(x_j^{\mathrm{f}})) \qquad \text{ES\_7\_ENS}$$

$$y_j^{\mathrm{a}} = g(x_j^{\mathrm{a}})$$

IES: Analytic and Ensemble $\nabla J(x)$

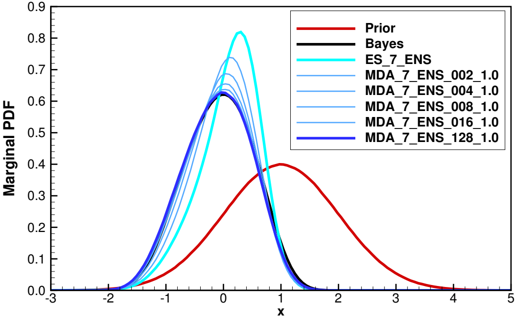


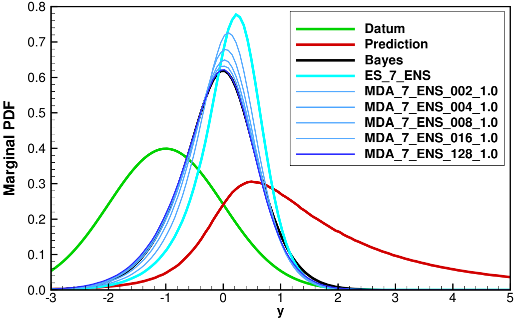


$$x_{j,i+1} = x_{j,i} - \gamma \frac{C_{dd}(x_{j,i} - x_j^{\mathrm{f}}) + g'(x_{j,i})C_{xx}(g(x_{j,i}) - d_j)}{g'(x_{j,i})C_{xx}g'(x_{j,i}) + C_{dd}}$$ IES_7_ANA

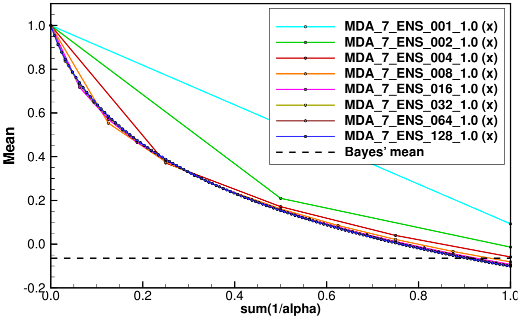
$$x_{j,i+1} = x_{j,i} - \gamma \frac{C_{dd}(x_{j,i} - x_j^{\mathrm{f}}) + C_{xy}^{\mathrm{e},i}\left(C_{xx}^{\mathrm{e},i}\right)^{-1} C_{xx}^{\mathrm{e}}(g(x_{j,i}) - d_j)}{C_{yy}^{\mathrm{e},i} + C_{dd}}$$ IES_7_ENS

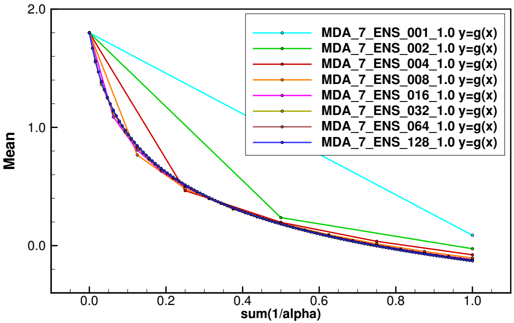
ESMDA convergence (2, 4, 8, 16, 32, 64, 128 steps)



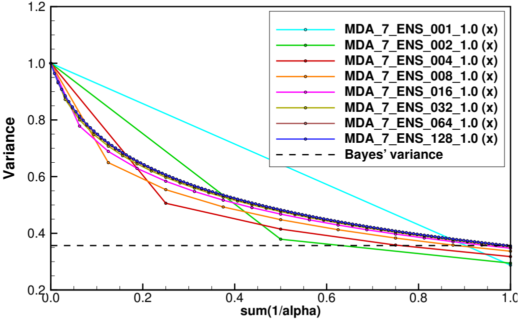


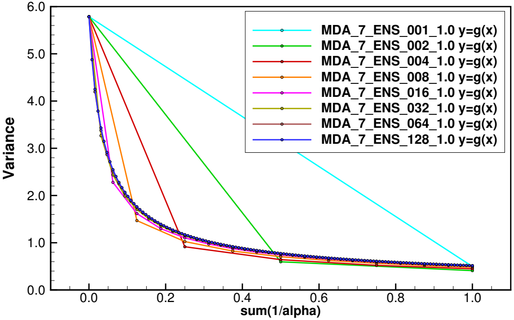
ESMDA convergence
(2, 4, 8, 16, 32, 64, 128 steps)



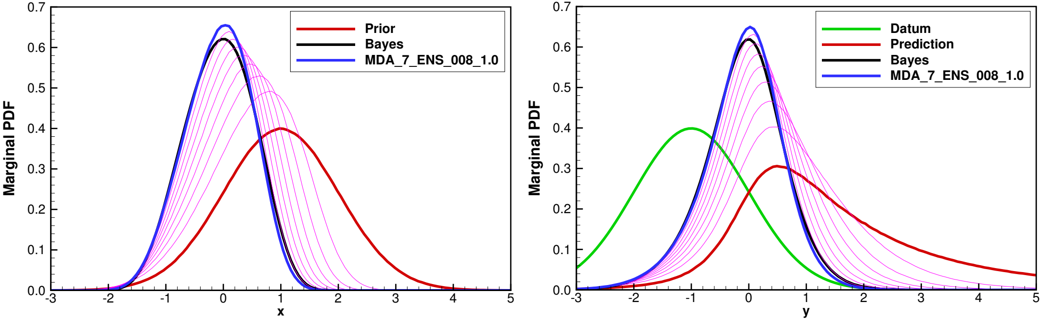


ESMDA convergence
(2, 4, 8, 16, 32, 64, 128 steps)

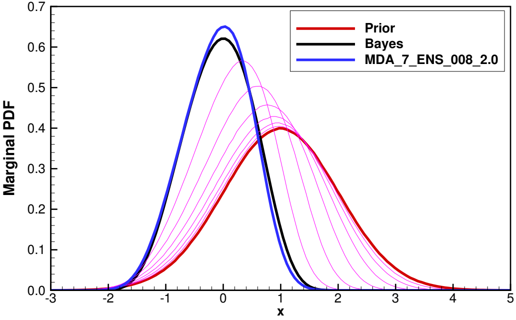


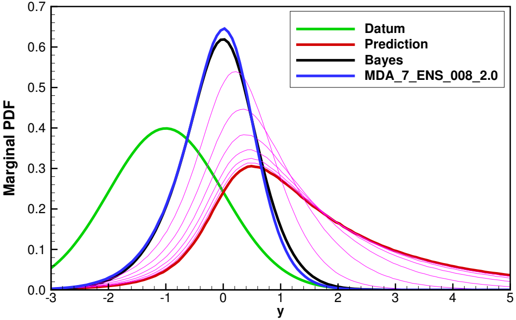


ESMDA–8 : uniform α



ESMDA–8 : geometrical $\alpha(i) = \alpha(i-1)/2$





Acknowledgement

- Statoil for funding this work.

Takeaways

Iterative smoothers like IES and ESMDA beat ES

- Solve different problems but which method is the best?
- What does ESMDA and IES converge to in the nonlinear case?
- What are the optimal number of ESMDA steps and weights?
- The results in this presentation are from Evensen (2018).

Anderson, J. L. A local least squares framework for ensemble filtering. *Mon. Weather Rev.*, 131:634–642, 2003. doi:10.1175/1520-0493(2003)131<0634:ALLSFF>2.0.CO;2.

Anderson, J. L. An adaptive covariance inflation error correction algorithm for ensemble filters. *Tellus, Ser. A*, 59:210–224, 2007a. doi:10.1111/j.1600-0870.2006.00216.x.

Anderson, J. L. Exploring the need for localization in the ensemble data assimilation using a hierarchical ensemble filter. *Physica D*, 230: 99–111, 2007b. doi:10.1016/j.physd.2006.02.011.

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

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

Bishop, C. H. and D. Hodyss. Flow-adaptive moderation of spurious ensemble correlations and its use in ensemble-based data assimilation. *Q. J. R. Meteorol. Soc.*, 133:2029–2044, 2007. doi:10.1002/qj.169.

Bishop, C. H., B. J. Etherton, and S. J. Majumdar. Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Weather Rev.*, 129:420–436, 2001. doi:10.1175/1520-0493(2001)129<0420:ASWTET>2.0.CO;2.

Brusdal, K., J. Brankart, G. Halberstadt, G. Evensen, P. Brasseur, P. van Leeuwen, E. Dombrowsky, and J. Verron. A demonstration of ensemble-based assimilation methods with a layered ogcm from the perspective of operational ocean forecasting systems. *Journal of Marine Systems*, 40-41:253–289, 2003. doi:10.1016/S0924-7963(03)00021-6.

Burgers, G., P. J. van Leeuwen, and G. Evensen. Analysis scheme in the ensemble Kalman filter. *Mon. Weather Rev.*, 126:1719–1724, 1998. doi:10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2.

Chen, Y. and D. S. Oliver. Levenberg-Marquardt forms of the iterative ensemble smoother for efficient history matching and uncertainty quantification. *Computat Geosci*, 17:689–703, 2013. doi:10.1007/s10596-013-9351-5.

Emerick, A. A. and A. C. Reynolds. Ensemble smoother with multiple data assimilation. *Computers and Geosciences*, 55:3–15, 2013. doi:10.1016/j.cageo.2012.03.011.

Evensen, G. The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, 53:343–367, 2003. doi:10.1007/s10236-003-0036-9.

Evensen, G. Sampling strategies and square root analysis schemes for the EnKF. *Ocean Dynamics*, 54:539–560, 2004. doi:10.1007/s10236-004-0099-2.

Evensen, G. The ensemble Kalman filter for combined state and parameter estimation. *IEEE Control Systems Magazine*, 29(3):83–104, 2009a. doi:10.1109/MCS.2009.932223.

Evensen, G. *Data Assimilation: The Ensemble Kalman Filter*. Springer, 2nd edition, 2009b. doi:10.1007/978-3-642-03711-5.

Evensen, G. Analysis of iterative ensemble smoothers for solving inverse problems. *Computat Geosci*, 22(3):885–908, 2018. doi:10.1007/s10596-018-9731-y.

Evensen, G. and P. J. van Leeuwen. An ensemble Kalman smoother for nonlinear dynamics. *Mon. Weather Rev.*, 128:1852–1867, 2000. doi:10.1175/1520-0493(2000)128<1852:AEKSFN>2.0.CO;2.

Fertig, E. J., B. R. Hunt, E. Ott, and I. Szunyogh. Assimilating non-local observations with a local ensemble Kalman filter. *Tellus, Ser. A*, 59: 719–730, 2007. doi:10.1111/j.1600-0870.2007.00260.x.

Hamill, T. M., J. S. Whitaker, and C. Snyder. Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Weather Rev.*, 129:2776–2790, 2001. doi:10.1175/1520-0493(2001)129<2776:DDFOBE>2.0.CO;2.

Haugen, V. E. and G. Evensen. Assimilation of SLA and SST data into an OGCM for the Indian ocean. *Ocean Dynamics*, 52:133–151, 2002.

Haugen, V. E., O. M. Johannessen, and G. Evensen. Indian Ocean: Validation of the Miami Isopycnic Coordinate Ocean Model and ENSO events during 1958–1998. *J. Geophys. Res.*, 107(C5):11–1–11–23, 2002. doi:10.1007/s10236-002-0014-7.

Houtekamer, P. L. and H. L. Mitchell. A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Weather Rev.*, 129: 123–137, 2001. doi:10.1175/1520-0493(2001)129<0123:ASEKFF>2.0.CO;2.

Li, H., E. Kalnay, and T. Miyoshi. Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter. *Q. J. R. Meteorol. Soc.*, 135:523–533, 2009.

Ott, E., B. Hunt, I. Szunyogh, A. V. Zimin, E. Kostelich, M. Corazza, E. Kalnay, D. J. Patil, and J. A. Yorke. A local ensemble Kalman filter for atmospheric data assimilation. *Tellus, Ser. A*, 56A:415–428, 2004. doi:10.3402/tellusa.v56i5.14462.

Pham, D. T. Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Mon. Weather Rev.*, 129:1194–1207, 2001.

Sacher, W. and P. Bartello. Sampling errors in ensemble Kalman filtering. Part I: Theory. *MWR*, 136:3035–3049, 2008.

Sakov, P. and L. Bertino. Relation between two common localization methods for the EnKF. *Computational Geosciences*, 15:225–237, 2011. doi:10.1007/s10596-010-9202-6.

Skjervheim, J.-A., G. Evensen, J. Hove, and J. Vabø. An ensemble smoother for assisted history matching. *SPE conference paper*, 2011. doi:10.2118/141929-MS.

van Leeuwen, P. J. and G. Evensen. Data assimilation and inverse methods in terms of a probabilistic formulation. *Mon. Weather Rev.*, 124: 2898–2913, 1996. doi:10.1175/1520-0493(1996)124<2898:DAAIMI>2.0.CO;2.

Wang, X. and C. H. Bishop. A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes. *J. Atmos. Sci.*, 60: 1140–1158, 2003.

Whitaker, J. S. and T. M. Hamill. Ensemble data assimilation without perturbed observations. *Mon. Weather Rev.*, 130:1913–1924, 2002. doi:10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2.