Data-Assimilation Fundamentals:

Introduction and ensemble methods

Geir Evensen

NORCE

Seminar at DIFA of UNIBO Apr. 7th and May 19th, 2022

Available from `https://github.com/geirev/Data-Assimilation-Fundamentals.git`

Springer Textbooks in Earth Sciences, Geography and Environment

Geir Evensen · Femke C. Vossepoel · Peter Jan van Leeuwen

Data Assimilation Fundamentals

This open-access textbook's significant contribution is the unified derivation of data-assimilation techniques from a common fundamental and optimal starting point, namely Bayes' theorem. Unique for this book is the "top-down" derivation of the assimilation methods. It starts from Bayes theorem and gradually introduces the assumptions and approximations needed to arrive at today's popular data-assimilation methods. This strategy is the opposite of most textbooks and reviews on data assimilation that typically take a bottom-up approach to derive a particular assimilation method. E.g., the derivation of the Kalman Filter from control theory and the derivation of the ensemble Kalman Filter as a low-rank approximation of the standard Kalman Filter. The bottom-up approach derives the assimilation methods from different mathematical principles, making it difficult to compare them. Thus, it is unclear which assumptions are made to derive an assimilation method and sometimes even which problem it aspires to solve. The book's top-down approach allows categorizing data-assimilation methods based on the approximations used. This approach enables the user to choose the most suitable method for a particular problem or application. Have you ever wondered about the difference between the ensemble 4DVar and the "ensemble randomized likelihood" (EnRML) methods? Do you know the differences between the ensemble smoother and the ensemble-Kalman smoother? Would you like to understand how a particle flow is related to a particle filter? In this book, we will provide clear answers to several such questions. The book provides the basis for an advanced course in data assimilation. It focuses on the unified derivation of the methods and illustrates their properties on multiple examples. It is suitable for graduate students, post-docs, scientists, and practitioners working in data assimilation.

ISBN 978-3-030-96708-6

9 783030 967086

▸ springer.com

Evensen · Vossepoel · Leeuwen

Data Assimilation Fundamentals

TEXTBOOK

Geir Evensen · Femke C. Vossepoel
Peter Jan van Leeuwen

Data Assimilation Fundamentals

A Unified Formulation of the State and Parameter Estimation Problem

OPEN ACCESS

Springer

Simple scalar DA example

Given the problem

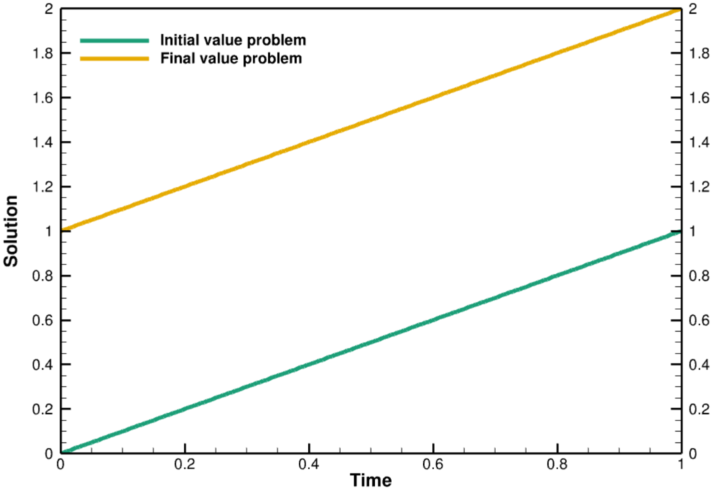
$$\frac{dx}{dt} = 1 \qquad \text{Linear model} \tag{1}$$

$$x(0) = 0 \qquad \text{Initial condition} \tag{2}$$

$$x(1) = 2 \qquad \text{Final condition} \tag{3}$$

- Overdetermined problem.
- No solution.

Initial- and final-value solutions



Allowing for errors in model and conditions

$$\frac{dx}{dt} = 1 + q(t) \tag{4}$$

$$x(0) = 0 + a \tag{5}$$

$$x(1) = 2 + b \tag{6}$$

- Underdetermined problem.
- Infinitively many solutions.

Impose a statistical assumption on the error terms

$$\begin{aligned}
&\text{The mean is zero:} && \text{The variance is known:} && \text{No cross correlations:} && (7)\\
&\overline{q(t)} = 0, && \overline{q(t_1)q(t_2)} = C\delta(t_1 - t_2), && \overline{q(t)a} = 0, && (8)\\
&\overline{a} = 0, && \overline{a^2} = C, && \overline{ab} = 0, && (9)\\
&\overline{b} = 0, && \overline{b^2} = C, && \overline{q(t)b} = 0. && (10)
\end{aligned}$$

We will search for a solution that

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

by minimizing the error terms.

Define a quadratic cost function

$$\mathcal{J}[x] = C^{-1}\int_0^1 \left(\frac{dx}{dt} - 1\right)^2 dt + C^{-1}\big(x(0) - 0\big)^2 + C^{-1}\big(x(1) - 2\big)^2 \tag{11}$$

Then x is an extremum if

$$\delta\mathcal{J}[x] = \mathcal{J}[x + \delta x] - \mathcal{J}[x] = \mathcal{O}\left(\delta x^2\right) \tag{12}$$

when $\delta x \to 0$.

$$\mathcal{J}[x + \delta x] = C^{-1}\int_0^1 \left(\frac{dx}{dt} - 1 + \frac{d\delta x}{dt}\right)^2 dt + C^{-1}\big(x(0) - 0 + \delta x(0)\big)^2 + C^{-1}\big(x(1) - 2 + \delta x(1)\big)^2 \tag{13}$$

Variation of cost function gives

$$\int_0^1 \frac{d\delta x}{dt}\left(\frac{dx}{dt}-1\right)dt \quad + \quad \delta x(0)\big(x(0)-0\big) \quad + \quad \delta x(1)\big(x(1)-2\big) = 0, \tag{14}$$

From integration by part we get

$$\delta x\left(\frac{dx}{dt}-1\right)\Big|_0^1 \quad - \quad \int_0^1 \delta x\frac{d^2x}{dt^2}dt \quad + \quad \delta x(0)\big(x(0)-0\big) \quad + \quad \delta x(1)\big(x(1)-2\big) = 0. \tag{15}$$

Minimum of cost function

This gives the following system of equations

$$\delta x(0)\left(-\frac{dx}{dt}+1+x\right)\bigg|_{t=0}=0, \tag{16}$$

$$\delta x(1)\left(\frac{dx}{dt}-1+x-2\right)\bigg|_{t=1}=0, \tag{17}$$

$$\int_0^1 \delta x\left(\frac{d^2x}{dt^2}\right)dt=0, \tag{18}$$

or since δx is arbitrary....

Euler-Lagrange equation

The Euler–Lagrange equation

$$\frac{d^2x}{dt^2} = 0, \tag{19}$$

$$\frac{dx}{dt} - x = 1 \quad \text{for } t = 0, \tag{20}$$

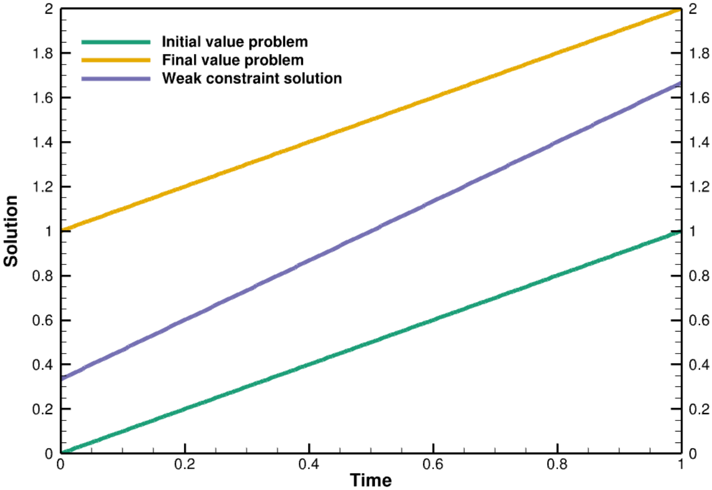
$$\frac{dx}{dt} + x = 3 \quad \text{for } t = 1, \tag{21}$$

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

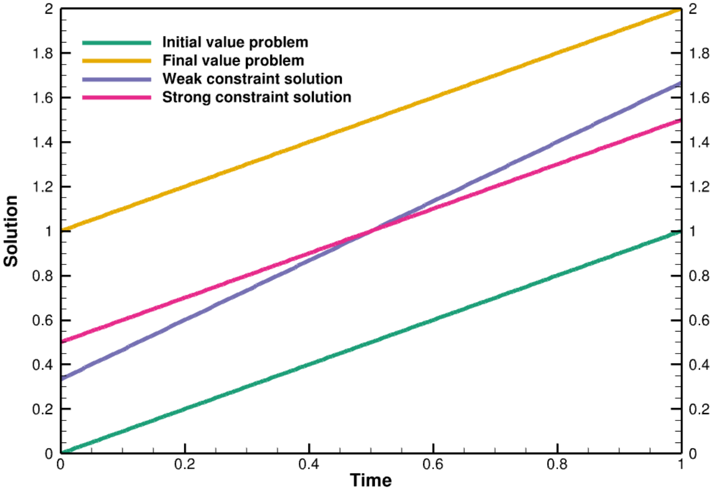
$$x = c_1 t + c_2, \tag{22}$$

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

Weak-constraint solution



Strong-constraint solution



Summary

- A well-posed model with conditions has a unique solution.
- Additional conditions makes the 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.
- The Euler-Lagrange equation defines the Least-squares solution.
- The problem becomes a boundary-value problem in time.
- Weak-constraint solution: almost satisfies dynamics and data.
- Strong-constraint solution: satisfies dynamics, and close to data.

Probabilistic formulation

Initial conditions

$$f\big(x(0)\big) \propto \exp\left(-\frac{a^2}{C}\right) = \exp\left(-\frac{\big(x(0)-0\big)^2}{C}\right) \tag{23}$$

Model evolution

$$f\big(x|x(0)\big) \propto \exp\left(-\int_0^1 \frac{q^2}{C}dt\right) = \exp\left(-\int_0^1 \frac{1}{C}\left(\frac{dx}{dt}-1\right)^2 dt\right) \tag{24}$$

A measurement

$$f\big(d|x(1)\big) \propto \exp\left(-\frac{b^2}{C}\right) = \exp\left(-\frac{\big(x(1)-2\big)^2}{C}\right) \tag{25}$$

Bayes' theorem for the scalar model with Gaussian priors

$$\begin{aligned}
f(x|d) &= \frac{f(d|x)\,f(x)}{f(d)} && (26)\\
&\propto f\big(d|x(1)\big)\,\Big(f\big(x|x(0)\big)\,f\big(x(0)\big)\Big) && (27)\\
&= \exp\left(-\frac{\big(x(1)-2\big)^2}{C}\right)\exp\left(-\int_0^1 \frac{1}{C}\left(\frac{dx}{dt}-1\right)^2 dt\right)\exp\left(-\frac{\big(x(0)-0\big)^2}{C}\right) && (28)\\
&= \exp\big(-\mathcal{J}[x]\big) && (29)
\end{aligned}$$

Hence, maximizing the probability is equivalent to minimizing the cost function.

Why Bayes'?

Bayes' formula is the optimal starting point and arises as the first-order optimality condition from the joint minimization of the Kullback-Leibler (KL) divergence between a posterior and prior distribution and the mean-square errors of the data represented by the likelihood.

Bayes' formula elegantly shows how to update prior information when new information becomes available.

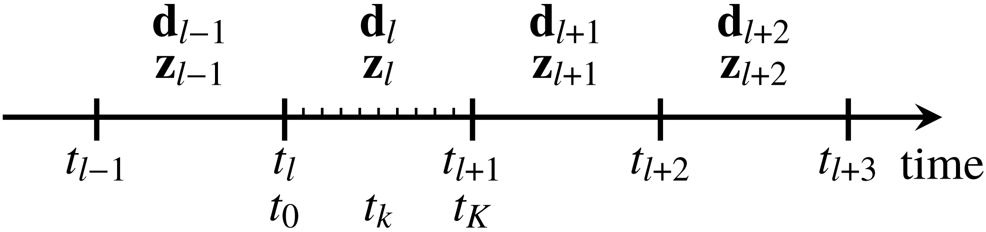
One of the strengths of Bayes' theorem is that it does not try to solve the ill-defined problem of "inverting observations" but instead updates prior knowledge.

We start from Bayes' theorem

$$\boxed{f(\mathcal{Z}|\mathcal{D}) = \frac{f(\mathcal{D}|\mathcal{Z})f(\mathcal{Z})}{f(\mathcal{D})}. \qquad (30)}$$

- $\mathcal{Z} = (\mathbf{z}_0, \mathbf{z}_1, \ldots, \mathbf{z}_L)$ is the vector of state variables on all the assimilation windows.
- $\mathcal{D} = (\mathbf{d}_1, \ldots, \mathbf{d}_L)$ is the vector containing all the measurements.

Split time into data-assimilation windows



- We consider the DA problem for one single window.
- Errors propagate from one window to the next by ensemble integrations.

Model is Markov process

Approximation 1 (Model is 1st-order Markov process)

We assume the dynamical model is a 1st-order Markov process.

$$f(\mathbf{z}_l|\mathbf{z}_{l-1}, \mathbf{z}_{l-2}, \ldots, \mathbf{z}_0) = f(\mathbf{z}_l|\mathbf{z}_{l-1}), \tag{31}$$

Independent measurements

Approximation 2 (Independent measurements)

We assume that measurements are independent between different assimilation windows.

Independent measurements have uncorrelated errors

$$f(\mathcal{D}|\mathcal{Z}) = \prod_{l=1}^{L} f(\mathbf{d}_l|\mathbf{z}_l). \tag{32}$$

Recursive form of Bayes

$$f(\mathbf{z}_1, \mathbf{z}_0 | \mathbf{d}_1) = \frac{f(\mathbf{d}_1|\mathbf{z}_1) f(\mathbf{z}_1|\mathbf{z}_0) f(\mathbf{z}_0)}{f(\mathbf{d}_1)}, \tag{33}$$

$$f(\mathbf{z}_2, \mathbf{z}_1, \mathbf{z}_0 | \mathbf{d}_1, \mathbf{d}_2) = \frac{f(\mathbf{d}_2|\mathbf{z}_2) f(\mathbf{z}_2|\mathbf{z}_1) f(\mathbf{z}_1, \mathbf{z}_0|\mathbf{d}_1)}{f(\mathbf{d}_2)}, \tag{34}$$

$$\vdots \tag{35}$$

$$f(\mathcal{Z}|\mathcal{D}) = \frac{f(\mathbf{d}_L|\mathbf{z}_L) f(\mathbf{z}_L|\mathbf{z}_{L-1}) f(\mathbf{z}_{L-1}, \ldots, \mathbf{z}_0|\mathbf{d}_{L-1}, \ldots, \mathbf{d}_1)}{f(\mathbf{d}_L)}. \tag{36}$$

Filtering assumption

Approximation 3 (Filtering assumption)

We approximate the full smoother solution with a sequential data-assimilation solution. We only update the solution in the current assimilation window, and we do not project the measurement's information backward in time from one assimilation window to the previous ones.

Recursive Bayes' for filtering

$$f(\mathbf{z}_1|\mathbf{d}_1) = \frac{f(\mathbf{d}_1|\mathbf{z}_1)\int f(\mathbf{z}_1|\mathbf{z}_0)f(\mathbf{z}_0)\,d\mathbf{z}_0}{f(\mathbf{d}_1)} = \frac{f(\mathbf{d}_1|\mathbf{z}_1)f(\mathbf{z}_1)}{f(\mathbf{d}_1)}, \tag{37}$$

$$f(\mathbf{z}_2|\mathbf{d}_1,\mathbf{d}_2) = \frac{f(\mathbf{d}_2|\mathbf{z}_2)\int f(\mathbf{z}_2|\mathbf{z}_1)f(\mathbf{z}_1|\mathbf{d}_1)\,d\mathbf{z}_1}{f(\mathbf{d}_2)} = \frac{f(\mathbf{d}_2|\mathbf{z}_2)f(\mathbf{z}_2|\mathbf{d}_1)}{f(\mathbf{d}_2)}, \tag{38}$$

$$\vdots$$

$$f(\mathbf{z}_L|\mathcal{D}) = \frac{f(\mathbf{d}_L|\mathbf{z}_L)\int f(\mathbf{z}_L|\mathbf{z}_{L-1})f(\mathbf{z}_{L-1}|\mathbf{d}_{L-1},\ldots,\mathbf{d}_1)\,d\mathbf{z}_{L-1}}{f(\mathbf{d}_L)} \tag{39}$$

$$= \frac{f(\mathbf{d}_L|\mathbf{z}_L)f(\mathbf{z}_L|\mathbf{d}_{L-1})}{f(\mathbf{d}_L)}. \tag{40}$$

Or just Bayes' for the assimilation window

$$\boxed{f(\mathbf{z}|\mathbf{d}) = \frac{f(\mathbf{d}|\mathbf{z})f(\mathbf{z})}{f(\mathbf{d})}, \qquad (41)}$$

Discrete model with uncertain inputs

$$\mathbf{x}_k = \mathbf{m}(\mathbf{x}_{k-1}, \boldsymbol{\theta}, \mathbf{u}_k, \mathbf{q}_k). \tag{42}$$

- $\mathbf{x}_k$ is the model state.
- $\boldsymbol{\theta}$ are model parameters.
- $\mathbf{u}_k$ are model controls.
- $\mathbf{q}_k$ are model errors.
- Define $\mathbf{x} = (\mathbf{x}_0, \ldots, \mathbf{x}_K)$ as model state over the assimilation window.
- Define $\mathbf{q} = (\mathbf{q}_0, \ldots, \mathbf{q}_K)$ as model errors over the assimilation window.
- Define $\mathbf{u} = (\mathbf{u}_0, \ldots, \mathbf{u}_K)$ as model forcing over the assimilation window.
- Define $\mathbf{z} = (\mathbf{x}, \boldsymbol{\theta}, \mathbf{u}, \mathbf{q})$ as state vector for assimilation problem.

General smoother formulation

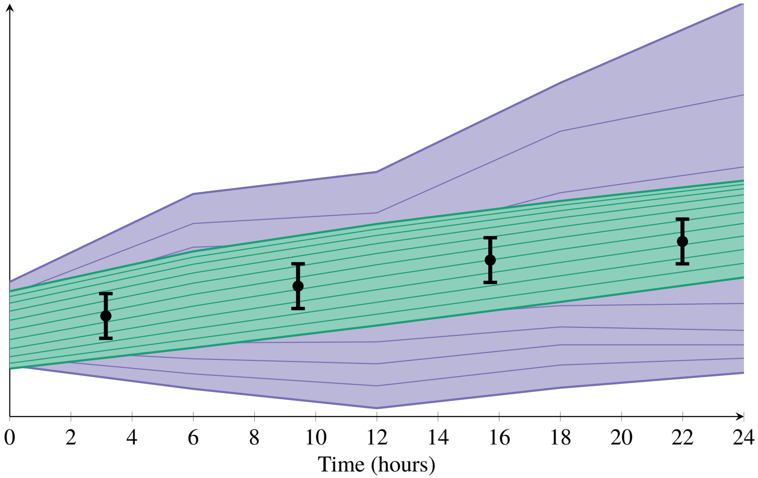
- Solve for model solution over an assimilation window $\mathbf{x}$.
- Condition on measurements distributed over the assimilation window.

Predicted measurements $\mathbf{y}$

$$\mathbf{y} = \mathbf{g}(\mathbf{z}) = \mathbf{h}(\mathbf{x}) = \mathbf{h}\big(\mathbf{m}(\mathbf{x}_0, \mathbf{q})\big) \tag{43}$$

- Measurement operator $\mathbf{h}$.
- Ensemble smoother (ES) solution, weak constraint 4DVar, Representer method.

General smoother formulation



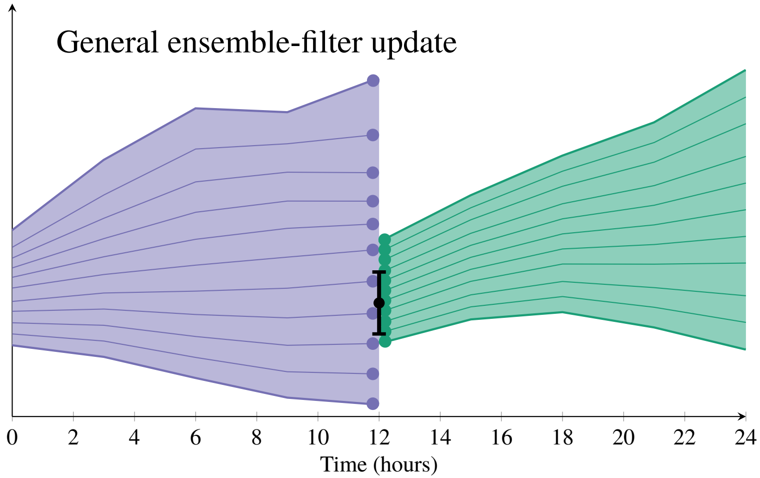
General filter formulation

- Solve for model solution at the end of an assimilation window $\mathbf{x}_K$.
- Condition on measurements at the end of the assimilation window.

$$\mathbf{y} = \mathbf{g}(\mathbf{z}) = \mathbf{h}(\mathbf{x}_K). \tag{44}$$

- Kalman filters
- EnKF (also allows for measurments distributed over the assimilation window)
- Particle filter

General filter formulation



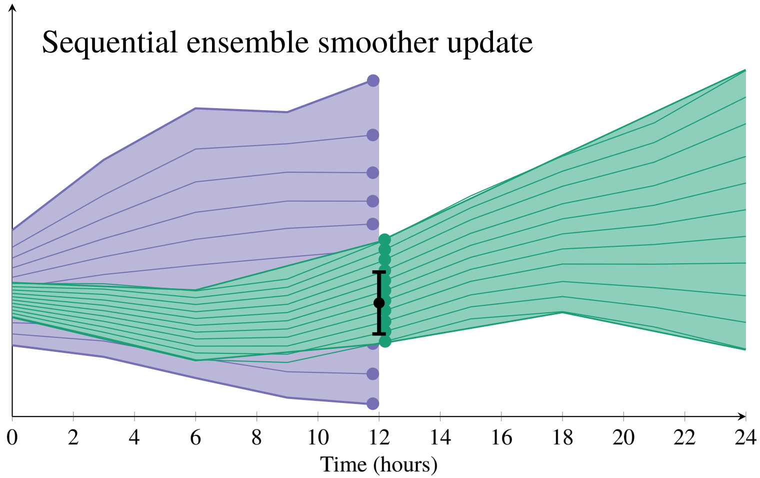
Recursive smoother formulation

- Solve for model solution in the whole (and previous) assimilation window(s) $\mathbf{x}$.
- Condition on measurements at the end of the assimilation window.

$$\mathbf{y} = \mathbf{g}(\mathbf{z}) = \mathbf{h}(\mathbf{x}_K), \tag{45}$$

- Ensemble Kalman Smoother (EnKS)

Recursive smoother formulation



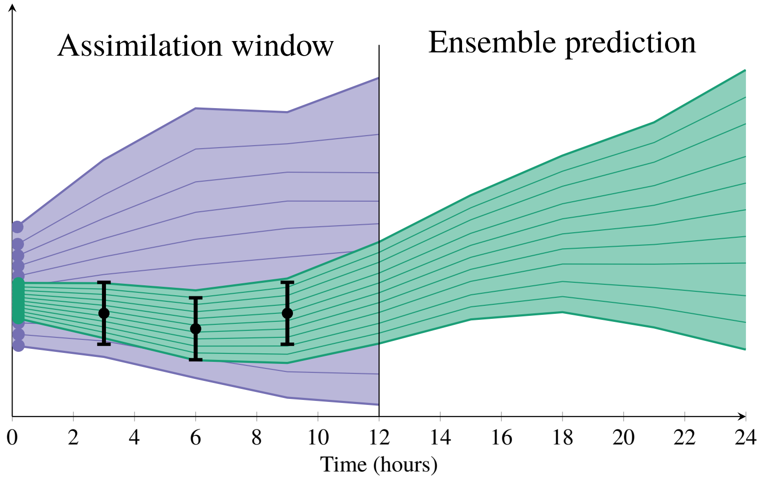
Smoother for perfect models

- Solve for model solution at the beginning of an assimilation window $\mathbf{x}_0$.
- Condition on measurements distributed over the assimilation window.

$$\mathbf{y} = \mathbf{g}(\mathbf{z}) = \mathbf{h}\big(\mathbf{m}(\mathbf{x}_0)\big). \tag{46}$$

- Strong constraint 4DVar.
- Iterative ensemble smoothers (EnRML, ESMDA).
- Parameter estimation (replace $\mathbf{x}_0$ with $\boldsymbol{\theta}$).

Smoother for perfect models



Bayes' theorem related to the predicted measurements

We introduce nonlinearity through the likelihood

$$f(\mathbf{z}|\mathbf{d}) = \frac{f\big(\mathbf{d}|\mathbf{g}(\mathbf{z})\big)f(\mathbf{z})}{f(\mathbf{d})}. \tag{47}$$

Gaussian assumption

Approximation 4 (Gaussian prior and likelihood)

We assume that the prior distributions of the state vector's components $\mathbf{z}$ *and observation errors* $\boldsymbol{\epsilon}$ *are both Gaussian distributed.*

$$f(\mathbf{z}|\mathbf{d}) \propto \exp\left\{-\mathcal{J}(\mathbf{z})\right\}, \tag{48}$$

Leads to a cost-function formulation for the MAP solution

Cost function

$$\mathcal{J}(\mathbf{z}) = \frac{1}{2}\left(\mathbf{z}-\mathbf{z}^{\mathrm{f}}\right)^{\mathrm{T}}\mathbf{C}_{zz}^{-1}\left(\mathbf{z}-\mathbf{z}^{\mathrm{f}}\right) + \frac{1}{2}\left(\mathbf{g}(\mathbf{z})-\mathbf{d}\right)^{\mathrm{T}}\mathbf{C}_{dd}^{-1}\left(\mathbf{g}(\mathbf{z})-\mathbf{d}\right). \tag{49}$$

The gradient set to zero

$$\mathbf{C}_{zz}^{-1}\left(\mathbf{z}^{\mathrm{a}}-\mathbf{z}^{\mathrm{f}}\right) + \nabla_{\mathbf{z}}\mathbf{g}\left(\mathbf{z}^{\mathrm{a}}\right)\mathbf{C}_{dd}^{-1}\left(\mathbf{g}\left(\mathbf{z}^{\mathrm{a}}\right)-\mathbf{d}\right) = 0. \tag{50}$$

- There is no explicit solution of the gradient equation.

Gauss-Newton methods solves for the MAP estimate

Gauss-Newton iteration

$$\mathbf{z}^{i+1} = \mathbf{z}^{i} - \gamma^{i}\left(\mathbf{C}_{zz}^{-1} + {\mathbf{G}^{i}}^{\mathrm{T}}\mathbf{C}_{dd}^{-1}\mathbf{G}^{i}\right)^{-1}\left(\mathbf{C}_{zz}^{-1}(\mathbf{z}^{i} - \mathbf{z}^{\mathrm{f}}) + {\mathbf{G}^{i}}^{\mathrm{T}}\mathbf{C}_{dd}^{-1}\left(\mathbf{g}(\mathbf{z}^{i}) - \mathbf{d}\right)\right). \tag{51}$$

- The incremental formulation is sometimes more convenient.

Incremental Gauss-Newton methods

Quadratic cost function for the increments

$$\mathcal{J}(\delta\mathbf{z}) = \frac{1}{2}\left(\delta\mathbf{z} - \boldsymbol{\xi}^i\right)^{\mathrm{T}} \mathbf{C}_{zz}^{-1}\left(\delta\mathbf{z} - \boldsymbol{\xi}^i\right) + \frac{1}{2}\left(\mathbf{G}^i\delta\mathbf{z} - \boldsymbol{\eta}^i\right)^{\mathrm{T}} \mathbf{C}_{dd}^{-1}\left(\mathbf{G}^i\delta\mathbf{z} - \boldsymbol{\eta}^i\right). \tag{52}$$

with

$$\mathbf{z}^{i+1} = \mathbf{z}^i + \delta\mathbf{z}, \tag{53}$$

$$\boldsymbol{\eta}^i = \mathbf{d} - \mathbf{g}\left(\mathbf{z}^i\right), \tag{54}$$

$$\boldsymbol{\xi}^i = \mathbf{z}^{\mathrm{f}} - \mathbf{z}^i. \tag{55}$$

- Sequence of linear iterates.
- Solved by SC-4DVar, WC-4DVar, and Representer method.

Linearization leads to an axpproximate explicit solution

Approximation 5 (Linearization)

Linearize $\mathbf{g}(\mathbf{z})$ *around the prior estimate* $\mathbf{z}^{\mathrm{f}}$,

$$\mathbf{g}(\mathbf{z}) \approx \mathbf{g}\left(\mathbf{z}^{\mathrm{f}}\right) + \mathbf{G}\left(\mathbf{z} - \mathbf{z}^{\mathrm{f}}\right), \tag{56}$$

and approximate the gradient evaluated at the prior estimate

$$\nabla_{\mathbf{z}}\mathbf{g}(\mathbf{z}) \approx \mathbf{G}^{\mathrm{T}}, \tag{57}$$

where we have defined

$$\mathbf{G}^{\mathrm{T}} = \nabla_{\mathbf{z}}\mathbf{g}(\mathbf{z})\Big|_{\mathbf{z}=\mathbf{z}^{\mathrm{f}}}. \tag{58}$$

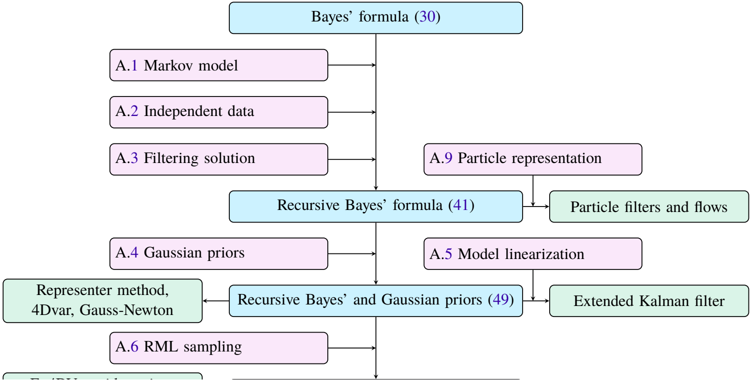
$\mathbf{G}$ *is the tangent-linear operator of* $\mathbf{g}(\mathbf{z})$ *and* $\mathbf{G}^{\mathrm{T}}$ *is its adjoint.*

Extended Kalman Filter solution

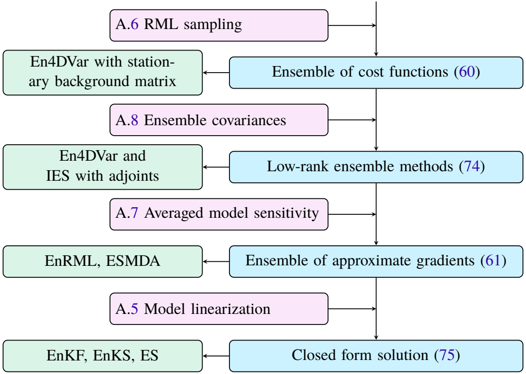
The linearization in Approx. 5 leads to closed form solution of (50)

$$\mathbf{z}^{\mathrm{a}} = \mathbf{z}^{\mathrm{f}} + \mathbf{C}_{zz}\mathbf{G}^{\mathrm{T}}\left(\mathbf{G}\mathbf{C}_{zz}\mathbf{G}^{\mathrm{T}} + \mathbf{C}_{dd}\right)^{-1}\left(\mathbf{d} - \mathbf{g}\left(\mathbf{z}^{\mathrm{f}}\right)\right). \tag{59}$$

Overview of approximations and methods



Overview of approximations and methods



Randomized Maximum Likelihood sampling

Approximation 6 (RML sampling)

In the weakly nonlinear case, we can approximately sample the posterior pdf with Gaussian priors by minimizing the ensemble of cost functions defined by Eq. (60).

ps: it's really Randomized MAP sampling, or rather just approximate sampling of the posterior pdf.

RML minimizes and ensemble of cost functions

Ensemble of cost functions

$$\mathcal{J}(\mathbf{z}_j) = \frac{1}{2}\left(\mathbf{z}_j - \mathbf{z}_j^{\mathrm{f}}\right)^{\mathrm{T}} \mathbf{C}_{zz}^{-1}\left(\mathbf{z}_j - \mathbf{z}_j^{\mathrm{f}}\right) + \frac{1}{2}\left(\mathbf{g}(\mathbf{z}_j) - \mathbf{d}_j\right)^{\mathrm{T}} \mathbf{C}_{dd}^{-1}\left(\mathbf{g}(\mathbf{z}_j) - \mathbf{d}_j\right), \tag{60}$$

Ensemble of gradients set to zero

$$\mathbf{C}_{zz}^{-1}\left(\mathbf{z}_j - \mathbf{z}_j^{\mathrm{f}}\right) + \nabla_{\mathbf{z}}\mathbf{g}\left(\mathbf{z}_j\right)\mathbf{C}_{dd}^{-1}\left(\mathbf{g}(\mathbf{z}_j) - \mathbf{d}_j\right) = 0. \tag{61}$$

Thus, we must solve N independent minimizations.

Solutions methods using the tangent linear model $\mathbf{G}$

Ensemble of incremental 4DVars

$$\mathcal{J}(\delta\mathbf{z}_j) = \frac{1}{2}\left(\delta\mathbf{z}_j - \boldsymbol{\xi}_j^i\right)^{\mathrm{T}} \mathbf{C}_{zz}^{-1}\left(\delta\mathbf{z}_j - \boldsymbol{\xi}_j^i\right) + \frac{1}{2}\left(\mathbf{G}_j^i\delta\mathbf{z}_j - \boldsymbol{\eta}_j^i\right)^{\mathrm{T}} \mathbf{C}_{dd}^{-1}\left(\mathbf{G}_j^i\delta\mathbf{z}_j - \boldsymbol{\eta}_j^i\right). \tag{62}$$

Ensemble of GN iterations

$$\mathbf{z}_j^{i+1} = \mathbf{z}_j^i - \gamma\left(\mathbf{C}_{zz}^{-1} + \mathbf{G}_j^{i\,\mathrm{T}}\mathbf{C}_{dd}^{-1}\mathbf{G}_j^i\right)^{-1}\left(\mathbf{C}_{zz}^{-1}\left(\mathbf{z}_j^i - \mathbf{z}_j^{\mathrm{f}}\right) + \mathbf{G}_j^{i\,\mathrm{T}}\mathbf{C}_{dd}^{-1}\left(\mathbf{g}\left(\mathbf{z}_j^i\right) - \mathbf{d}_j\right)\right), \tag{63}$$

The linear Approximation 5 leads to an Ensemble of Kalman-filter updates

$$\mathbf{z}_j^{\mathrm{a}} = \mathbf{z}_j^{\mathrm{f}} + \mathbf{C}_{zz}\mathbf{G}_j^{\mathrm{T}}\left(\mathbf{G}_j\mathbf{C}_{zz}\mathbf{G}_j^{\mathrm{T}} + \mathbf{C}_{dd}\right)^{-1}\left(\mathbf{d}_j - \mathbf{g}\left(\mathbf{z}_j^{\mathrm{f}}\right)\right). \tag{64}$$

Use an averaged model sensitivity to avoid adjoints

Approximation 7 (Best-fit averaged model sensitivity)

Interpret $\mathbf{G}_j$ in Eq. (64) and $\mathbf{G}_j^i$ in Eq. (63) as the sensitivity matrix in linear regression and represent them using the definition

$$\mathbf{G}_j \approx \mathbf{G} \triangleq \mathbf{C}_{yz}\mathbf{C}_{zz}^{-1}. \tag{65}$$

We approximate the individual model sensitivities with a common averaged sensitivity used for all realizations.

Ensemble representation of covariances

Approximation 8 (Ensemble approximation)

It is possible to approximately represent a covariance matrix by a low-rank ensemble of states with fewer realizations than the state dimension.

$$\mathbf{Z} = \left(\mathbf{z}_1,\ \mathbf{z}_2,\ \dots\ ,\ \mathbf{z}_N\right). \tag{66}$$

Define the projection $\mathbf{\Pi} \in \Re^{N \times N}$ as

$$\mathbf{\Pi} = \left(\mathbf{I} - \frac{1}{N}\mathbf{1}\mathbf{1}^{\mathrm{T}}\right) \Big/ \sqrt{N-1}, \tag{67}$$

Ensemble anomalies

$$\mathbf{A} = \mathbf{Z}\mathbf{\Pi}, \tag{68}$$

Ensemble matrices

Ensemble covariance

$$\overline{\mathbf{C}}_{zz} = \mathbf{A}\mathbf{A}^{\mathrm{T}}. \tag{69}$$

Measurement error covariance matrix

$$\overline{\mathbf{C}}_{dd} = \mathbf{E}\mathbf{E}^{\mathrm{T}}. \tag{70}$$

Ensemble of model-predicted measurements

$$\boldsymbol{\Upsilon} = \mathbf{g}(\mathbf{Z}), \tag{71}$$

with anomalies

$$\mathbf{Y} = \boldsymbol{\Upsilon}\boldsymbol{\Pi}, \tag{72}$$

Solution is confined to ensemble subspace

Search for the solution in the ensemble subspace

$$\mathbf{Z}^{\mathrm{a}} = \mathbf{Z}^{\mathrm{f}} + \mathbf{A}\mathbf{W}. \tag{73}$$

Cost function in ensemble subspace

$$\mathcal{J}(\mathbf{w}_j) = \frac{1}{2}\mathbf{w}_j^{\mathrm{T}}\mathbf{w}_j + \frac{1}{2}\Big(\mathbf{g}\big(\mathbf{z}_j^{\mathrm{f}} + \mathbf{A}\mathbf{w}_j\big) - \mathbf{d}_j\Big)^{\mathrm{T}}\overline{\mathbf{C}}_{dd}^{-1}\Big(\mathbf{g}\big(\mathbf{z}_j^{\mathrm{f}} + \mathbf{A}\mathbf{w}_j\big) - \mathbf{d}_j\Big), \tag{74}$$

Ensemble Kalman filter update

$$\mathbf{Z}^{\mathrm{a}} = \mathbf{Z}^{\mathrm{f}} + \mathbf{A}\mathbf{Y}^{\mathrm{T}}\left(\mathbf{Y}\mathbf{Y}^{\mathrm{T}} + \mathbf{E}\mathbf{E}^{\mathrm{T}}\right)^{-1}\left(\mathbf{D} - \mathbf{g}\left(\mathbf{Z}^{\mathrm{f}}\right)\right). \tag{75}$$

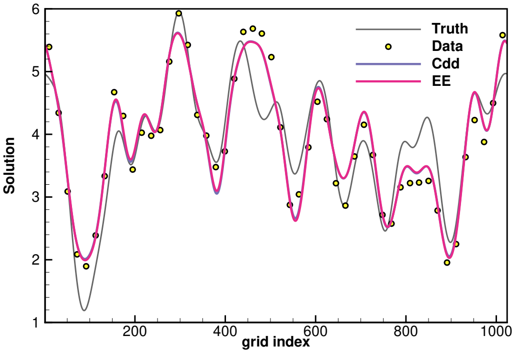
Alternative interpretation using

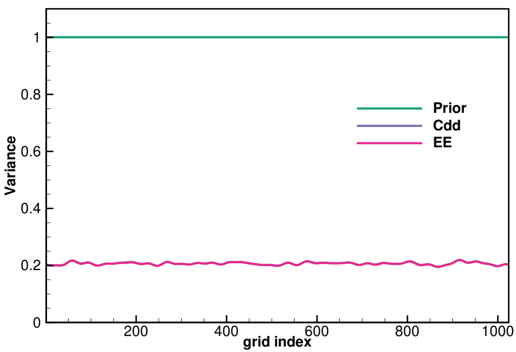
$$\mathbf{W} = \mathbf{Y}^{\mathrm{T}}\left(\mathbf{Y}\mathbf{Y}^{\mathrm{T}} + \mathbf{E}\mathbf{E}^{\mathrm{T}}\right)^{-1}\left(\mathbf{D} - \mathbf{g}\left(\mathbf{Z}^{\mathrm{f}}\right)\right), \tag{76}$$

to get the analysis ensemble from a linear combination of the forecast ensemble

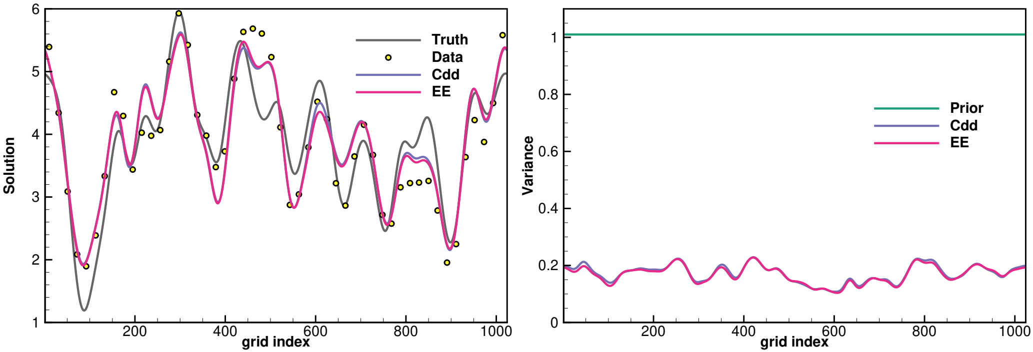
$$\mathbf{Z}^{\mathrm{a}} = \mathbf{Z}^{\mathrm{f}}\left(\mathbf{I} + \mathbf{W}/\sqrt{N-1}\right). \tag{77}$$

EnKF update with large ensemble size $N = 2000$ and $m = 50$

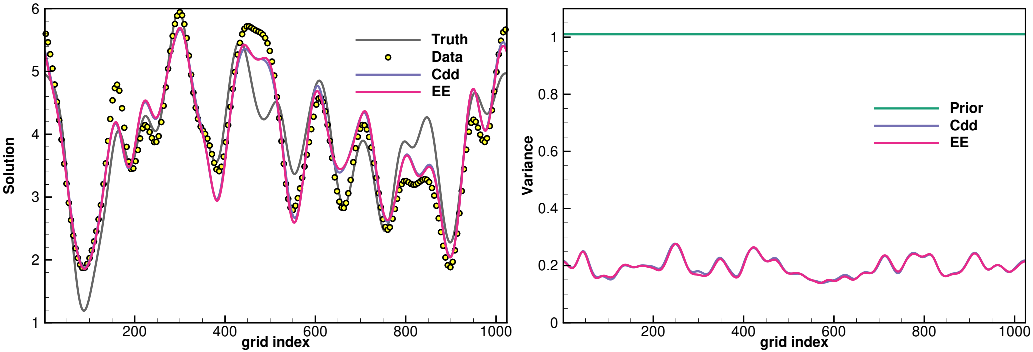




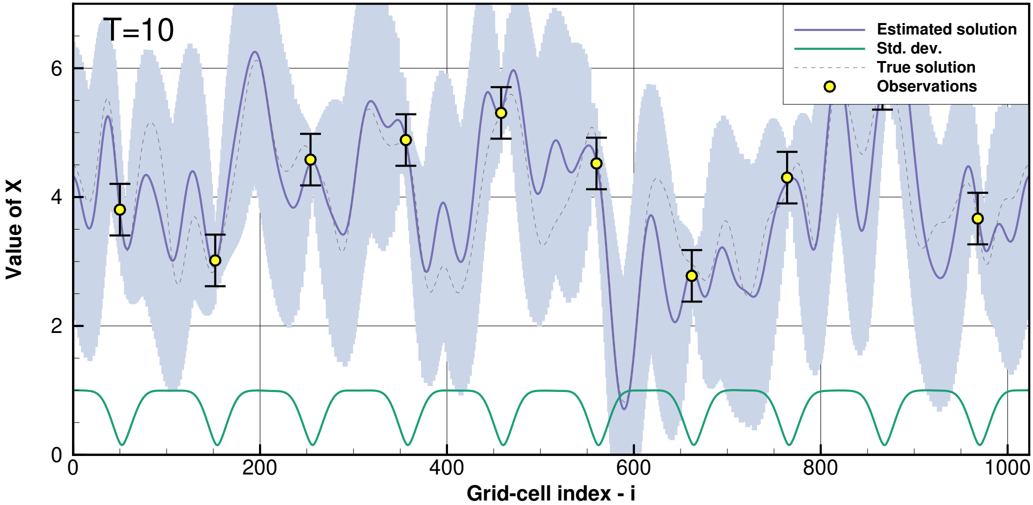
EnKF update with normal ensemble size $N = 100$ and $m = 50$



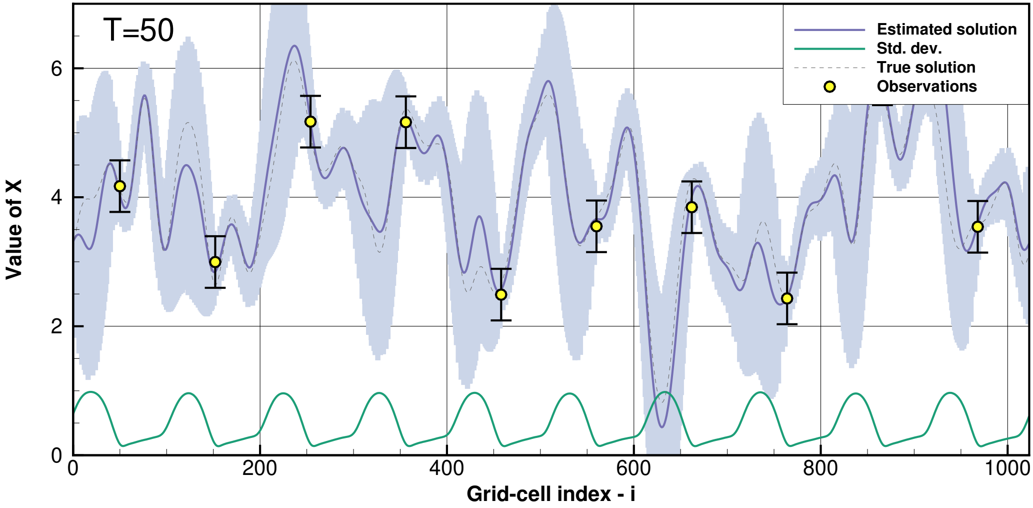
EnKF update with $N = 100$ and many measurements $m = 200$



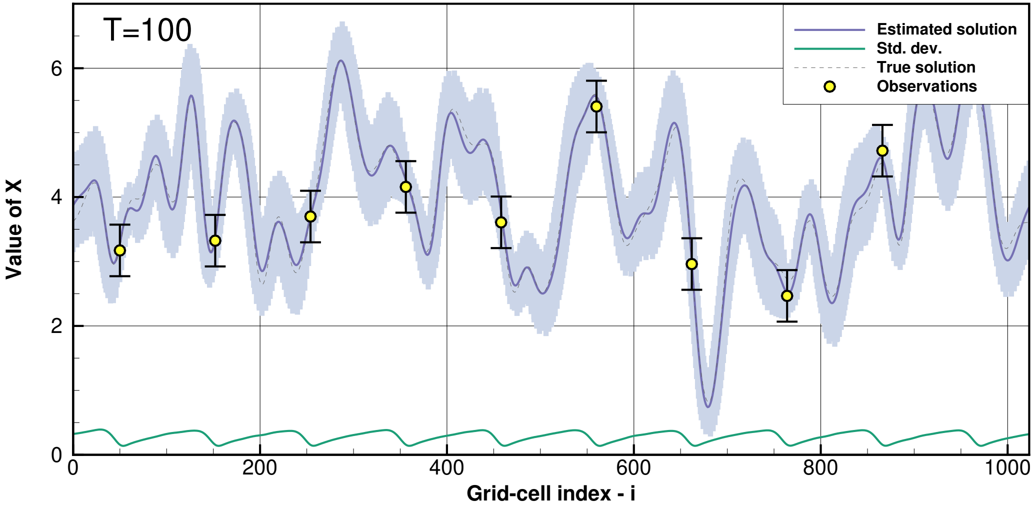
EnKF with the advection equation after two updates



EnKF with the advection equation after ten updates



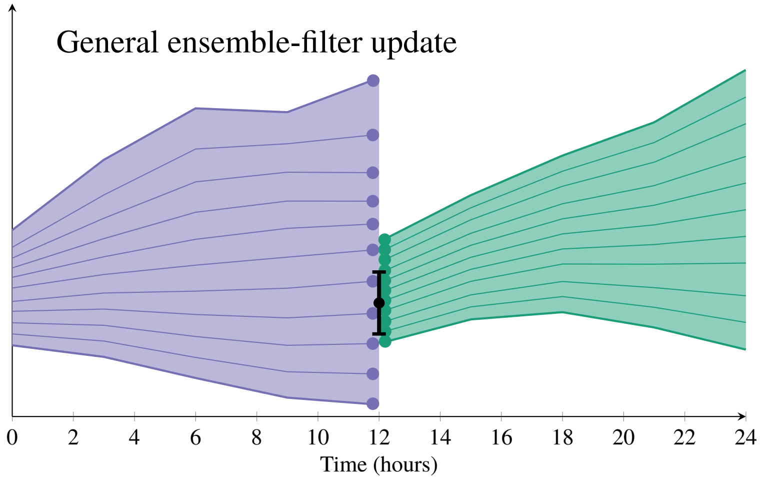
EnKF with the advection equation after twenty updates



EnKF with the advection equation: Animation

Animation

General filter formulation



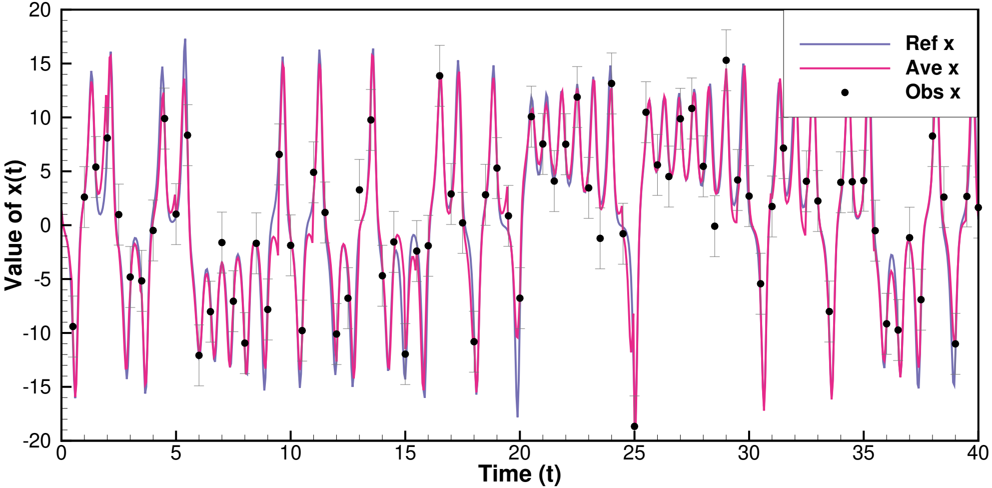
EnKF with the Lorenz model

$$\frac{\partial x}{\partial t} = \sigma(y - x), \tag{78}$$

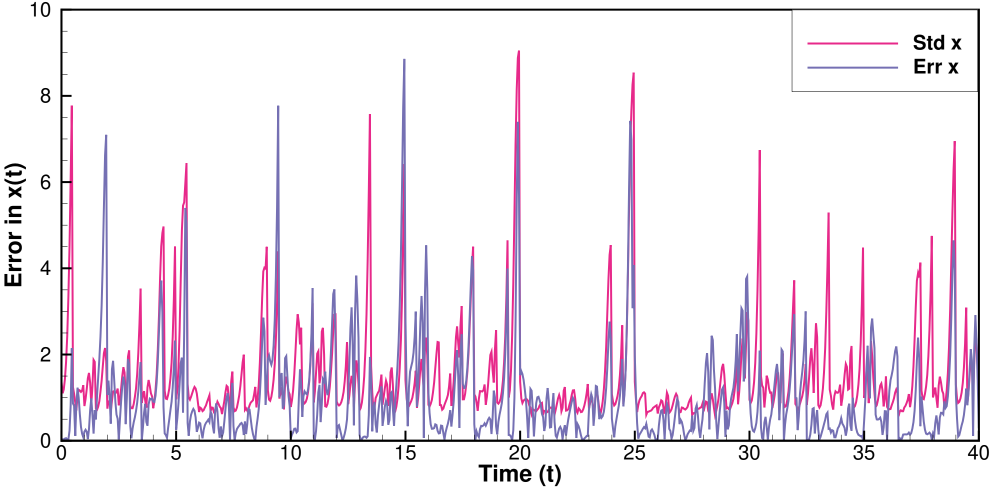
$$\frac{\partial y}{\partial t} = \rho x - y - xz, \tag{79}$$

$$\frac{\partial z}{\partial t} = xy - \beta z. \tag{80}$$

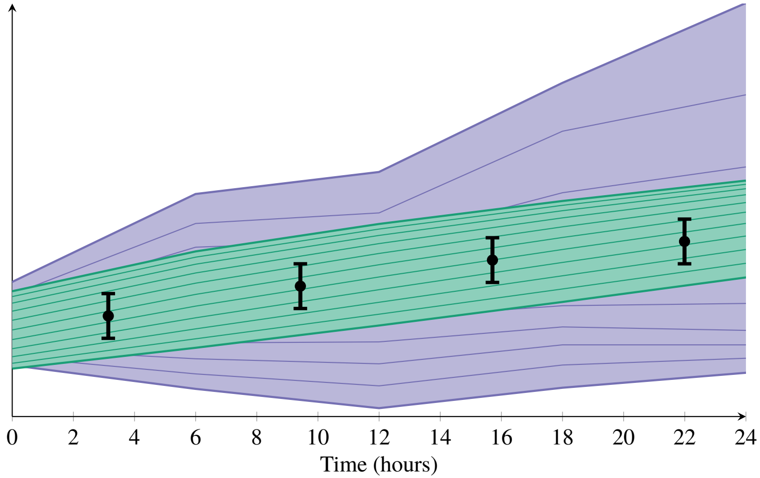
EnKF with the Lorenz model: estimate



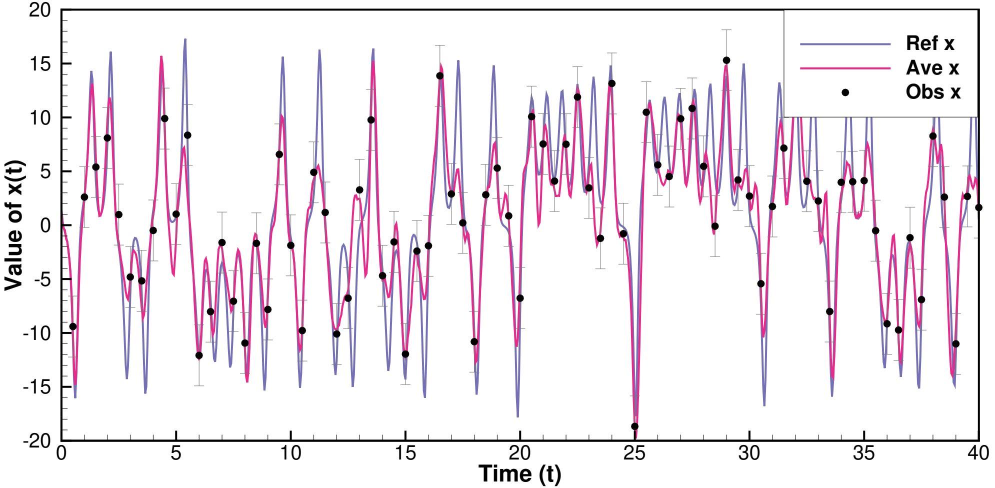
EnKF with the Lorenz model: error estimate



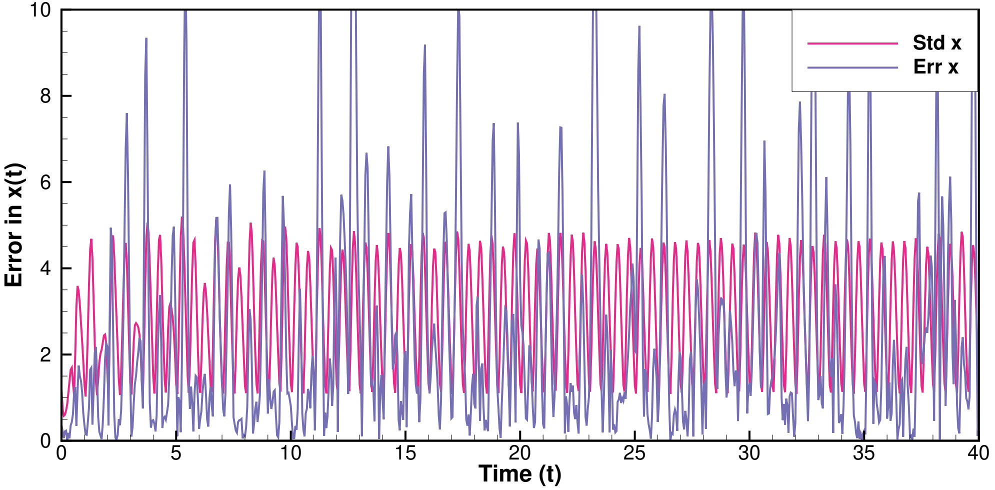
General smoother formulation



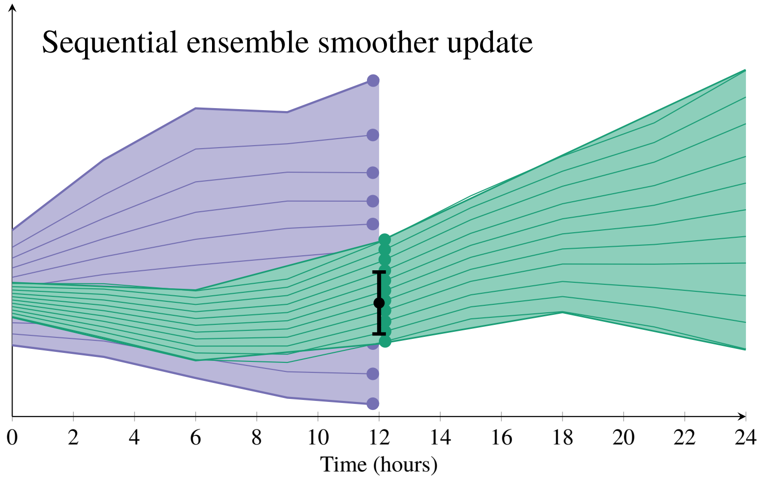
ES with the Lorenz model: estimate



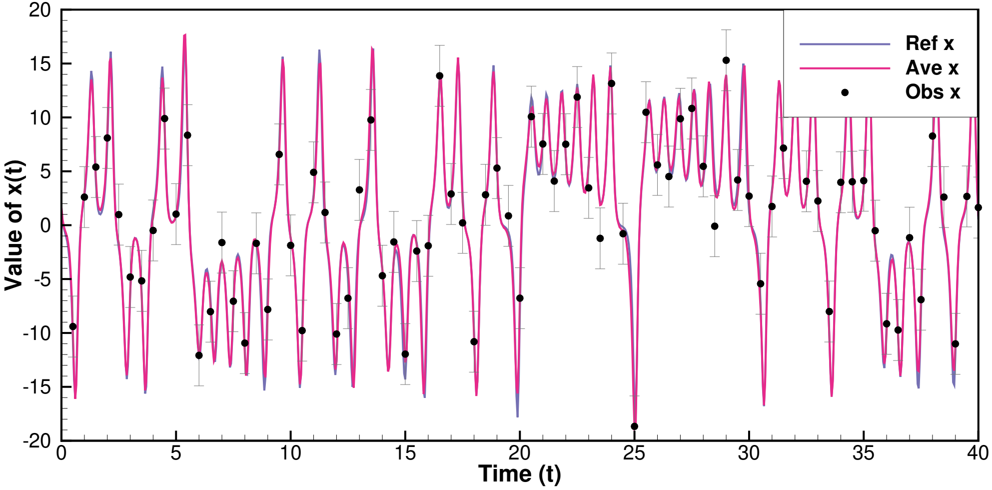
ES with the Lorenz model: error estimate



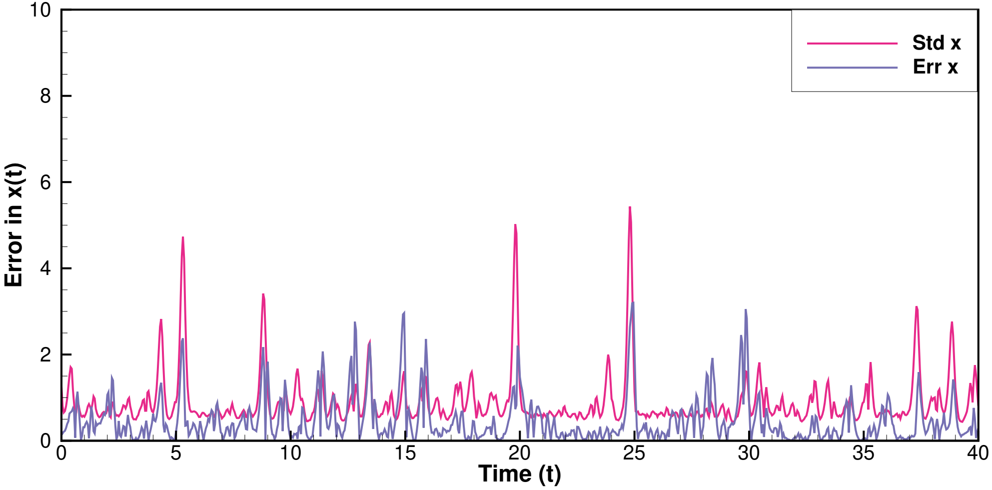
Recursive smoother formulation



EnKS with the Lorenz model: estimate



EnKS with the Lorenz model: error estimate



Summary

Presentation follows new text book on data assimilation:

- Top-down approach for deriving the most popular methods from Bayes'.
- We introduced the important assumptions and applied approximations.
- We illustrated ensemble methods on simple problems.
- Next seminar will present iterative ensemble smoothers and their applications.
- Codes for examples are available from `https://github.com/geirev/`

Particle filters and flows

Approximation 9 (Particle representation of the pdfs)

It is possible to approximate a probability density function by a finite ensemble of N model states (or particles) as

$$f(\mathbf{z}) \approx \sum_{j=1}^{N} \frac{1}{N}\delta(\mathbf{z}-\mathbf{z}_j), \tag{81}$$

where $\delta(\cdot)$ denotes the Dirac-delta function.

Scalar models

$$y = g(x, q) = x + 0.3x^3 + q, \tag{82}$$

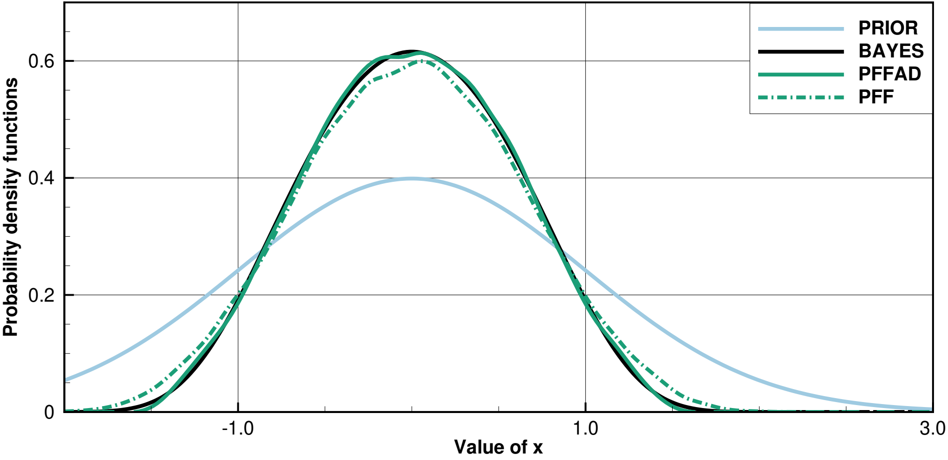
$$y = g(x, q) = 1 + \mathsf{sin}(x) + q, \tag{83}$$

Scalar examples

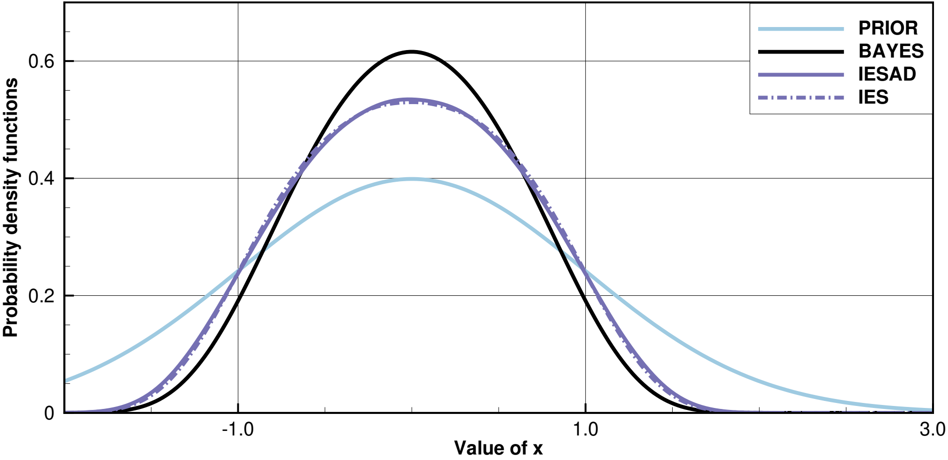
1. Model Eq. (82), $x_j^{\mathrm{f}} = \mathcal{N}(x^{\mathrm{f}} = 0.0, C_{xx} = 1.0)$, $d_j = \mathcal{N}(0.0, 1.0)$, and $q = \mathcal{N}(0, C_{qq} = 0.25)$.
2. Model Eq. (82), $x_j^{\mathrm{f}} = \mathcal{N}(x^{\mathrm{f}} = 1.0, C_{xx} = 1.0)$, $d_j = \mathcal{N}(0.0, -1.0)$, and $q = \mathcal{N}(0, C_{qq} = 0.25)$.
3. Model Eq. (83), $x_j^{\mathrm{f}} = \mathcal{N}(x^{\mathrm{f}} = 1.0, C_{xx} = 1.0)$, $d_j = \mathcal{N}(0.0, 1.0)$, and $q = \mathcal{N}(0, C_{qq} = 0.25)$.

Ensemble size $N = 10^7$.

Case 1: Particle flow

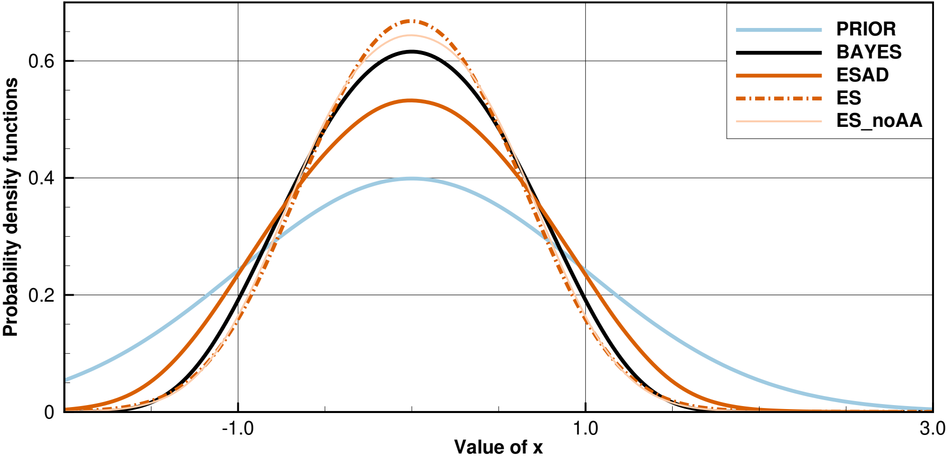


Case 1: IES



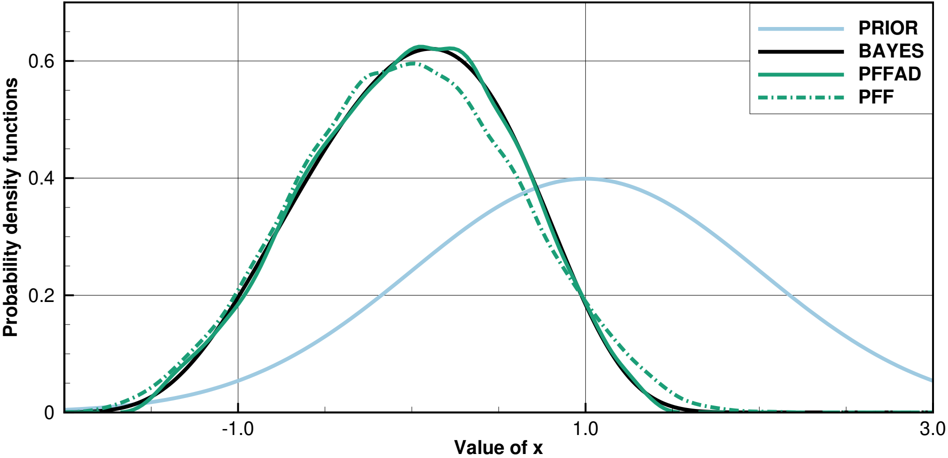
PRIOR
BAYES
IESAD
IES
0.6
0.4
0.2
0
-1.0
1.0
3.0
Probability density functions
Value of x

Case 1: EnKF

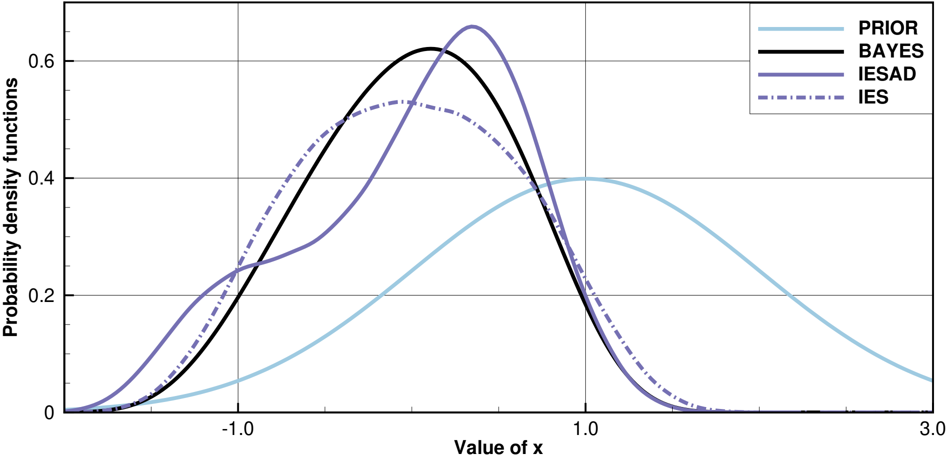


PRIOR
BAYES
ESAD
ES
ES_noAA
Probability density functions
0.6
0.4
0.2
0
-1.0
1.0
3.0
Value of x

Case 2: Particle flow

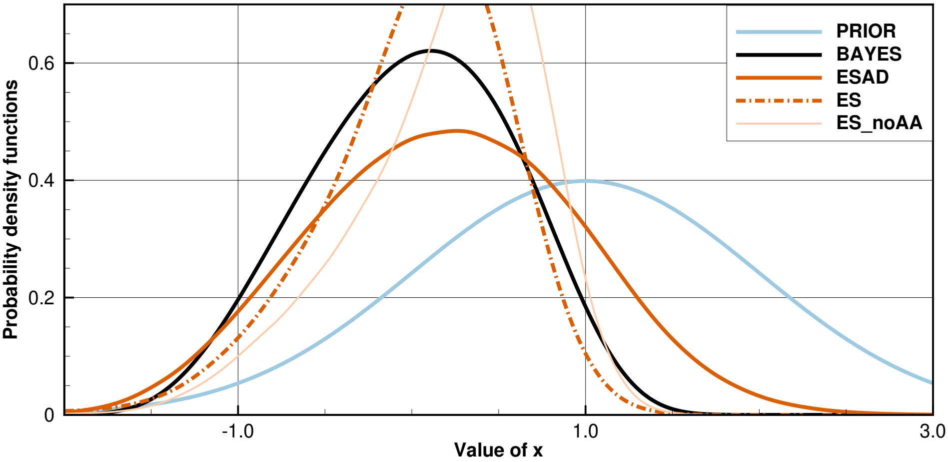


Case 2: IES



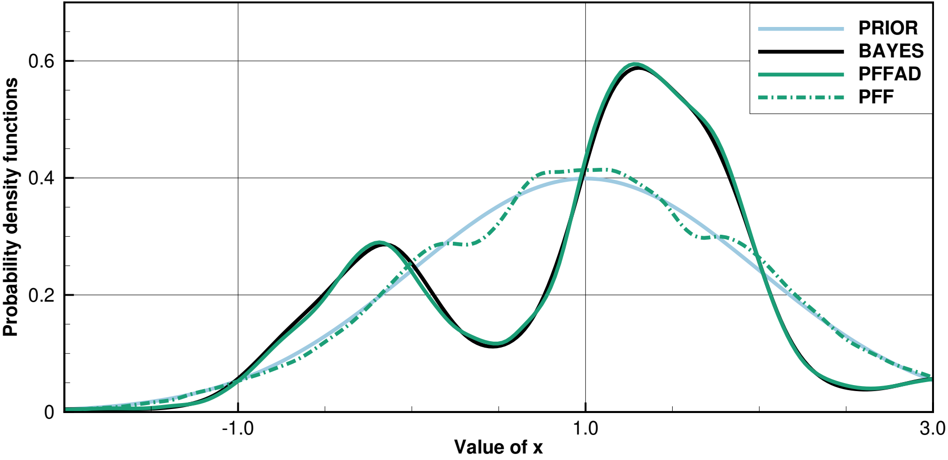
PRIOR
BAYES
IESAD
IES
Probability density functions
0
0.2
0.4
0.6
-1.0
1.0
3.0
Value of x

Case 2: EnKF

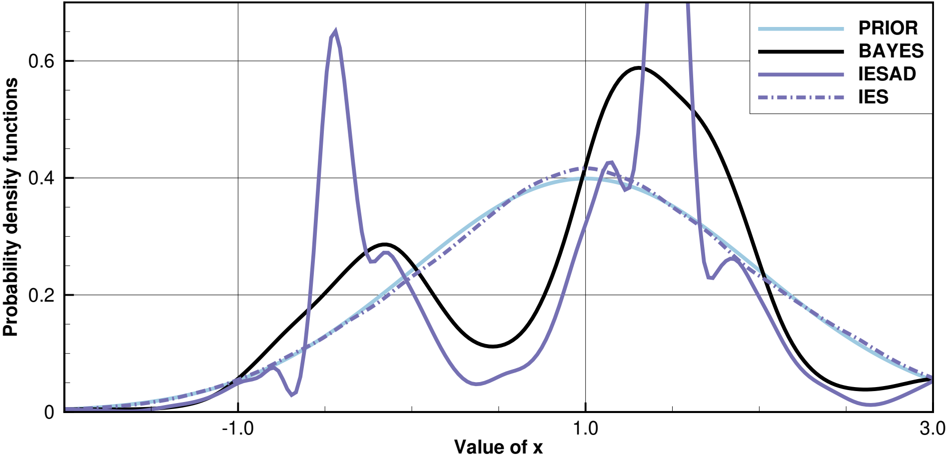


PRIOR
BAYES
ESAD
ES
ES_noAA
Probability density functions
0.6
0.4
0.2
0
-1.0
1.0
3.0
Value of x

Case 3: Particle flow

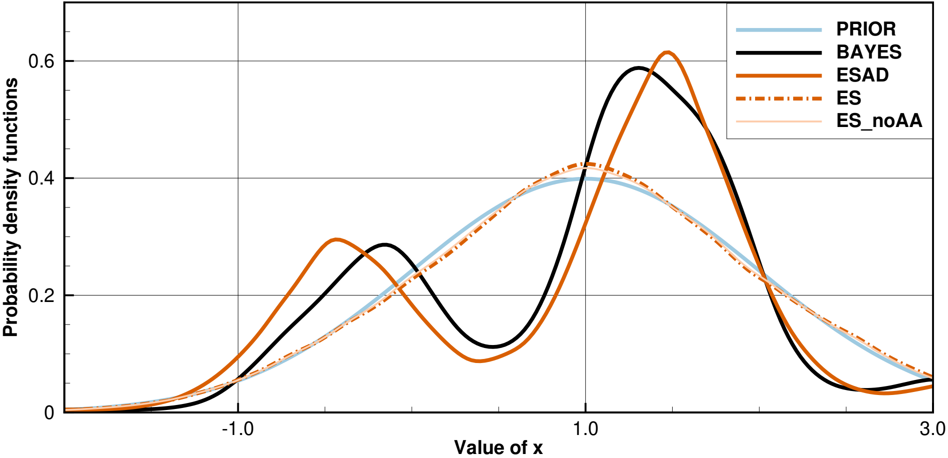


Case 3: IES



PRIOR
BAYES
IESAD
IES
Probability density functions
0.6
0.4
0.2
0
-1.0
1.0
3.0
Value of x

Case 3: EnKF


PRIOR
BAYES
ESAD
ES
ES_noAA
Probability density functions
0
0.2
0.4
0.6
-1.0
1.0
3.0
Value of x

Summary

- New text book on data assimilation.
- Top-down approach for deriving the most popular methods from Bayes'.
- Introduces the important assumptions and applied approximations.
- Discusses methods based on what they solve and and which approximations they use.
- Discusses the severity of the approximations.
- Many simple examples illustrate the various methods.