

Data-Assimilation Fundamentals:

A Unified Formulation of the State and Parameter Estimation Problem

Geir Evensen, Femke C. Vossepoel, and Peter Jan van Leeuwen



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

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.

Evensen · Vossepoel · Leeuwen



Data Assimilation Fundamentals

Except where otherwise noted, this book is licensed under a Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.



► springer.com



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 example from Evensen (2009)

Simple scalar DA example

Given the problem

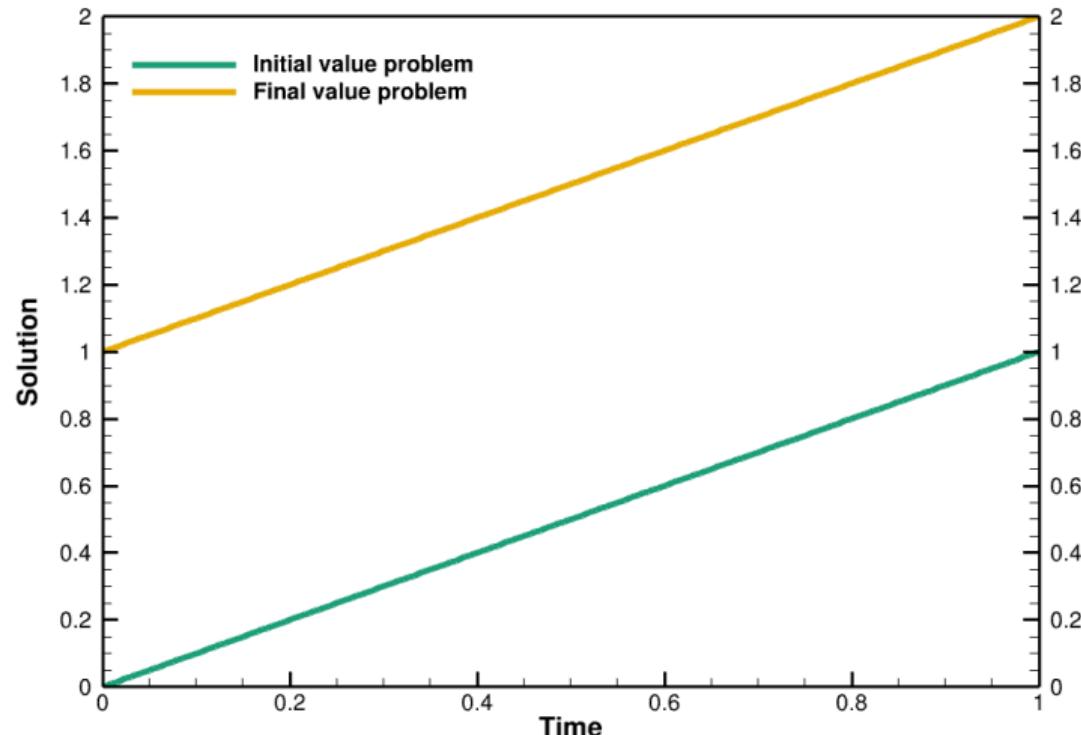
$$\frac{dx}{dt} = 1 \quad \text{Linear model} \quad (1)$$

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

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

- Overdetermined problem.
- No solution.

Initial- and final-value solutions



Allowing for errors in model and conditions

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

$$x(0) = 0 + a \quad (5)$$

$$x(1) = 2 + b \quad (6)$$

- Underdetermined problem.
- Infinitively many solutions.

Impose a statistical assumption on the error terms

The mean is zero:

$$\overline{q(t)} = 0,$$

$$\overline{a} = 0,$$

$$\overline{b} = 0,$$

The variance is known:

$$\overline{q(t_1)q(t_2)} = C\delta(t_1 - t_2),$$

$$\overline{a^2} = C,$$

$$\overline{b^2} = C,$$

No cross correlations:

$$\overline{q(t)a} = 0, \quad (7)$$

$$\overline{ab} = 0, \quad (8)$$

$$\overline{q(t)b} = 0. \quad (9)$$

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

$$\mathcal{J}[x] = C^{-1} \int_0^1 \left(\frac{dx}{dt} - 1 \right)^2 dt + C^{-1} (x(0) - 0)^2 + C^{-1} (x(1) - 2)^2 \quad (10)$$

Then x is an extremum if

$$\delta \mathcal{J}[x] = \mathcal{J}[x + \delta x] - \mathcal{J}[x] = O(\delta x^2) \quad (11)$$

when $\delta x \rightarrow 0$.

$$\mathcal{J}[x + \delta x] = C^{-1} \int_0^1 \left(\frac{dx}{dt} - 1 + \frac{d\delta x}{dt} \right)^2 dt + C^{-1} (x(0) - 0 + \delta x(0))^2 + C^{-1} (x(1) - 2 + \delta x(1))^2 \quad (12)$$

Variation of cost function gives

$$\int_0^1 \frac{d\delta x}{dt} \left(\frac{dx}{dt} - 1 \right) dt + \delta x(0)(x(0) - 0) + \delta x(1)(x(1) - 2) = 0, \quad (13)$$

From integration by part we get

$$\delta x \left(\frac{dx}{dt} - 1 \right) \Big|_0^1 - \int_0^1 \delta x \frac{d^2 x}{dt^2} dt + \delta x(0)(x(0) - 0) + \delta x(1)(x(1) - 2) = 0. \quad (14)$$

Minimum of cost function

This gives the following system of equations

$$\delta x(0) \left. \left(-\frac{dx}{dt} + 1 + x \right) \right|_{t=0} = 0, \quad (15)$$

$$\delta x(1) \left. \left(\frac{dx}{dt} - 1 + x - 2 \right) \right|_{t=1} = 0, \quad (16)$$

$$\int_0^1 \delta x \left(\frac{d^2x}{dt^2} \right) dt = 0, \quad (17)$$

or since δx is arbitrary....

Euler-Lagrange equation

The Euler–Lagrange equation

$$\frac{d^2x}{dt^2} = 0, \tag{18}$$

$$\frac{dx}{dt} - x = 1 \quad \text{for } t = 0, \tag{19}$$

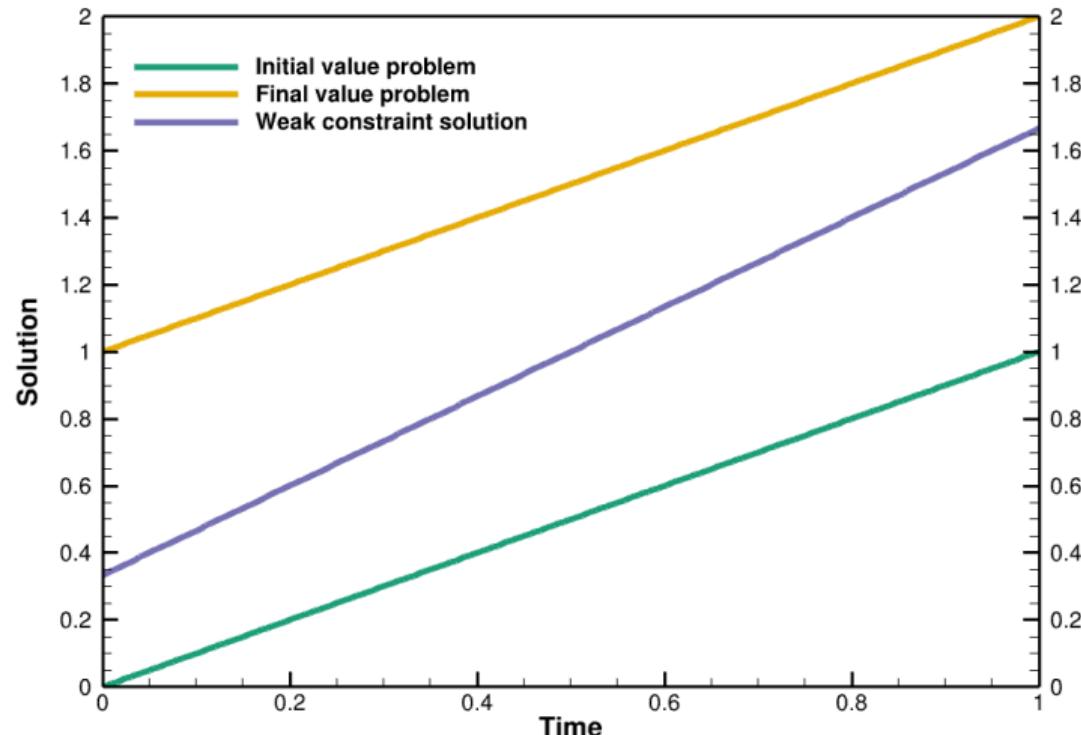
$$\frac{dx}{dt} + x = 3 \quad \text{for } t = 1, \tag{20}$$

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

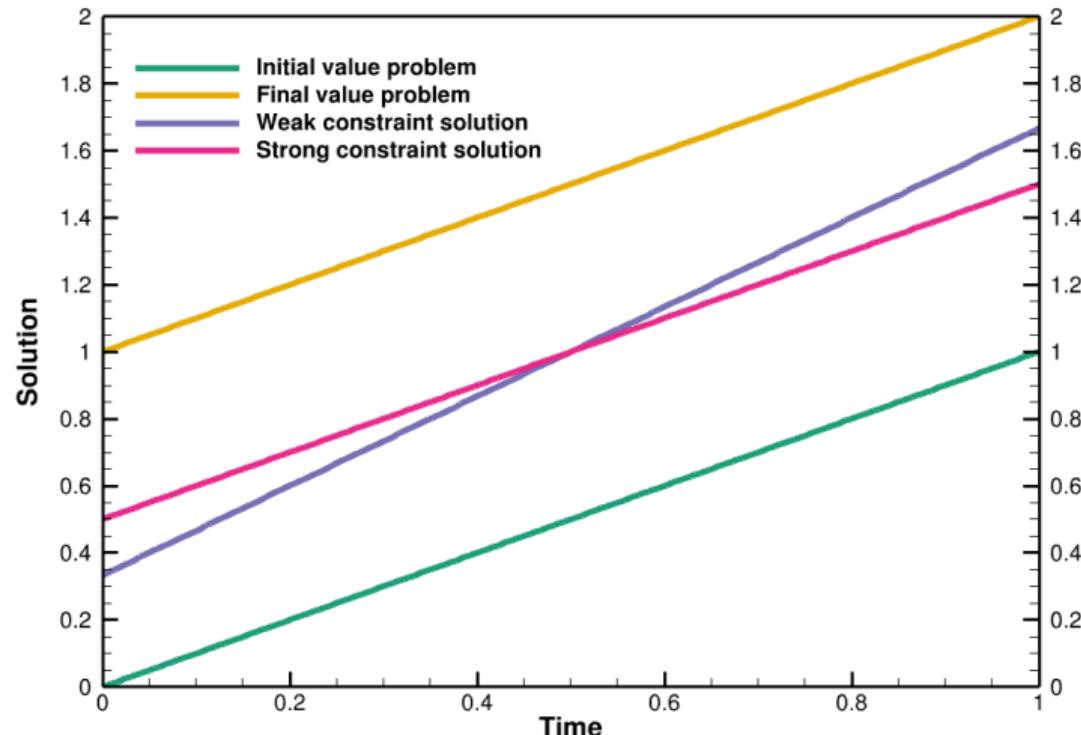
$$x = c_1 t + c_2, \tag{21}$$

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.
- 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 with Gaussian priors

Initial conditions

$$f(x(0)) \propto \exp\left(-\frac{a^2}{C}\right) = \exp\left(-\frac{(x(0) - 0)^2}{C}\right) \quad (22)$$

Model evolution

$$f(x|x(0)) \propto \exp\left(-\int_0^1 \frac{q^2}{C} dt\right) = \exp\left(-\int_0^1 \frac{1}{C} \left(\frac{dx}{dt} - 1\right)^2 dt\right) \quad (23)$$

A measurement

$$f(d|x(1)) \propto \exp\left(-\frac{b^2}{C}\right) = \exp\left(-\frac{(x(1) - 2)^2}{C}\right) \quad (24)$$

Bayes' theorem for the scalar model with Gaussian priors

$$f(x|d) = \frac{f(d|x)f(x)}{f(d)} \quad (25)$$

$$\propto f(d|x(1)) \left(f(x|x(0))f(x(0)) \right) \quad (26)$$

$$= \exp\left(-\frac{(x(1) - 2)^2}{C}\right) \exp\left(-\int_0^1 \frac{1}{C} \left(\frac{dx}{dt} - 1\right)^2 dt\right) \exp\left(-\frac{(x(0) - 0)^2}{C}\right) \quad (27)$$

$$= \exp(-\mathcal{J}[x]) \quad (28)$$

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

We start from Bayes formula

Why Bayes Theorem?

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

Properties of a probability density 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

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

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

Bayes' theorem

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

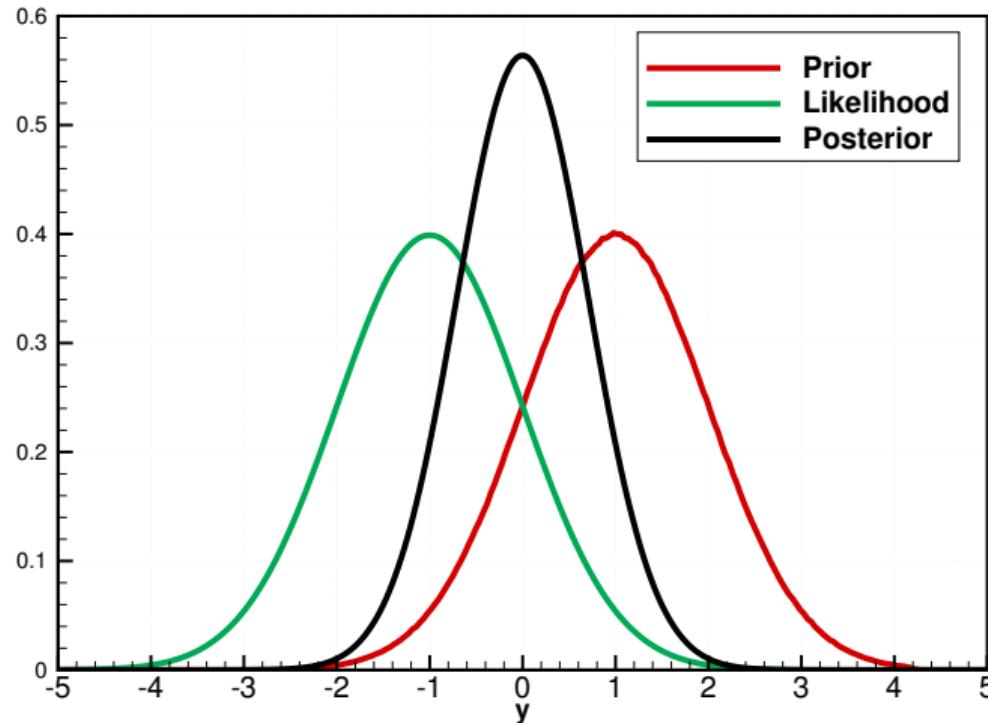
What is the likelihood function: $f(d|x)$

- 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.

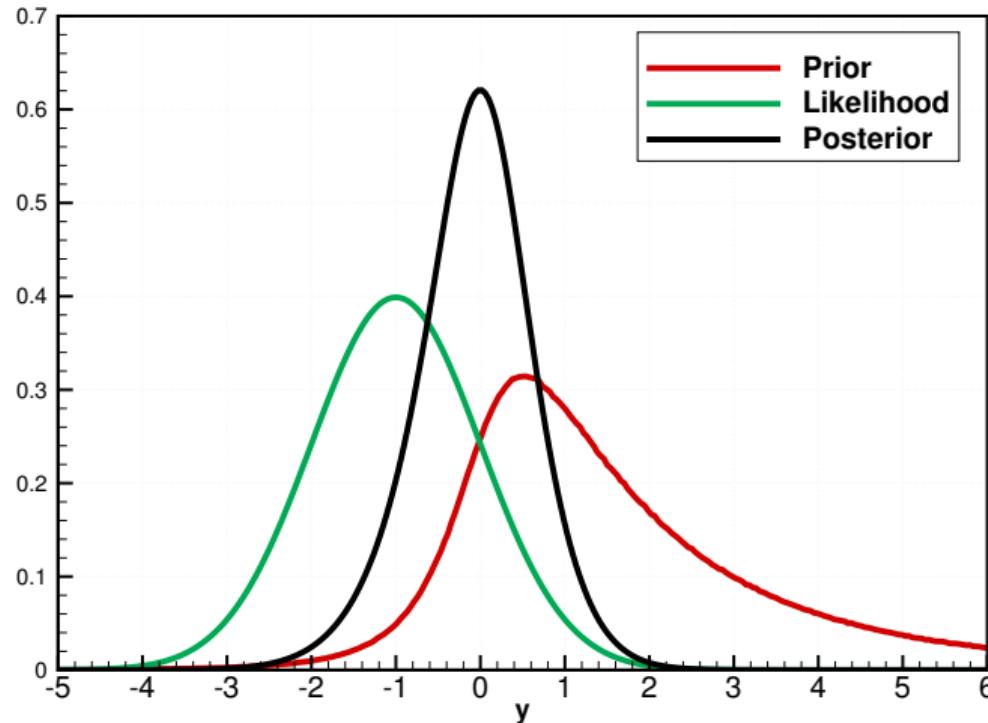
Likelihood is the plausibility of a particular distribution explaining the given data. The higher the likelihood of a distribution, the more likely it is to explain the observed data.

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



Example of using Bayes' theorem



Bayes' formula is the optimal starting point

Bayes' formula 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.

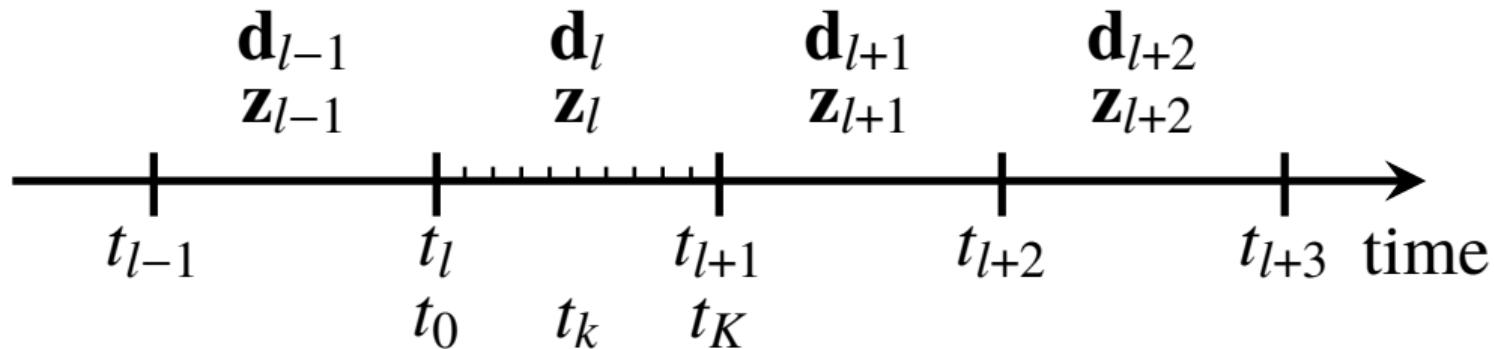
One of the strengths of Bayes' formula is that it does not try to solve the ill-defined problem of “inverting observations” but instead updates prior knowledge.

We start from Bayes' theorem

$$f(\mathcal{Z}|\mathcal{D}) = \frac{f(\mathcal{D}|\mathcal{Z})f(\mathcal{Z})}{f(\mathcal{D})}. \quad (29)$$

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

$$f(\mathbf{z}_l | \mathbf{z}_{l-1}, \mathbf{z}_{l-2}, \dots, \mathbf{z}_0) = f(\mathbf{z}_l | \mathbf{z}_{l-1}), \quad (30)$$

Independent measurements

Approximation 2 (Independent measurements)

We assume that measurements are independent between different assimilation windows.

Independent measurements have uncorrelated errors

$$f(\mathcal{D}|\mathcal{Z}) = \prod_{l=1}^L f(\mathbf{d}_l|\mathbf{z}_l). \quad (31)$$

Bayes becomes

$$f(\mathcal{D}|\mathcal{Z}) \propto \prod_{l=1}^L f(\mathbf{d}_l|\mathbf{z}_l) \prod_{l=1}^L f(\mathbf{z}_l|\mathbf{z}_{l-1}) f(\mathbf{z}_0). \quad (32)$$

Recursive form of Bayes

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

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

$$\vdots \quad (35)$$

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

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

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

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

⋮

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

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

Or just Bayes' for the assimilation window

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

Discrete model with uncertain inputs

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

- \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 $\mathbf{x} = (\mathbf{x}_0, \dots, \mathbf{x}_K)$ as model state over the assimilation window.
- Define $\mathbf{q} = (\mathbf{q}_0, \dots, \mathbf{q}_K)$ as model errors over the assimilation window.
- Define $\mathbf{u} = (\mathbf{u}_0, \dots, \mathbf{u}_K)$ 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

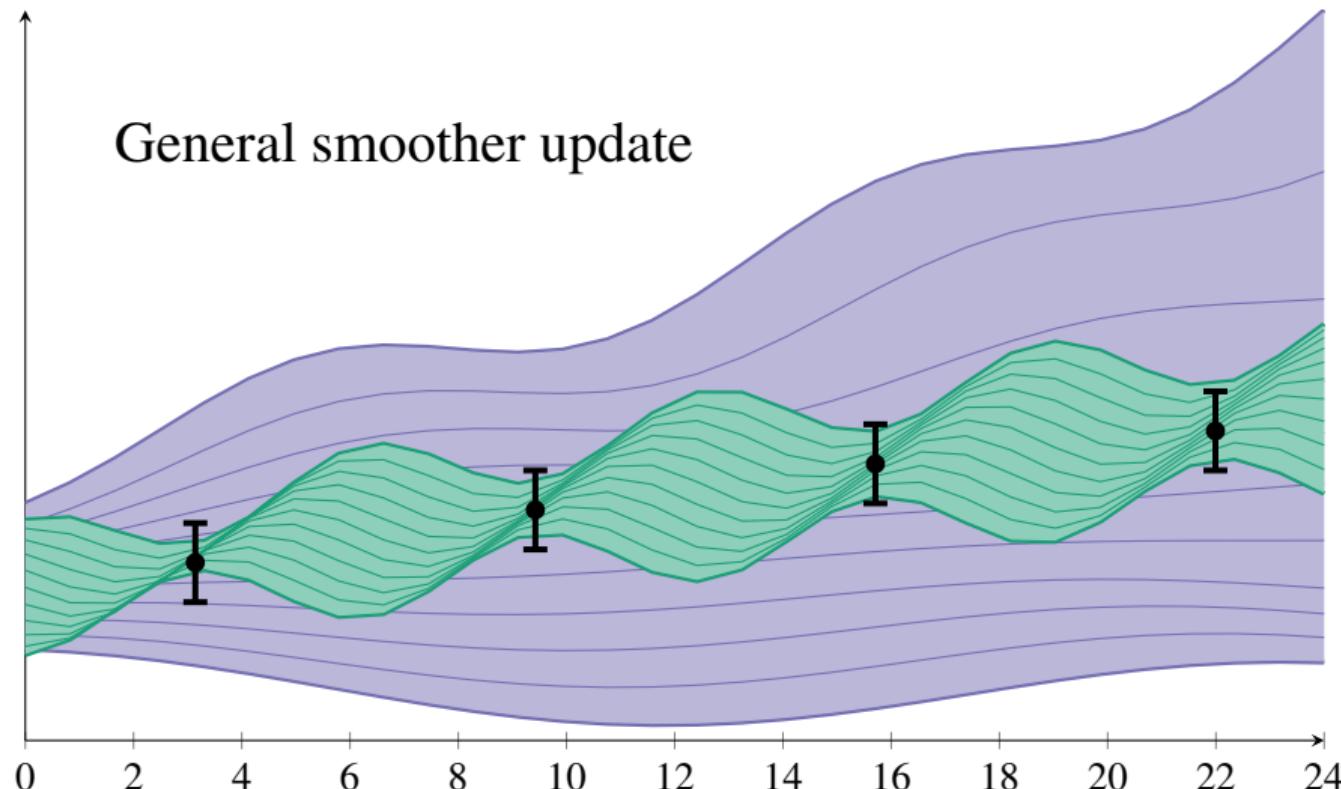
- Solve for model solution over an assimilation window \mathbf{x} .
- Condition on measurements distributed over the assimilation window.

Predicted measurements \mathbf{y}

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

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

General smoother formulation



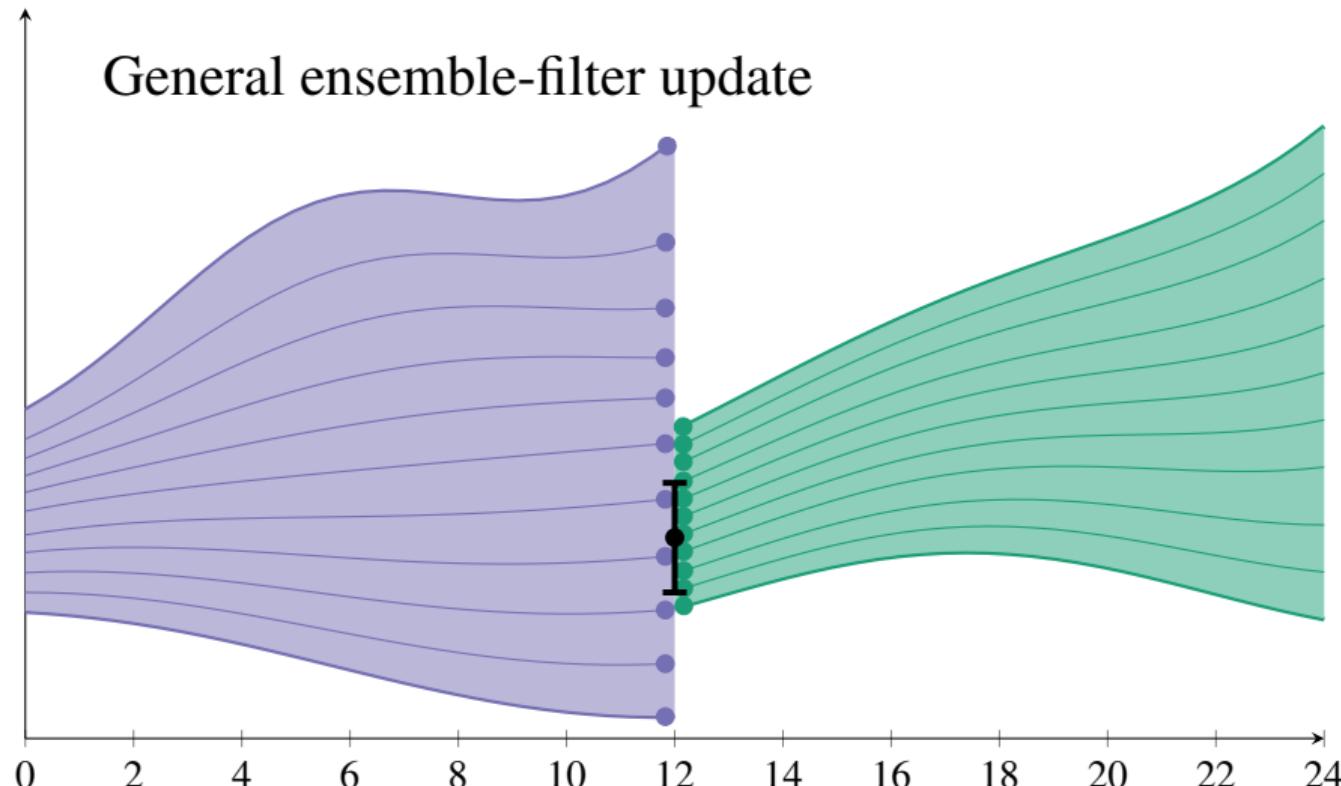
General filter formulation

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

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

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

General filter formulation



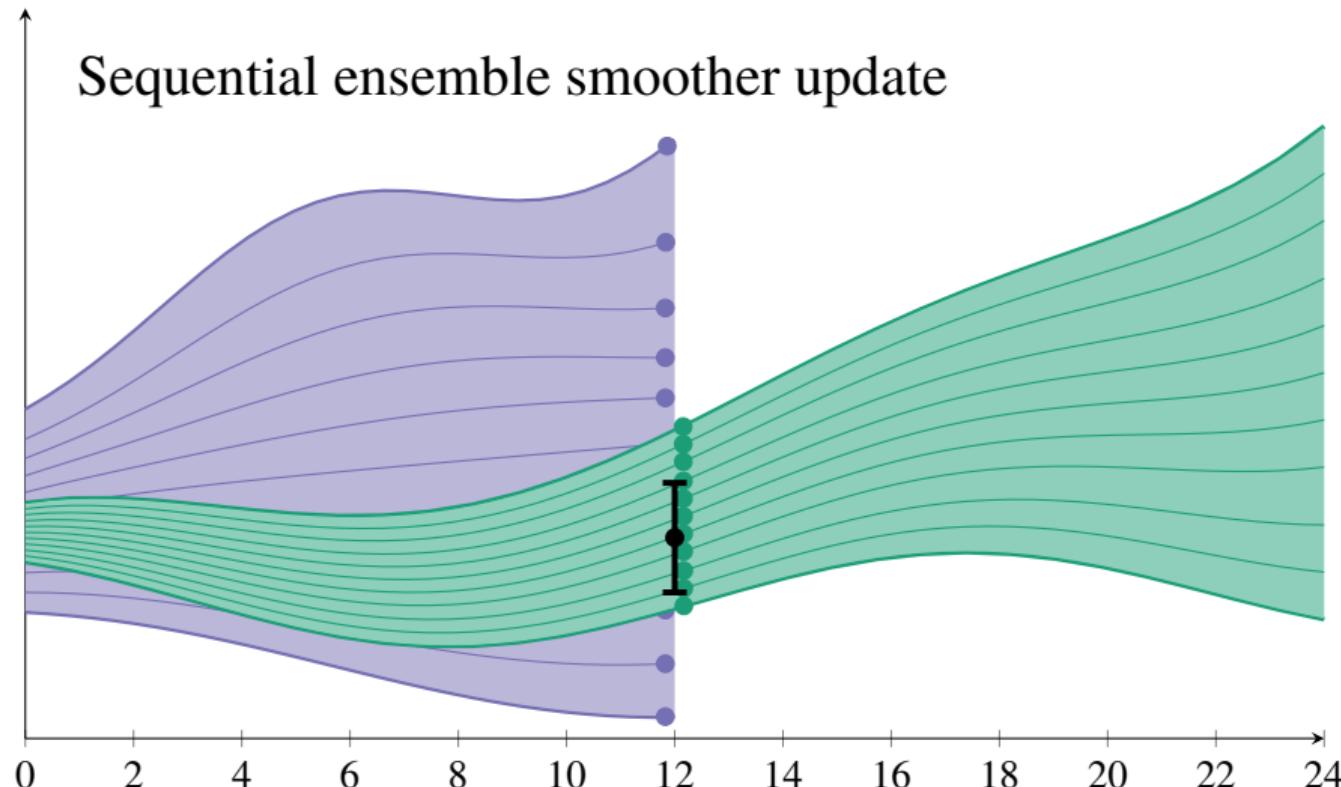
Recursive smoother formulation

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

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

- Ensemble Kalman Smoother (EnKS)

Recursive smoother formulation



Smoothen 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.

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

- Strong constraint 4DVar.
- Iterative ensemble smoothers (EnRML, ESMDA).

Nonlinear measurement functional

- Sequential DA with nonlinear data at the end of the assimilation window.
- Solve for model solution at the end of an assimilation window \mathbf{x}_K .

$$\mathbf{y} = \mathbf{g}(\mathbf{z}) = \mathbf{h}(\mathbf{x}_K). \quad (47)$$

- EnKF.

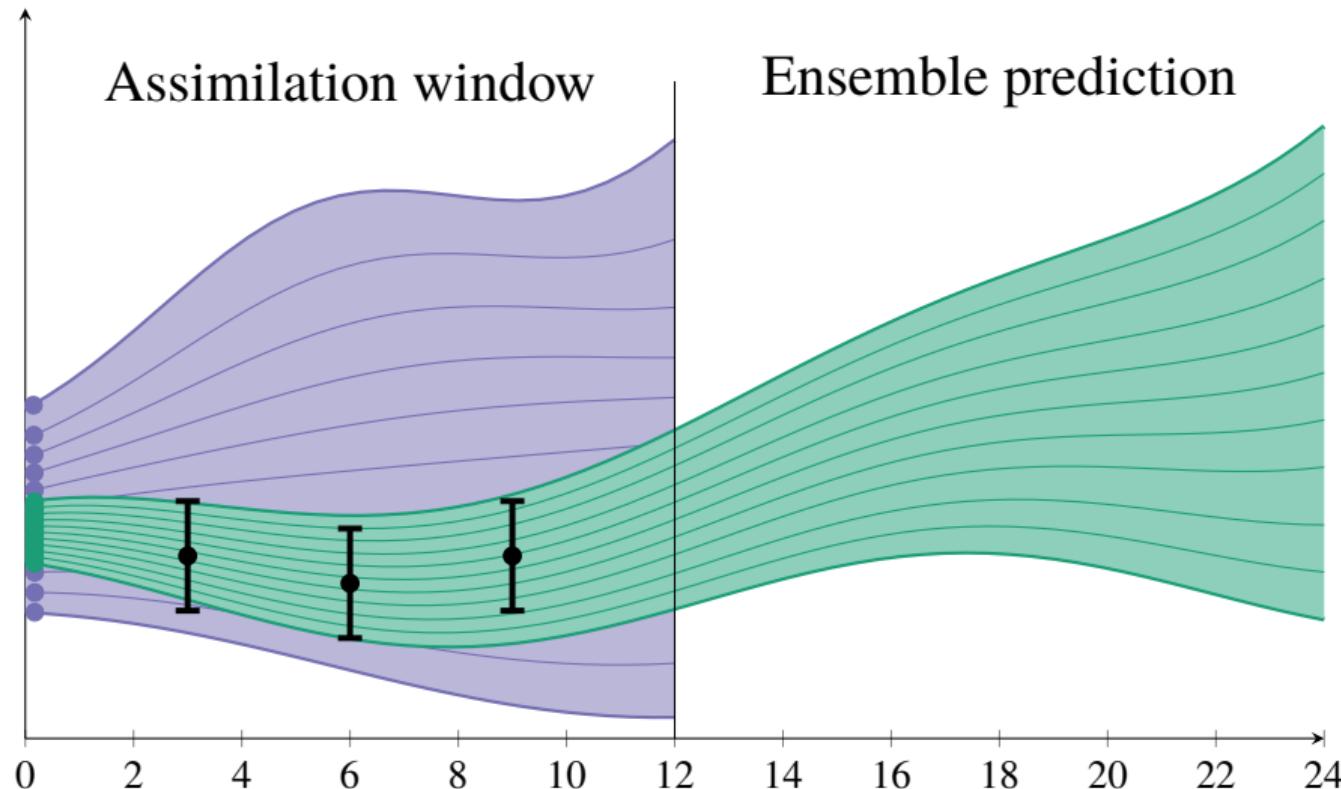
Parameter estimation

- Solve for uncertain input parameters.
- Condition on measurements distributed over the assimilation window.

$$\mathbf{y} = \mathbf{g}(\mathbf{z}) = \mathbf{h}(\mathbf{m}(\boldsymbol{\theta})). \quad (48)$$

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

Smoother for perfect models



Deriving the marginal posterior pdf

Nonlinear “perfect” model and measurements

$$\mathbf{y} = \mathbf{g}(\mathbf{x}) \quad \mathbf{d} \leftarrow \mathbf{y} + \mathbf{e}$$

Bayesian formulation

$$f(\mathbf{x}, \mathbf{y}|\mathbf{d}) \propto f(\mathbf{d}|\mathbf{y})f(\mathbf{y}|\mathbf{x})f(\mathbf{x})$$

Model pdf

$$f(\mathbf{y}|\mathbf{x}) = \delta(\mathbf{y} - \mathbf{g}(\mathbf{x}))$$

Marginal pdf

$$f(\mathbf{x}|\mathbf{d}) \propto \int f(\mathbf{d}|\mathbf{y})f(\mathbf{y}|\mathbf{x})f(\mathbf{x})d\mathbf{y} = f(\mathbf{d}|\mathbf{g}(\mathbf{x}))f(\mathbf{x})$$

Bayes' theorem related to the predicted measurements

We introduce nonlinearity through the likelihood

$$f(\mathbf{z}|\mathbf{d}) = \frac{f(\mathbf{d}|\mathbf{g}(\mathbf{z}))f(\mathbf{z})}{f(\mathbf{d})}. \quad (49)$$

The MAP solution

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.

$$f(\mathbf{z}|\mathbf{d}) \propto \exp\{-\mathcal{J}(\mathbf{z})\}, \quad (50)$$

Leads to a cost-function formulation for the MAP solution

Cost function

$$\mathcal{J}(\mathbf{z}) = \frac{1}{2} (\mathbf{z} - \mathbf{z}^f)^T \mathbf{C}_{zz}^{-1} (\mathbf{z} - \mathbf{z}^f) + \frac{1}{2} (\mathbf{g}(\mathbf{z}) - \mathbf{d})^T \mathbf{C}_{dd}^{-1} (\mathbf{g}(\mathbf{z}) - \mathbf{d}). \quad (51)$$

The gradient set to zero

$$\mathbf{C}_{zz}^{-1} (\mathbf{z}^a - \mathbf{z}^f) + \nabla_{\mathbf{z}} \mathbf{g}(\mathbf{z}^a) \mathbf{C}_{dd}^{-1} (\mathbf{g}(\mathbf{z}^a) - \mathbf{d}) = 0. \quad (52)$$

- There is no explicit solution of the gradient equation.

Gauss-Newton methods solves for the MAP estimate

Gauss-Newton iteration

$$\mathbf{z}^{i+1} = \mathbf{z}^i - \gamma^i \left(\mathbf{C}_{zz}^{-1} + \mathbf{G}^{iT} \mathbf{C}_{dd}^{-1} \mathbf{G}^i \right)^{-1} \left(\mathbf{C}_{zz}^{-1} (\mathbf{z}^i - \mathbf{z}^f) + \mathbf{G}^{iT} \mathbf{C}_{dd}^{-1} (\mathbf{g}(\mathbf{z}^i) - \mathbf{d}) \right). \quad (53)$$

- The incremental formulation is sometimes more convenient.

Incremental Gauss-Newton methods

Quadratic cost function for the increments

$$\mathcal{J}(\delta\mathbf{z}) = \frac{1}{2}(\delta\mathbf{z} - \boldsymbol{\xi}^i)^T \mathbf{C}_{zz}^{-1} (\delta\mathbf{z} - \boldsymbol{\xi}^i) + \frac{1}{2}(\mathbf{G}^i \delta\mathbf{z} - \boldsymbol{\eta}^i)^T \mathbf{C}_{dd}^{-1} (\mathbf{G}^i \delta\mathbf{z} - \boldsymbol{\eta}^i). \quad (54)$$

with

$$\mathbf{z}^{i+1} = \mathbf{z}^i + \delta\mathbf{z}, \quad (55)$$

$$\boldsymbol{\eta}^i = \mathbf{d} - \mathbf{g}(\mathbf{z}^i), \quad (56)$$

$$\boldsymbol{\xi}^i = \mathbf{z}^f - \mathbf{z}^i. \quad (57)$$

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

Standard strong constraint 4DVar

Standard SC-4DVar

Model with initial condition and poorly known parameter

$$\mathbf{x}_0 = \mathbf{x}_0^f + \mathbf{x}'_0, \quad (58)$$

$$\boldsymbol{\theta} = \boldsymbol{\theta}^f + \boldsymbol{\theta}', \quad (59)$$

$$\mathbf{x}_{k+1} = \mathbf{m}(\mathbf{x}_k, \boldsymbol{\theta}), \quad (60)$$

Measurements

$$\mathbf{d} = \mathbf{h}(\mathbf{x}) + \mathbf{e}. \quad (61)$$

Problem formulation

State vector and covariance matrix

$$\mathbf{z} = \begin{pmatrix} \mathbf{x}_0 \\ \boldsymbol{\theta} \end{pmatrix} \quad \text{and} \quad \mathbf{C}_{zz} = \begin{pmatrix} \mathbf{C}_{x_0 x_0} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_{\theta \theta} \end{pmatrix}, \quad (62)$$

SC-4DVar costfunction

$$\mathcal{J}(\mathbf{z}) = \frac{1}{2} (\mathbf{z} - \mathbf{z}^f)^T \mathbf{C}_{zz}^{-1} (\mathbf{z} - \mathbf{z}^f) + \frac{1}{2} (\mathbf{h}(\mathbf{x}) - \mathbf{d})^T \mathbf{C}_{dd}^{-1} (\mathbf{h}(\mathbf{x}) - \mathbf{d}), \quad (63)$$

Solve for initial condition and parameter that minimize the cost function

Lagrangian formulation

$$\begin{aligned}
 \mathcal{L}(\mathbf{x}_0, \dots, \mathbf{x}_{K+1}, \boldsymbol{\theta}, \boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_{K+1}) = & \frac{1}{2} (\mathbf{x}_0 - \mathbf{x}_0^f)^T \mathbf{C}_{x_0 x_0}^{-1} (\mathbf{x}_0 - \mathbf{x}_0^f) \\
 & + \frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\theta}^f)^T \mathbf{C}_{\boldsymbol{\theta} \boldsymbol{\theta}}^{-1} (\boldsymbol{\theta} - \boldsymbol{\theta}^f) \\
 & + \frac{1}{2} (\mathbf{h}(\mathbf{x}) - \mathbf{d})^T \mathbf{C}_{dd}^{-1} (\mathbf{h}(\mathbf{x}) - \mathbf{d}) \\
 & + \sum_{k=0}^K \boldsymbol{\lambda}_{k+1}^T (\mathbf{x}_{k+1} - \mathbf{m}(\mathbf{x}_k, \boldsymbol{\theta})). \tag{64}
 \end{aligned}$$

We include the perfect model by introducing a Lagrangian multiplier $\boldsymbol{\lambda}$.

Gradient of Lagrangian

$$\nabla_{\mathbf{x}_k} \mathcal{L}(\mathbf{x}, \boldsymbol{\theta}, \boldsymbol{\lambda}) = \mathbf{H}_k^T \mathbf{C}_{dd}^{-1} (\mathbf{h}(\mathbf{x}) - \mathbf{d}) + \boldsymbol{\lambda}_k - \mathbf{M}_{x,k}^T \boldsymbol{\lambda}_{k+1}, \quad (65)$$

$$\nabla_{\mathbf{x}_{K+1}} \mathcal{L}(\mathbf{z}, \mathbf{x}, \boldsymbol{\lambda}) = \boldsymbol{\lambda}_{K+1}, \quad (66)$$

$$\begin{aligned} \nabla_{\mathbf{x}_0} \mathcal{L}(\mathbf{x}, \boldsymbol{\theta}, \boldsymbol{\lambda}) &= \mathbf{C}_{zz}^{-1} (\mathbf{x}_0 - \mathbf{x}_0^f) - \mathbf{M}_{x,0}^T \boldsymbol{\lambda}_1 \\ &= \mathbf{C}_{zz}^{-1} (\mathbf{x}_0 - \mathbf{x}_0^f) - \boldsymbol{\lambda}_0, \end{aligned} \quad (67)$$

$$\nabla_{\boldsymbol{\theta}} \mathcal{L}(\mathbf{x}, \boldsymbol{\theta}, \boldsymbol{\lambda}) = \mathbf{C}_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1} (\boldsymbol{\theta} - \boldsymbol{\theta}^f) - \sum_{k=0}^K \mathbf{M}_{\boldsymbol{\theta},k}^T \boldsymbol{\lambda}_{k+1}, \quad (68)$$

$$\nabla_{\boldsymbol{\lambda}_k} \mathcal{L}(\mathbf{x}, \boldsymbol{\theta}, \boldsymbol{\lambda}) = \mathbf{x}_{k+1} - \mathbf{m}(\mathbf{x}_k, \boldsymbol{\theta}). \quad (69)$$

Euler-Lagrange equation(s)

Forward model

$$\mathbf{x}_0 = \mathbf{x}_0^f + \mathbf{C}_{x_0 x_0} \boldsymbol{\lambda}_0, \quad (70)$$

$$\boldsymbol{\theta} = \boldsymbol{\theta}^f + \mathbf{C}_{\boldsymbol{\theta} \boldsymbol{\theta}} \sum_{k=0}^K \mathbf{M}_{\boldsymbol{\theta}, k}^T \boldsymbol{\lambda}_{k+1}, \quad (71)$$

$$\mathbf{x}_{k+1} = \mathbf{m}(\mathbf{x}_k, \boldsymbol{\theta}), \quad (72)$$

Backward model for the adjoint variable

$$\boldsymbol{\lambda}_{K+1} = 0, \quad (73)$$

$$\boldsymbol{\lambda}_k = \mathbf{M}_{x, k}^T \boldsymbol{\lambda}_{k+1} - \mathbf{H}_k^T \mathbf{C}_{dd}^{-1} (\mathbf{h}(\mathbf{x}) - \mathbf{d}). \quad (74)$$

Coupled two-point boundary-value problem in time.

SC-4DVar algorithm

```

1: Input:  $\mathbf{z}^f \in \Re^n$ ;  $\mathbf{d} \in \Re^m$                                 ▷ Prior initial conditions and observations
2:  $\mathbf{x}_0 = \mathbf{x}_0^f$                                          ▷ Initialization of  $\mathbf{x}_0$ 
3:  $\boldsymbol{\theta} = \boldsymbol{\theta}^f$                                          ▷ Initialization of  $\boldsymbol{\theta}$ 
4: repeat                                                 ▷ Iteration loop
5:   for  $k = 0 : K$  do                               ▷ Integrate forward model
6:      $\mathbf{x}_{k+1} = \mathbf{m}(\mathbf{x}_k, \boldsymbol{\theta})$ 
7:   end for
8:    $\lambda_{K+1} = 0$ 
9:   for  $k = K : 0$  do                               ▷ Integrate backward adjoint model
10:     $\lambda_k = \mathbf{M}_{x,k}^T \lambda_{k+1} - \mathbf{H}_k^T \mathbf{C}_{dd}^{-1} (\mathbf{Hx} - \boldsymbol{\eta})$ 
11:   end for
12:    $\mathbf{x}_0 \leftarrow \mathbf{x}_0 - \gamma \mathbf{B} \nabla_{x_0} \mathcal{L}(\mathbf{x}, \boldsymbol{\theta}, \lambda)$  ▷ Update  $\mathbf{x}_0$  using Eq. (67)
13:    $\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} - \gamma \mathbf{B} \nabla_{\boldsymbol{\theta}} \mathcal{L}(\mathbf{x}, \boldsymbol{\theta}, \lambda)$  ▷ Update  $\boldsymbol{\theta}$  using Eq. (68)
14: until convergence

```

Incremental strong constraint 4DVar

Incremental SC-4DVar

Nonlinear model

$$\mathbf{x}_0 = \mathbf{x}_0^f + \mathbf{x}'_0, \quad (75)$$

$$\mathbf{x}_{k+1} = \mathbf{m}(\mathbf{x}_k). \quad (76)$$

State variable

$$\mathbf{z} = \mathbf{x}_0 \quad (77)$$

We compute updates

$$\mathbf{z}^{i+1} = \mathbf{z}^i + \delta\mathbf{z}, \quad (78)$$

Solving for the increments

Inner incremental SC-4DVar costfunction

$$\mathcal{J}(\delta\mathbf{z}) = \frac{1}{2} (\delta\mathbf{z} - \boldsymbol{\xi}^i)^T \mathbf{C}_{zz}^{-1} (\delta\mathbf{z} - \boldsymbol{\xi}^i) + \frac{1}{2} (\mathbf{H}^i \delta\mathbf{x} - \boldsymbol{\eta}^i)^T \mathbf{C}_{dd}^{-1} (\mathbf{H}^i \delta\mathbf{x} - \boldsymbol{\eta}^i), \quad (79)$$

Linear model for the increments

$$\delta\mathbf{x}_0 = \delta\mathbf{z}^i, \quad (80)$$

$$\delta\mathbf{x}_{k+1} = \mathbf{M}_k^i \delta\mathbf{x}_k. \quad (81)$$

Prior increment

$$\boldsymbol{\xi}^i = \mathbf{x}_0^f - \mathbf{x}_0^i, \quad (82)$$

Innovation

$$\boldsymbol{\eta}^i = \mathbf{d} - \mathbf{h}(\mathbf{x}^i). \quad (83)$$

Lagrangian formulation

$$\begin{aligned}
 \mathcal{L}(\delta\mathbf{z}, \delta\mathbf{x}, \delta\boldsymbol{\lambda}) = & \frac{1}{2} (\delta\mathbf{z} - \boldsymbol{\xi}^i)^T \mathbf{C}_{zz}^{-1} (\delta\mathbf{z} - \boldsymbol{\xi}^i) \\
 & + \frac{1}{2} (\mathbf{H}^i \delta\mathbf{x} - \boldsymbol{\eta}^i)^T \mathbf{C}_{dd}^{-1} (\mathbf{H}^i \delta\mathbf{x} - \boldsymbol{\eta}^i) \\
 & + \sum_{k=0}^K \delta\boldsymbol{\lambda}_{k+1}^T (\delta\mathbf{x}_{k+1} - \mathbf{M}_k^i \delta\mathbf{x}_k).
 \end{aligned} \tag{84}$$

Introduces the linear model for the increments using a Lagrangian multiplier.

Gradient of Lagrangian

$$\nabla_{\delta \mathbf{x}_k} \mathcal{L}(\delta \mathbf{z}, \delta \mathbf{x}, \delta \boldsymbol{\lambda}) = \mathbf{H}_k^{i^T} \mathbf{C}_{dd}^{-1} (\mathbf{H}^i \delta \mathbf{x} - \boldsymbol{\eta}^i) + \delta \boldsymbol{\lambda}_k - \mathbf{M}_k^{i^T} \delta \boldsymbol{\lambda}_{k+1}, \quad (85)$$

$$\nabla_{\delta \mathbf{x}_K} \mathcal{L}(\delta \mathbf{z}, \delta \mathbf{x}, \delta \boldsymbol{\lambda}) = \delta \boldsymbol{\lambda}_{K+1}, \quad (86)$$

$$\begin{aligned} \nabla_{\delta \mathbf{z}} \mathcal{L}(\delta \mathbf{z}, \delta \mathbf{x}, \delta \boldsymbol{\lambda}) &= \mathbf{C}_{zz}^{-1} (\delta \mathbf{z} - \boldsymbol{\xi}^i) - \mathbf{M}_0^{i^T} \delta \boldsymbol{\lambda}_1 \\ &= \mathbf{C}_{zz}^{-1} (\delta \mathbf{z} - \boldsymbol{\xi}^i) - \delta \boldsymbol{\lambda}_0, \end{aligned} \quad (87)$$

$$\nabla_{\delta \boldsymbol{\lambda}_k} \mathcal{L}(\delta \mathbf{z}, \delta \mathbf{x}, \delta \boldsymbol{\lambda}) = \delta \mathbf{x}_{k+1} - \mathbf{M}_k^i \delta \mathbf{x}_k. \quad (88)$$

Euler-Lagrange equation(s)

Forward model

$$\delta \mathbf{x}_0 = \xi^i + \mathbf{C}_{zz} \delta \lambda_0, \quad (89)$$

$$\delta \mathbf{x}_{k+1} - \mathbf{M}_k^i \delta \mathbf{x}_k = 0, \quad (90)$$

Adjoint model

$$\delta \lambda_{K+1} = 0, \quad (91)$$

$$\delta \lambda_k - \mathbf{M}_k^{i^T} \delta \lambda_{k+1} = \mathbf{H}_k^{i^T} \mathbf{C}_{dd}^{-1} (\mathbf{H}^i \delta \mathbf{x} - \boldsymbol{\eta}^i). \quad (92)$$

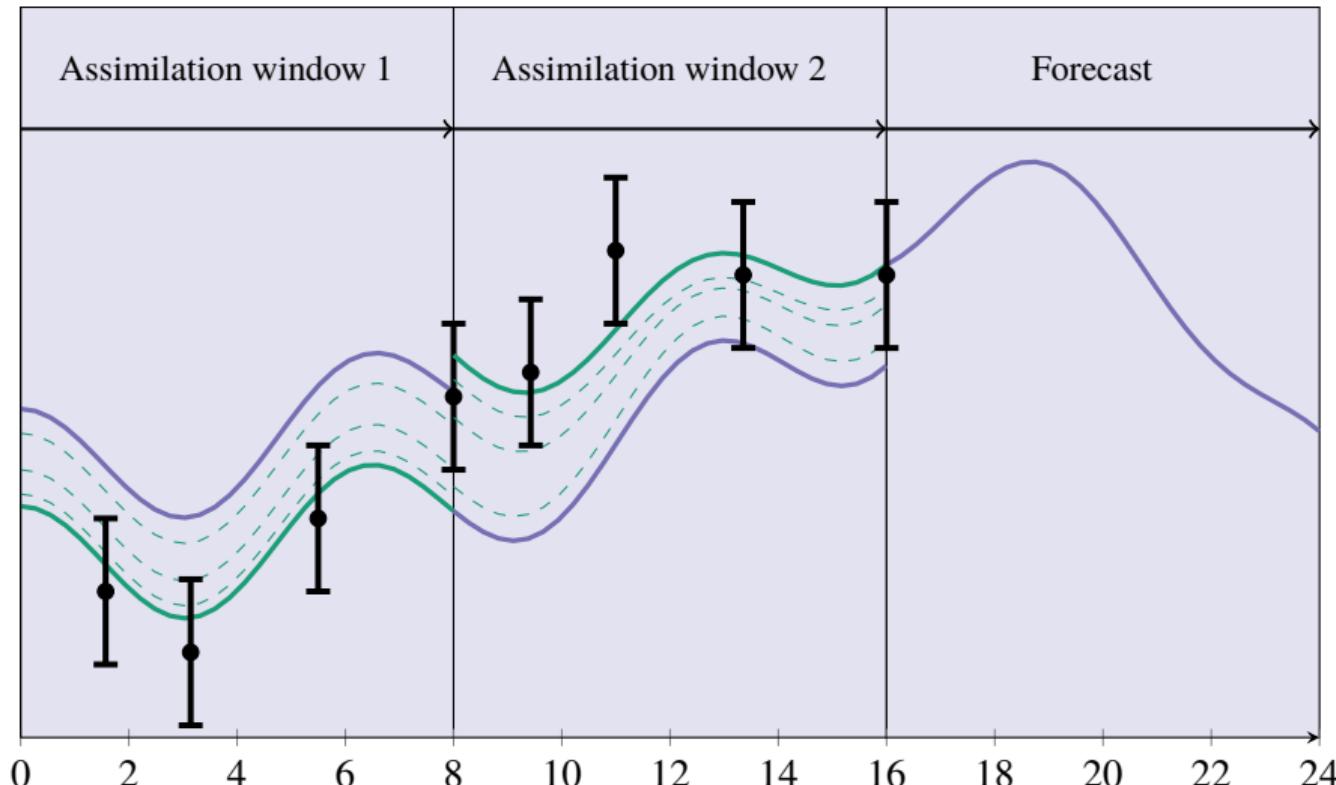
Incremental 4DVar algorithm

```

1: Input:  $\mathbf{x}_0^f \in \mathbb{R}^n$ ;  $\mathbf{d} \in \mathbb{R}^m$ ;  $\mathbf{C}_{dd} \in \mathbb{R}^{m \times m}$ ;  $\mathbf{C}_{zz} \in \mathbb{R}^{n \times n}$                                 ▷ Prior inputs
2:  $\mathbf{x}_0^0 = \mathbf{x}_0^f$                                                                ▷ Initialization of  $\mathbf{x}_0$ 
3:  $\delta\mathbf{z} = 0$                                                                     ▷ Initialization of  $\delta\mathbf{z}$ 
4:  $i = 1$ 
5: repeat                                                                           ▷ Iteration loop
6:    $\mathbf{x}_0^i = \mathbf{x}_0^{i-1} + \delta\mathbf{z}$                                          ▷ Update initial condition
7:    $\xi^i = \mathbf{x}_0^f - \mathbf{x}_0^i$                                                  ▷ Current increment
8:   for  $k = 0 : K$  do                                                       ▷ Integrate forward model
9:      $\mathbf{x}_{k+1}^i = \mathbf{m}(\mathbf{x}_k^i)$                                          ▷ Nonlinear model prediction
10:    end for
11:     $\eta^i = \mathbf{d} - \mathbf{h}(\mathbf{x}^i)$                                          ▷ Current innovation vector
12:    repeat                                                                           ▷ Iteration loop
13:       $\delta\mathbf{x}_0 = \delta\mathbf{z}$ 
14:      for  $k = 0 : K$  do                                                       ▷ Integrate forward model
15:         $\delta\mathbf{x}_{k+1} = \mathbf{M}_k^i \delta\mathbf{x}_k$ 
16:      end for
17:       $\delta\lambda_{K+1} = 0$ 
18:      for  $k = K : 0$  do                                                       ▷ Integrate backward adjoint model
19:         $\delta\lambda_k = \mathbf{M}_k^{i^T} \delta\lambda_{k+1} - \mathbf{H}_k^{i^T} \mathbf{C}_{dd}^{-1} (\mathbf{H}^i \delta\mathbf{x} - \eta^i)$ 
20:      end for
21:       $\delta\mathbf{z} \leftarrow \delta\mathbf{z} - \gamma \mathbf{B} (\delta\mathbf{z} - \xi^i - \mathbf{C}_{zz} \delta\lambda_0)$  ▷ Update  $\delta\mathbf{z}$  using gradient in Eq. (87)
22:    until convergence
23:     $i = i + 1$ 
24:  until convergence

```

Incremental 4DVar algorithm



Comments

- Incremental SC-4DVar is now the standard approach in weather prediction.
- The method solves a data assimilation problem over an assimilation window.
- The solution is the initial condition for the window and it defines the MAP solution.
- Do we know that the incremental formulation converges in all cases?
- We do not evolve error statistics in time.
- Finally, it is a strong-constraint solution that assumes the model is perfect.

Weak constraint 4DVar

WC-4DVar: State-space formulation

Model with additive errors

$$\mathbf{x}_k = \mathbf{m}(\mathbf{x}_{k-1}) + \mathbf{q}_k. \quad (93)$$

Define state vector and model error covariance

$$\mathbf{z} = \mathbf{x} = \begin{pmatrix} \mathbf{x}_0 \\ \vdots \\ \mathbf{x}_K \end{pmatrix}, \quad \text{and} \quad \mathbf{C}_{qq} = \begin{pmatrix} \mathbf{C}_{q_1 q_1} & \cdots & \mathbf{C}_{q_1 q_K} \\ \vdots & \ddots & \vdots \\ \mathbf{C}_{q_K q_1} & \cdots & \mathbf{C}_{q_K q_K} \end{pmatrix}, \quad (94)$$

WC-4DVar cost function

The weak-constraint cost function (generalized inverse formulation)

$$\begin{aligned}
 \mathcal{J}(\mathbf{x}) = & \frac{1}{2} (\mathbf{x}_0 - \mathbf{x}_0^f)^T \mathbf{C}_{x_0 x_0}^{-1} (\mathbf{x}_0 - \mathbf{x}_0^f) \\
 & + \frac{1}{2} (\mathbf{h}(\mathbf{x}) - \mathbf{d})^T \mathbf{C}_{dd}^{-1} (\mathbf{h}(\mathbf{x}) - \mathbf{d}) \\
 & + \frac{1}{2} \sum_{r=1}^K \sum_{s=1}^K (\mathbf{x}_r - \mathbf{m}(\mathbf{x}_{r-1}))^T \mathbf{C}_{qq}^{-1}(r,s) (\mathbf{x}_s - \mathbf{m}(\mathbf{x}_{s-1})). \tag{95}
 \end{aligned}$$

WC-4DVar: incremental form

Define update

$$\mathbf{x}^{i+1} = \mathbf{x}^i + \delta\mathbf{x}. \quad (96)$$

Linearized model

$$\begin{aligned} \mathbf{m}(\mathbf{x}_k^{i+1}) &= \mathbf{m}(\mathbf{x}_k^i + \delta\mathbf{x}_k) \\ &\approx \mathbf{m}(\mathbf{x}_k^i) + \mathbf{M}_k \delta\mathbf{x}_k, \end{aligned} \quad (97)$$

Linearized measurement operator

$$\begin{aligned} \mathbf{h}(\mathbf{x}^{i+1}) &= \mathbf{h}(\mathbf{x}^i + \delta\mathbf{x}) \\ &\approx \mathbf{h}(\mathbf{x}^i) + \mathbf{H} \delta\mathbf{x}, \end{aligned} \quad (98)$$

WC-4DVar: incremental form

Write the model residual in Eq. (95) as

$$\begin{aligned}
 \mathbf{x}_k^{i+1} - \mathbf{m}(\mathbf{x}_{k-1}^{i+1}) &\approx \mathbf{x}_k^{i+1} - \mathbf{m}(\mathbf{x}_{k-1}^i) - \mathbf{M}_{k-1}\delta\mathbf{x}_{k-1} \\
 &= \mathbf{x}_k^{i+1} - \mathbf{x}_k^i + \mathbf{x}_k^i - \mathbf{m}(\mathbf{x}_{k-1}^i) - \mathbf{M}_{k-1}\delta\mathbf{x}_{k-1} \\
 &= \delta\mathbf{x}_k - \mathbf{M}_{k-1}\delta\mathbf{x}_{k-1} + \boldsymbol{\xi}_k^i,
 \end{aligned} \tag{99}$$

where

$$\boldsymbol{\xi}_k^i = \mathbf{x}_k^i - \mathbf{m}(\mathbf{x}_{k-1}^i). \tag{100}$$

Innovations becomes

$$\boldsymbol{\eta}^i = \mathbf{d} - \mathbf{h}(\mathbf{x}^i). \tag{101}$$

WC incremental cost function

The weak-constraint incremental cost function

$$\begin{aligned}
 \mathcal{J}(\delta\mathbf{x}) = & \frac{1}{2} (\delta\mathbf{x}_0 - \boldsymbol{\xi}_0^i)^T \mathbf{C}_{x_0 x_0}^{-1} (\delta\mathbf{x}_0 - \boldsymbol{\xi}_0^i) \\
 & + \frac{1}{2} (\mathbf{H}^i \delta\mathbf{x} - \boldsymbol{\eta}^i)^T \mathbf{C}_{dd}^{-1} (\mathbf{H}^i \delta\mathbf{x} - \boldsymbol{\eta}^i) \\
 & + \frac{1}{2} \sum_{r=1}^K \sum_{s=1}^K (\delta\mathbf{x}_r - \mathbf{M}_{r-1} \delta\mathbf{x}_{r-1} + \boldsymbol{\xi}_r^i)^T \mathbf{C}_{qq}^{-1}(r, s) (\delta\mathbf{x}_s - \mathbf{M}_{s-1} \delta\mathbf{x}_{s-1} + \boldsymbol{\xi}_s^i).
 \end{aligned} \tag{102}$$

Linear problem with Gaussian priors

Minimizing the cost function for the increment

Gradient of the cost function to the model state $\delta\mathbf{x}_k$ gives

$$\begin{aligned} \nabla_{\delta\mathbf{x}_k} \mathcal{J}(\delta\mathbf{x}) &= \mathbf{H}_k^{i^T} \mathbf{C}_{dd}^{-1} (\mathbf{H}^i \delta\mathbf{x} - \boldsymbol{\eta}^i) \\ &\quad + \sum_{s=1}^K \mathbf{C}_{qq}^{-1}(k, s) (\delta\mathbf{x}_s - \mathbf{M}_{s-1}^i \delta\mathbf{x}_{s-1} + \boldsymbol{\xi}_s^i) \\ &\quad - \mathbf{M}_k^{i^T} \sum_{s=1}^K \mathbf{C}_{qq}^{-1}(k+1, s) (\delta\mathbf{x}_s - \mathbf{M}_{s-1}^i \delta\mathbf{x}_{s-1} + \boldsymbol{\xi}_s^i). \end{aligned} \tag{103}$$

Define an adjoint vector for each time step as

$$\delta\lambda_k = \sum_{s=1}^K \mathbf{C}_{qq}^{-1}(k, s) (\delta\mathbf{x}_s - \mathbf{M}_{s-1}^i \delta\mathbf{x}_{s-1} + \boldsymbol{\xi}_s^i), \tag{104}$$

and write Eq. (103) as

$$\nabla_{\delta\mathbf{x}_k} \mathcal{J}(\delta\mathbf{x}) = \mathbf{H}_k^{i^T} \mathbf{C}_{dd}^{-1} (\mathbf{H}^i \delta\mathbf{x} - \boldsymbol{\eta}^i) + \delta\lambda_k - \mathbf{M}_k^{i^T} \delta\lambda_{k+1}. \tag{105}$$

Euler-Lagrange equation(s)

Forward model

$$\delta \mathbf{x}_0 = \xi_0^i + \mathbf{C}_{x_0 x_0} \delta \lambda_0, \quad (106)$$

$$\delta \mathbf{x}_k - \mathbf{M}_{k-1}^i \delta \mathbf{x}_{k-1} = -\xi_k^i + \sum_{s=1}^K \mathbf{C}_{qq}(k, s) \delta \lambda_s, \quad (107)$$

Backward model

$$\delta \lambda_{K+1} = \mathbf{0}, \quad (108)$$

$$\delta \lambda_k - \mathbf{M}_k^{i^T} \delta \lambda_{k+1} = -\mathbf{H}_k^{i^T} \mathbf{C}_{dd}^{-1} (\mathbf{H}^i \delta \mathbf{x} - \boldsymbol{\eta}^i). \quad (109)$$

Representer method

We have a linear problem for the increments. Assume solution of the form

$$\delta\mathbf{x} = \delta\mathbf{x}^f + \sum_{p=1}^m b_p \mathbf{r}_p = \delta\mathbf{x}^f + \mathbf{R}\mathbf{b}, \quad (110)$$

$$\delta\boldsymbol{\lambda} = \delta\boldsymbol{\lambda}^f + \sum_{p=1}^m b_p \mathbf{s}_p = \delta\boldsymbol{\lambda}^f + \mathbf{S}\mathbf{b}. \quad (111)$$

Solution is a linear combination of influence functions (representers) \mathbf{r}_p , one for each measurement.

Euler-Lagrange equations: Prior

We get for the prior

$$\delta \mathbf{x}_0^f = \boldsymbol{\xi}_0^i, \quad (112)$$

$$\delta \mathbf{x}_k^f - \mathbf{M}_{k-1}^i \delta \mathbf{x}_{k-1}^f = -\boldsymbol{\xi}_k^i, \quad (113)$$

$$\delta \boldsymbol{\lambda}_{K+1}^f = 0, \quad (114)$$

$$\delta \boldsymbol{\lambda}_k^f - \mathbf{M}_k^{i^T} \delta \boldsymbol{\lambda}_{k+1}^f = 0. \quad (115)$$

which is just the prior model solution obtained from one model integration.

Derivation for the representers

$$\mathbf{R}_0 \mathbf{b} = \mathbf{C}_{x_0 x_0} \mathbf{M}_0^{i^T} \mathbf{S}_1 \mathbf{b}, \quad (116)$$

$$(\mathbf{R}_k \mathbf{b} - \mathbf{M}_{k-1}^i \mathbf{R}_{k-1} \mathbf{b}) = \sum_{s=1}^K \mathbf{C}_{qq}(k, s) \mathbf{S}_s \mathbf{b}, \quad (117)$$

$$\mathbf{S}_{K+1} \mathbf{b} = \mathbf{0}, \quad (118)$$

$$\mathbf{S}_k \mathbf{b} - \mathbf{M}_k^{i^T} \mathbf{S}_{k+1} \mathbf{b} = \mathbf{H}_k^{i^T} \mathbf{C}_{dd}^{-1} (\boldsymbol{\eta}^i - \mathbf{H}^i (\delta \mathbf{x}^f + \mathbf{R} \mathbf{b})). \quad (119)$$

Define \mathbf{b} as

$$\mathbf{b} = \mathbf{C}_{dd}^{-1} (\boldsymbol{\eta}^i - \mathbf{H}^i (\delta \mathbf{x}^f + \mathbf{R} \mathbf{b})), \quad (120)$$

such that Eq. (119) simplifies to

$$\mathbf{S}_k \mathbf{b} - \mathbf{M}_k^{i^T} \mathbf{S}_{k+1} \mathbf{b} = \mathbf{H}_k^{i^T} \mathbf{b}. \quad (121)$$

Euler-Lagrange equations: Representers

$$\mathbf{R}_0 = \mathbf{C}_{x_0 x_0} \mathbf{M}_0^{i^T} \mathbf{S}_1, \quad (122)$$

$$\mathbf{R}_k - \mathbf{M}_{k-1}^i \mathbf{R}_{k-1} = \sum_{s=1}^K \mathbf{C}_{qq}(k, s) \mathbf{S}_s, \quad (123)$$

$$\mathbf{S}_{K+1} = \mathbf{0}, \quad (124)$$

$$\mathbf{S}_k - \mathbf{M}_k^{i^T} \mathbf{S}_{k+1} = \mathbf{H}_k^{i^T}. \quad (125)$$

Requires $2m$ model integrations.

Euler-Lagrange equations: \mathbf{b}

Linear system for the representer coefficients \mathbf{b}

$$(\mathbf{H}^i \mathbf{R} + \mathbf{C}_{dd}) \mathbf{b} = \boldsymbol{\eta}^i - \mathbf{H}^i \delta \mathbf{x}^f. \quad (126)$$

Results in minimizing solution from $2m + 1$ model integrations.

More efficient solution method

Given \mathbf{b} we obtain the final solution from (109)

Adjoint equation forced by \mathbf{b}

$$\delta\lambda_k - \mathbf{M}_k^{i^T} \delta\lambda_{k+1} = \mathbf{H}_k^{i^T} \mathbf{b}. \quad (127)$$

followed by a forward model integration.

Thus, no need to store the representers \mathbf{R} , (only $\mathbf{H}^i \mathbf{R}$ is needed to solve for \mathbf{b}).

An even more efficient solution method

Use an iterative equation solver:

$$\mathbf{C}\mathbf{b} = \boldsymbol{\nu}. \quad (128)$$

Solving Eq. (128) is equivalent to minimizing the functional

$$\phi(\mathbf{b}) = \frac{1}{2} \mathbf{b}^T \mathbf{C} \mathbf{b} - \mathbf{b}^T \boldsymbol{\nu}, \quad (129)$$

which has a gradient

$$\nabla_{\mathbf{b}} \phi(\mathbf{b}) = \boldsymbol{\rho} = \mathbf{C}\mathbf{b} - \boldsymbol{\nu}. \quad (130)$$

Iterative solution

$$\mathbf{b}^{i+1} = \mathbf{b}^i - \gamma \boldsymbol{\rho}, \quad (131)$$

We don't need to know \mathbf{C} only the product $\mathbf{C}\mathbf{b}$.

Evaluate the product $\mathbf{R}\mathbf{v}$

Define $\mathbf{c} = \mathbf{R}\mathbf{v}$ and $\boldsymbol{\psi} = \mathbf{S}\mathbf{v}$, and write (116–119) as

$$\mathbf{c}_0 = \mathbf{C}_{x_0 x_0} \mathbf{M}_0^{i^T} \boldsymbol{\psi}_1, \quad (132)$$

$$\mathbf{c}_k = \mathbf{M}_{k-1}^i \mathbf{c}_{k-1} + \sum_{s=1}^K \mathbf{C}_{qq}(k, s) \boldsymbol{\psi}_s, \quad (133)$$

$$\boldsymbol{\psi}_{K+1} = \mathbf{0}, \quad (134)$$

$$\boldsymbol{\psi}_k = \mathbf{M}_k^{i^T} \boldsymbol{\psi}_{k+1} + \mathbf{H}_k^{i^T} \mathbf{v}. \quad (135)$$

By measuring this solution, we find for any nonzero vector \mathbf{v}

$$\mathbf{H}\mathbf{c} = \mathbf{H}\mathbf{R}\mathbf{v} = \mathcal{R}\mathbf{v}, \quad (136)$$

Comments

- Efficient solution for WC-4DVar problem by the Representer method.
- The method solves the data-assimilation problem over an assimilation window.
- The solution is the model state over the window and defines the MAP solution.
- **Do we know that the incremental formulation converges in all cases?**
- **We do not evolve error statistics in time.**
- **A prediction should be initialized from the end of the assimilation window.**
- An ensemble of 4DVars would allow for evolving error statistics in time.

Representer method with an Ekman flow model

Representer solution for an Ekman flow model

Model

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{k} \times \mathbf{u} = \frac{\partial}{\partial z} \left(A \frac{\partial \mathbf{u}}{\partial z} \right), \quad (137)$$

Initial condition

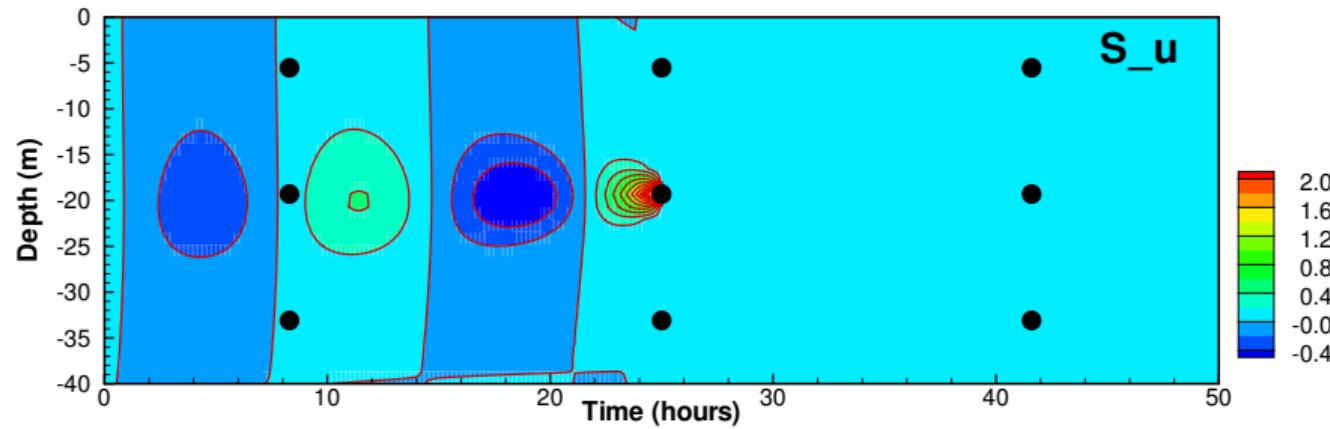
$$\mathbf{u}(z, 0) = \mathbf{u}_0, \quad (138)$$

boundary conditions

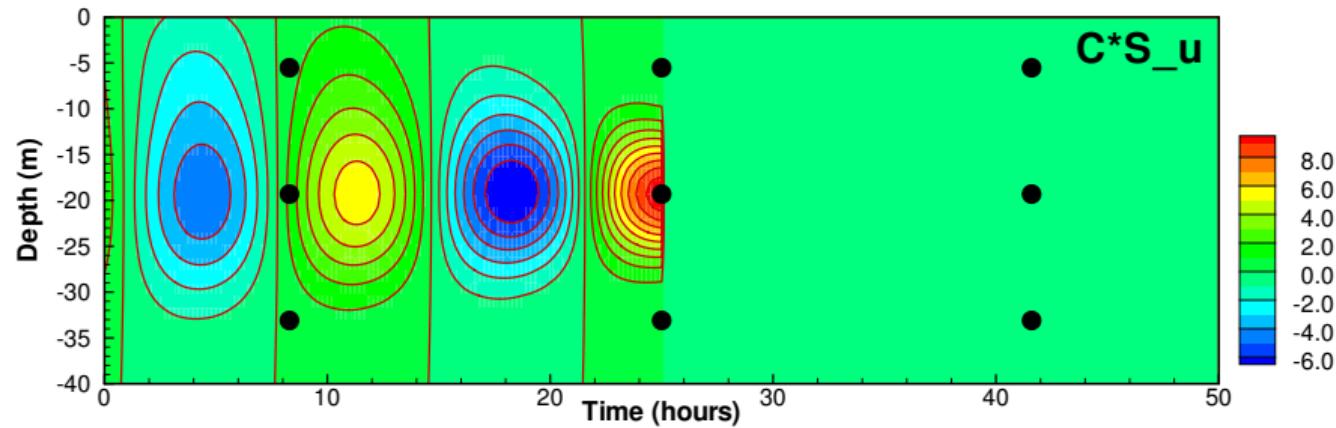
$$A \frac{\partial \mathbf{u}}{\partial z} \Big|_{z=0} = \left(c_d \sqrt{u_a^2 + v_a^2} \right) \mathbf{u}_a, \quad (139)$$

$$A \frac{\partial \mathbf{u}}{\partial z} \Big|_{z=-H} = \mathbf{0}, \quad (140)$$

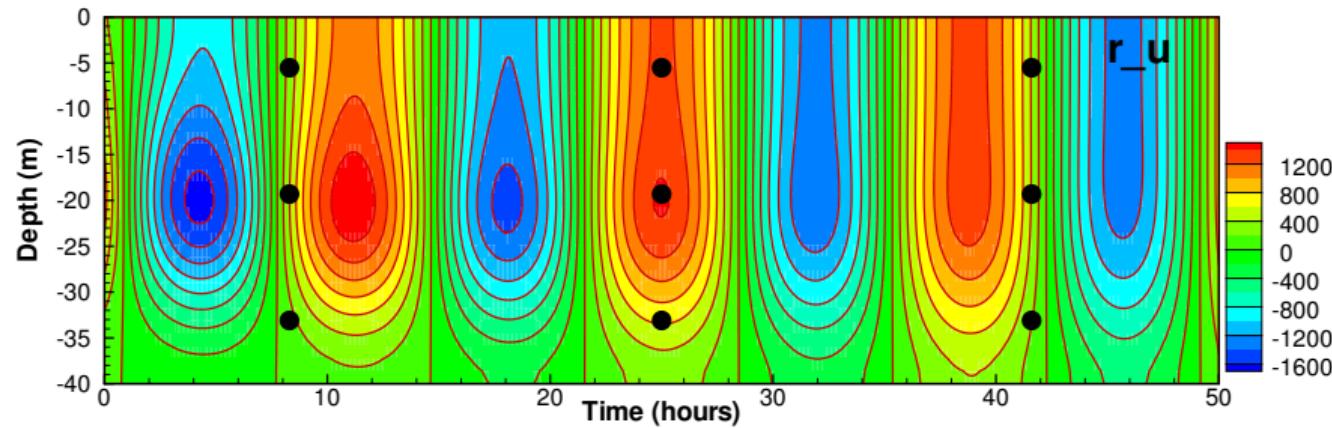
An adjoint representer s



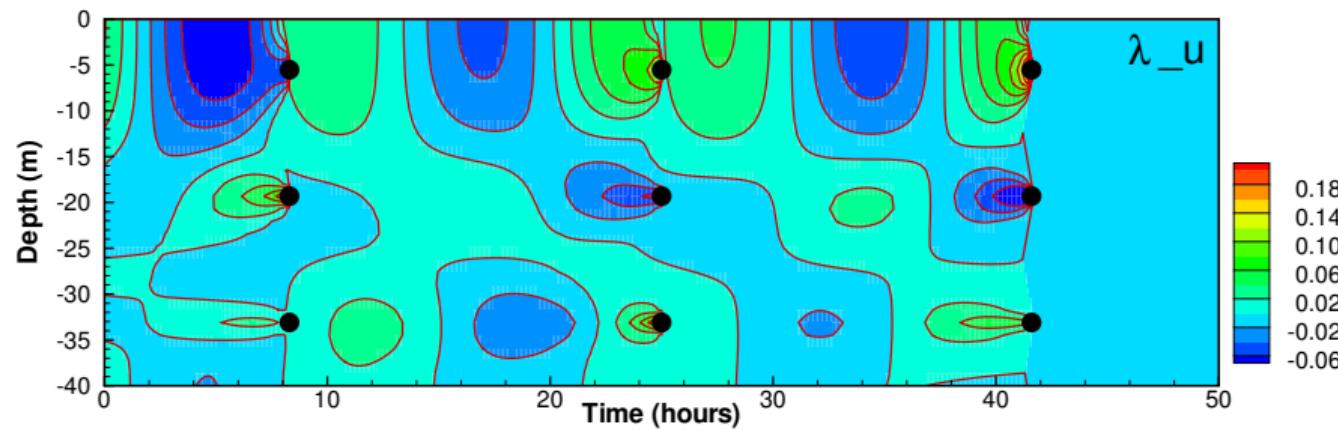
A convolution of an adjoint representer $\mathbf{C}_{qq} \circ \mathbf{s}$



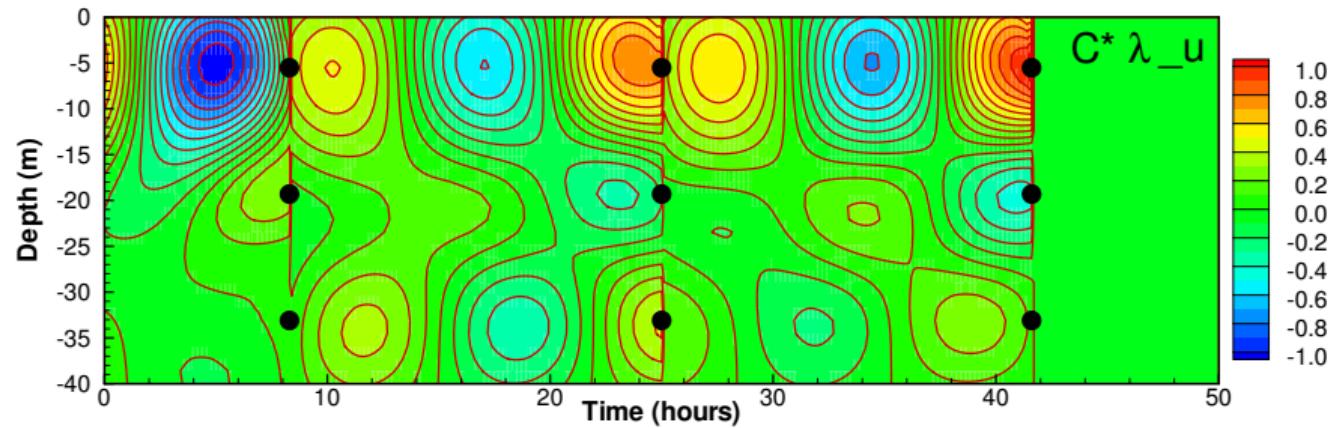
A representer \mathbf{r}



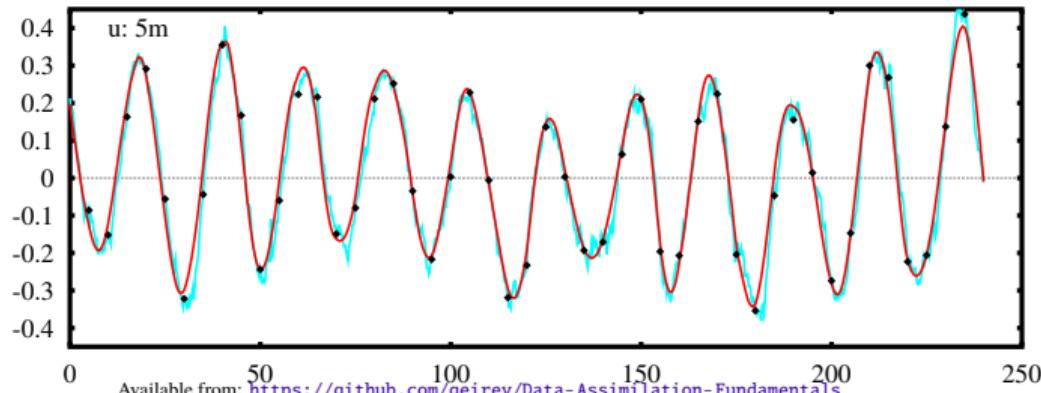
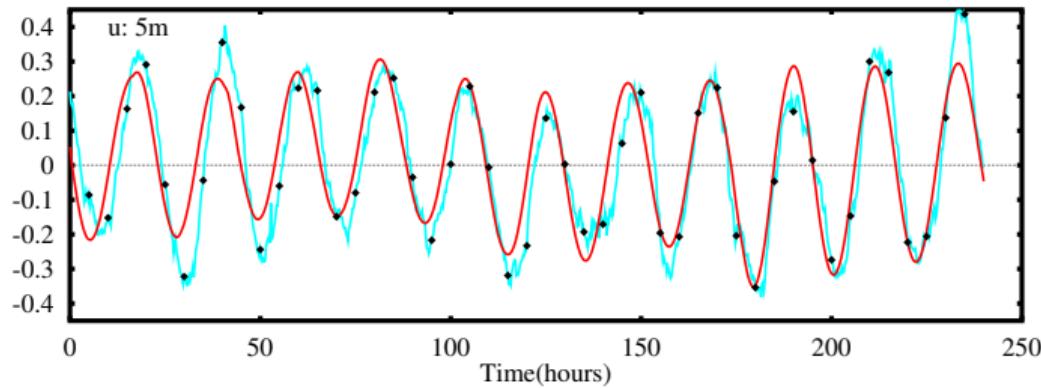
The adjoint λ



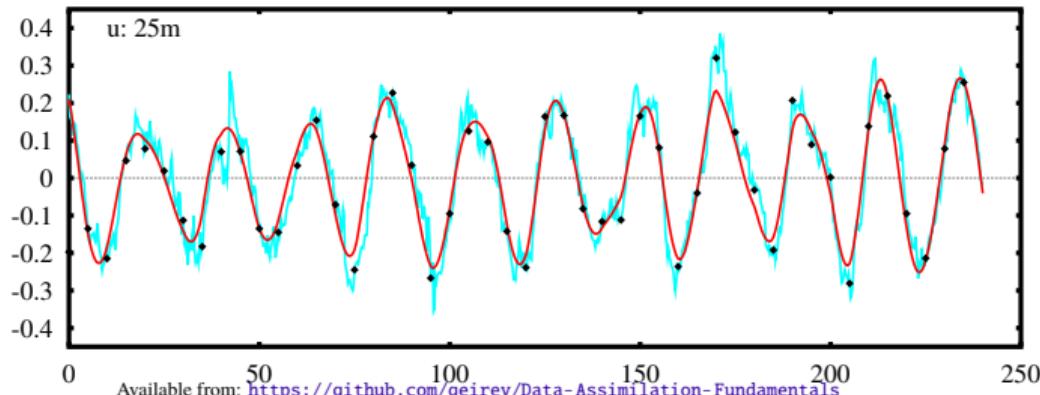
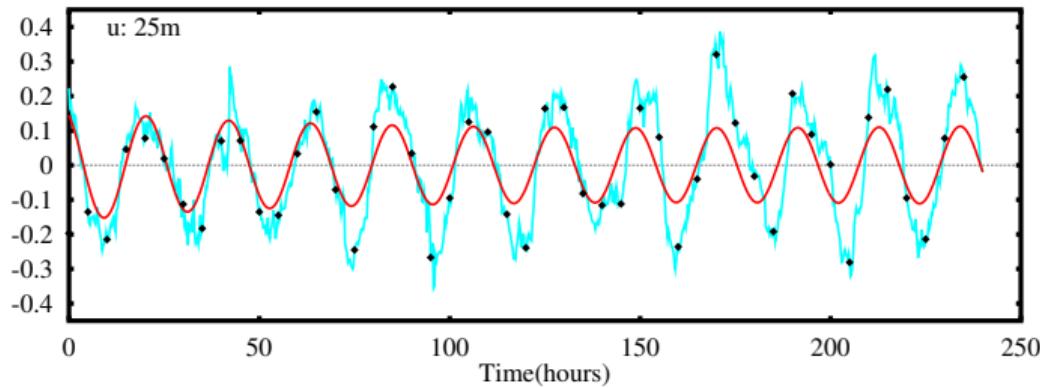
The convolution of the adjoint $\mathbf{C}_{qq} \circ \lambda$



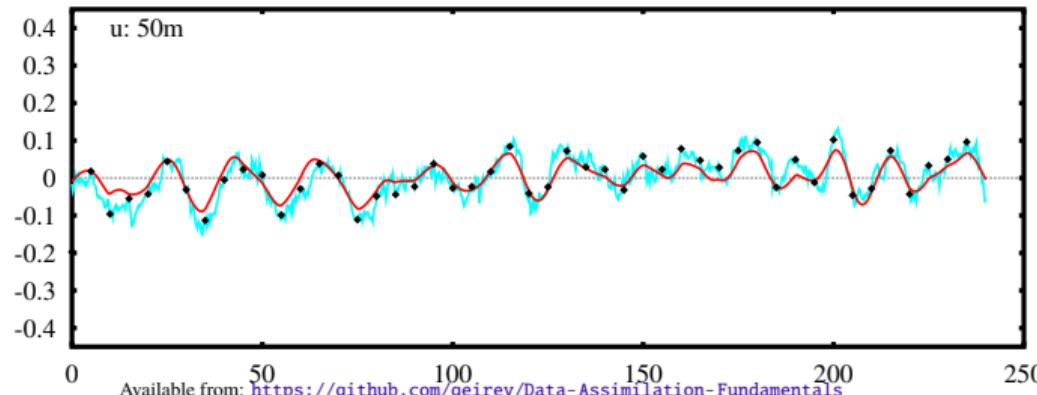
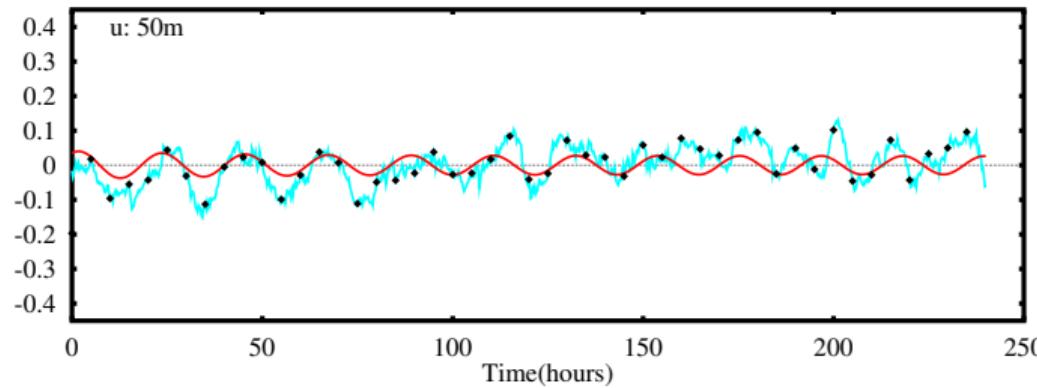
Strong and weak constraint solution



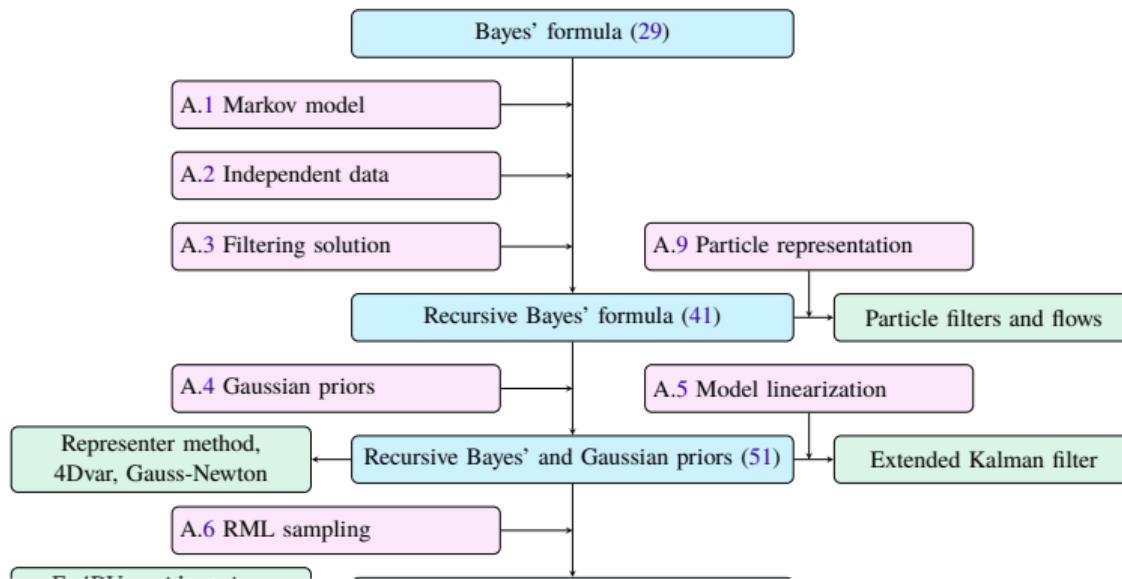
Strong and weak constraint solution



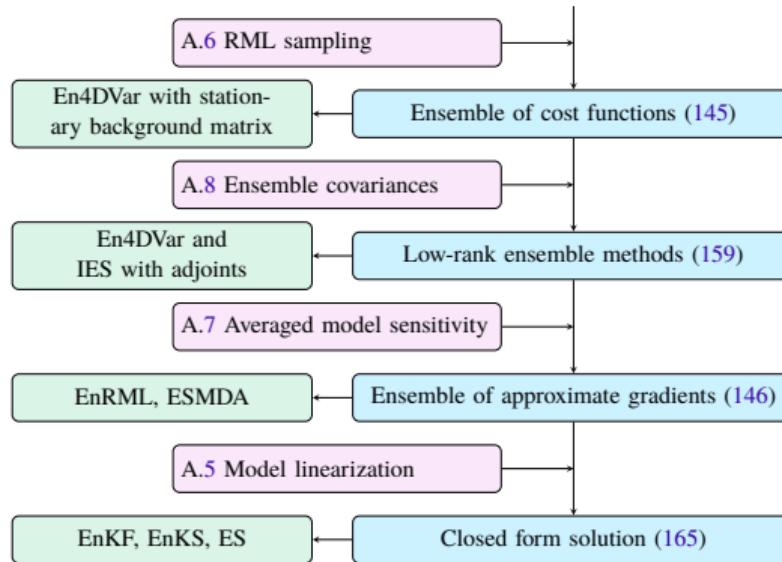
Strong and weak constraint solution



Overview of approximations and methods



Overview of approximations and methods



Kalman filters and 3D-Var

Linearization leads to an approximate explicit solution

Approximation 5 (Linearization)

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

$$\mathbf{g}(\mathbf{z}) \approx \mathbf{g}(\mathbf{z}^f) + \mathbf{G}(\mathbf{z} - \mathbf{z}^f), \quad (141)$$

and approximate the gradient evaluated at the prior estimate

$$\nabla_{\mathbf{z}} \mathbf{g}(\mathbf{z}) \approx \mathbf{G}^T, \quad (142)$$

where we have defined

$$\mathbf{G}^T = \nabla_{\mathbf{z}} \mathbf{g}(\mathbf{z}) \Big|_{\mathbf{z}=\mathbf{z}^f}. \quad (143)$$

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

Extended Kalman Filter solution

The linearization in Approx. 5 leads to closed form solution of (52)

$$\mathbf{z}^a = \mathbf{z}^f + \mathbf{C}_{zz} \mathbf{G}^T \left(\mathbf{G} \mathbf{C}_{zz} \mathbf{G}^T + \mathbf{C}_{dd} \right)^{-1} \left(\mathbf{d} - \mathbf{g}(\mathbf{z}^f) \right). \quad (144)$$

Randomized-maximum-likelihood sampling

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. (145).

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

RML minimizes an ensemble of cost functions

Ensemble of cost functions

$$\mathcal{J}(\mathbf{z}_j) = \frac{1}{2} (\mathbf{z}_j - \mathbf{z}_j^f)^T \mathbf{C}_{zz}^{-1} (\mathbf{z}_j - \mathbf{z}_j^f) + \frac{1}{2} (\mathbf{g}(\mathbf{z}_j) - \mathbf{d}_j)^T \mathbf{C}_{dd}^{-1} (\mathbf{g}(\mathbf{z}_j) - \mathbf{d}_j), \quad (145)$$

Ensemble of gradients set to zero

$$\mathbf{C}_{zz}^{-1} (\mathbf{z}_j - \mathbf{z}_j^f) + \nabla_{\mathbf{z}} \mathbf{g}(\mathbf{z}_j) \mathbf{C}_{dd}^{-1} (\mathbf{g}(\mathbf{z}_j) - \mathbf{d}_j) = 0. \quad (146)$$

Thus, we must solve N independent minimizations.

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

Ensemble of incremental 4DVars

$$\mathcal{J}(\delta \mathbf{z}_j) = \frac{1}{2} (\delta \mathbf{z}_j - \boldsymbol{\xi}_j^i)^T \mathbf{C}_{zz}^{-1} (\delta \mathbf{z}_j - \boldsymbol{\xi}_j^i) + \frac{1}{2} (\mathbf{G}_j^i \delta \mathbf{z}_j - \boldsymbol{\eta}_j^i)^T \mathbf{C}_{dd}^{-1} (\mathbf{G}_j^i \delta \mathbf{z}_j - \boldsymbol{\eta}_j^i). \quad (147)$$

Ensemble of GN iterations

$$\mathbf{z}_j^{i+1} = \mathbf{z}_j^i - \gamma \left(\mathbf{C}_{zz}^{-1} + \mathbf{G}_j^{iT} \mathbf{C}_{dd}^{-1} \mathbf{G}_j^i \right)^{-1} \left(\mathbf{C}_{zz}^{-1} (\mathbf{z}_j^i - \mathbf{z}_j^f) + \mathbf{G}_j^{iT} \mathbf{C}_{dd}^{-1} (\mathbf{g}(\mathbf{z}_j^i) - \mathbf{d}_j) \right), \quad (148)$$

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

$$\mathbf{z}_j^a = \mathbf{z}_j^f + \mathbf{C}_{zz} \mathbf{G}_j^T \left(\mathbf{G}_j \mathbf{C}_{zz} \mathbf{G}_j^T + \mathbf{C}_{dd} \right)^{-1} \left(\mathbf{d}_j - \mathbf{g}(\mathbf{z}_j^f) \right). \quad (149)$$

Comments

- SC-4DVar and WC-4DVar solves for the MAP estimate over the assimilation window.
- RML sampling using (En)SC-4DVar and (En)WC-4DVar would aim to sample the posterior.
- It is possible to propagate error statistics using ensemble integrations.
- We could have a consistent method using ensemble “background” covariances.
- Still uses the tangent linear and adjoint models.
- What is the benefit of computing update over a finite data-assimilation window?

Use an averaged model sensitivity to avoid adjoints

Approximation 7 (Best-fit averaged model sensitivity)

Interpret \mathbf{G}_j in Eq. (149) and \mathbf{G}_j^i in Eq. (148) as the sensitivity matrix in linear regression and represent them using the definition

$$\mathbf{G}_j \approx \mathbf{G} \triangleq \mathbf{C}_{yz} \mathbf{C}_{zz}^{-1}. \quad (150)$$

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

Low-rank ensemble methods

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.

$$\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_N). \quad (151)$$

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

$$\mathbf{\Pi} = \left(\mathbf{I} - \frac{1}{N} \mathbf{1} \mathbf{1}^T \right) / \sqrt{N-1}, \quad (152)$$

Ensemble anomalies

$$\mathbf{A} = \mathbf{Z} \mathbf{\Pi}, \quad (153)$$

Ensemble matrices

Ensemble covariance

$$\bar{\mathbf{C}}_{zz} = \mathbf{A}\mathbf{A}^T. \quad (154)$$

Measurement error covariance matrix

$$\bar{\mathbf{C}}_{dd} = \mathbf{E}\mathbf{E}^T. \quad (155)$$

Ensemble of model-predicted measurements

$$\mathbf{Y} = \mathbf{g}(\mathbf{Z}), \quad (156)$$

with anomalies

$$\mathbf{Y} = \mathbf{Y}\Pi, \quad (157)$$

Solution is confined to ensemble subspace

Search for the solution in the ensemble subspace

$$\mathbf{Z}^a = \mathbf{Z}^f + \mathbf{A}\mathbf{W}. \quad (158)$$

Cost function in ensemble subspace

$$\mathcal{J}(\mathbf{w}_j) = \frac{1}{2} \mathbf{w}_j^T \mathbf{w}_j + \frac{1}{2} (\mathbf{g}(\mathbf{z}_j^f + \mathbf{A}\mathbf{w}_j) - \mathbf{d}_j)^T \bar{\mathbf{C}}_{dd}^{-1} (\mathbf{g}(\mathbf{z}_j^f + \mathbf{A}\mathbf{w}_j) - \mathbf{d}_j), \quad (159)$$

Gradient and Hessian

Gradient

$$\nabla_{\mathbf{w}} \mathcal{J}(\mathbf{w}_j) = \mathbf{w}_j + (\mathbf{G}_j \mathbf{A})^T \bar{\mathbf{C}}_{dd}^{-1} (\mathbf{g}(\mathbf{z}_j^f + \mathbf{A}\mathbf{w}_j) - \mathbf{d}_j), \quad (160)$$

Approximate Hessian

$$\nabla_{\mathbf{w}} \nabla_{\mathbf{w}} \mathcal{J}(\mathbf{w}_j) \approx \mathbf{I} + (\mathbf{G}_j \mathbf{A})^T \bar{\mathbf{C}}_{dd}^{-1} (\mathbf{G}_j \mathbf{A}). \quad (161)$$

Tangent-linear model

$$\mathbf{G}_j = \left(\nabla_{\mathbf{w}} \mathbf{g} \Big|_{\mathbf{z}_j^f + \mathbf{A}\mathbf{w}_j} \right)^T \in \Re^{m \times n}, \quad (162)$$

Gauss-Newton iterations

$$\mathbf{w}_j^{i+1} = \mathbf{w}_j^i - \gamma \left\{ \mathbf{w}_j^i - (\mathbf{G}_j^i \mathbf{A})^T \left((\mathbf{G}_j^i \mathbf{A})(\mathbf{G}_j^i \mathbf{A})^T + \bar{\mathbf{C}}_{dd} \right)^{-1} \left((\mathbf{G}_j^i \mathbf{A})\mathbf{w}_j^i + \mathbf{d}_j - \mathbf{g}(\mathbf{z}_j^f + \mathbf{A}\mathbf{w}_j^i) \right) \right\}. \quad (163)$$

with linear regression approximation for model sensitivities

$$\mathbf{G}_j^i \approx \bar{\mathbf{G}}^i \triangleq \bar{\mathbf{C}}_{yz}^i \bar{\mathbf{C}}_{zz}^{i\dagger} = \mathbf{Y}^i \mathbf{A}^{i\dagger}, \quad (164)$$

Subspace EnRML

```

1: Input:  $\mathbf{Z} \in \mathbb{R}^{n \times N}$                                 ▷ Prior state-vector ensemble
2: Input:  $\mathbf{D} \in \mathbb{R}^{m \times N}$                                 ▷ Perturbed measurements
3:  $\mathbf{W}^{(0)} = 0$                                          ▷  $\mathbf{W} \in \mathbb{R}^{N \times N}$ 
4:  $\mathbf{\Pi} = \left(\mathbf{I} - \frac{1}{N}\mathbf{1}\mathbf{1}^T\right) / \sqrt{N-1}$     ▷  $\mathbf{\Pi} \in \mathbb{R}^{N \times N}$ 
5:  $\mathbf{E} = \mathbf{D}\mathbf{\Pi}$                                          ▷  $\mathbf{E} \in \mathbb{R}^{m \times N}$ 
6:  $i = 0$ 
7: repeat
8:    $\mathbf{Y}^i = \mathbf{g}(\mathbf{Z}^i)\mathbf{\Pi}$                                 ▷  $\mathbf{Y} \in \mathbb{R}^{m \times N}$ 
9:   if  $n < N - 1$  then
10:     $\mathbf{Y}^i = \mathbf{Y}^i \mathbf{A}^{i\dagger} \mathbf{A}^i$ 
11:   end if
12:    $\mathbf{\Omega}^i = \mathbf{I} + \mathbf{W}^i \mathbf{\Pi}$                                 ▷  $\mathbf{\Omega} \in \mathbb{R}^{N \times N}$ 
13:    $\mathbf{S}^i = \mathbf{Y}^i \mathbf{\Omega}^{i-1}$                                 ▷  $\mathbf{S} \in \mathbb{R}^{m \times N}$ 
14:    $\mathbf{W}^{i+1} = \mathbf{W}^i - \gamma \left( \mathbf{W}^i - \mathbf{S}^{i^T} \left( \mathbf{S}^i \mathbf{S}^{i^T} + \mathbf{E} \mathbf{E}^T \right)^{-1} \left( \mathbf{S}^i \mathbf{W}^i + \mathbf{D} - \mathbf{g}(\mathbf{Z}^i) \right) \right)$ 
15:    $\mathbf{Z}^{i+1} = \mathbf{Z} \left( \mathbf{I} + \mathbf{W}^{i+1} \right) / \sqrt{N-1}$ 
16:    $i = i + 1$ 
17: until convergence

```

EnKF update

Identical to one iteration of Subspace EnRML with step length $\gamma = 1.0$.

$$\mathbf{Z}^a = \mathbf{Z}^f + \mathbf{A} \mathbf{Y}^T (\mathbf{Y} \mathbf{Y}^T + \mathbf{E} \mathbf{E}^T)^{-1} (\mathbf{D} - \mathbf{g}(\mathbf{Z}^f)) \quad (165)$$

Alternative interpretation using

$$\mathbf{W} = \mathbf{Y}^T (\mathbf{Y} \mathbf{Y}^T + \mathbf{E} \mathbf{E}^T)^{-1} (\mathbf{D} - \mathbf{g}(\mathbf{Z}^f)), \quad (166)$$

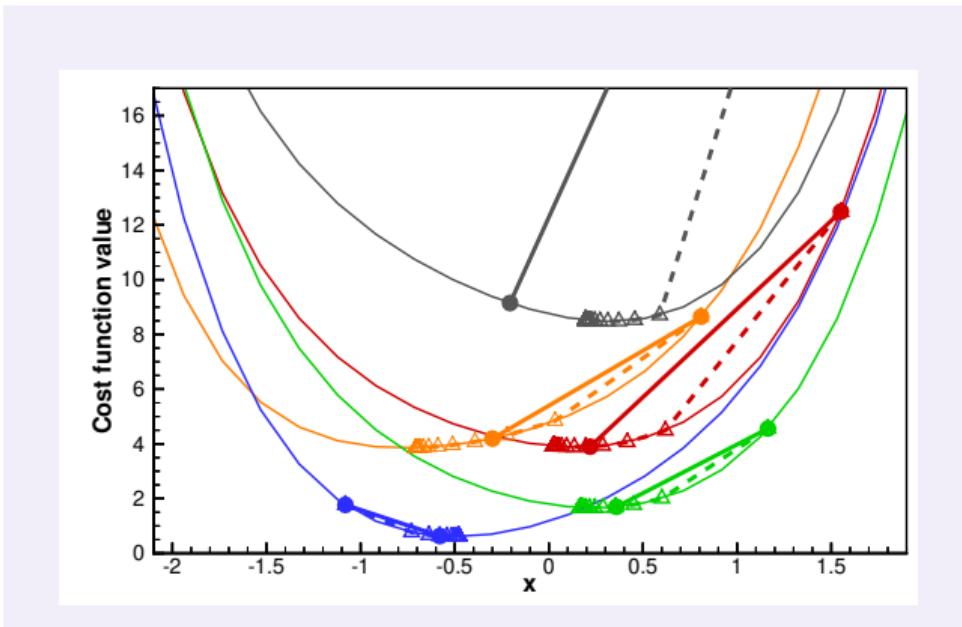
to get

$$\mathbf{Z}^a = \mathbf{Z}^f \left(\mathbf{I} + \mathbf{W} / \sqrt{N - 1} \right) \quad (167)$$

But note that

$$\mathbf{Y} = \begin{cases} \mathbf{Y} & \text{for } n \geq N - 1 \\ \mathbf{Y} \mathbf{A}^\dagger \mathbf{A} & \text{for } n < N - 1. \end{cases} \quad (168)$$

ES and IES illustration: Non-linear model



- IES gets closer to minimum than ES
- Approximate sampling of posterior pdf.

ESMDA uses tapering of likelihood

Approximate sampling of $f(\mathbf{x}|\mathbf{d})$ by gradually introducing the measurements (Neal, 1996)

$$\begin{aligned}
 f(\mathbf{x}|\mathbf{d}) &= f(\mathbf{d}|\mathbf{y})f(\mathbf{x}) \\
 &= f(\mathbf{d}|\mathbf{y})^{\left(\sum_{i=1}^N \frac{1}{\alpha_i}\right)} f(\mathbf{x}) \quad \text{with} \quad \sum_{i=1}^N \frac{1}{\alpha_i} = 1 \\
 &= f(\mathbf{d}|\mathbf{y})^{\frac{1}{\alpha_N}} \cdots f(\mathbf{d}|\mathbf{y})^{\frac{1}{\alpha_2}} f(\mathbf{d}|\mathbf{y})^{\frac{1}{\alpha_1}} f(\mathbf{x})
 \end{aligned}$$

We compute N ES steps with “inflated” observation errors.

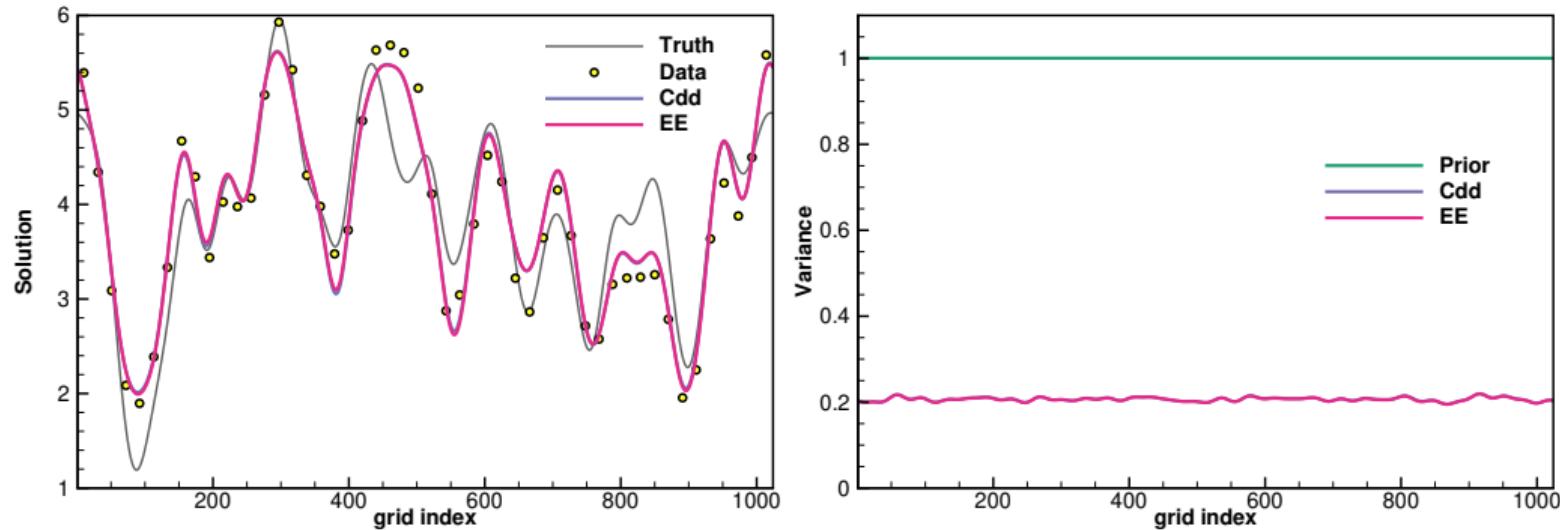
- Small updates reduce impact of the linear approximation.
- ESMDA is identical to ES in the linear case.
- Remember to resample measurement perturbations for each update step.

Some publications:

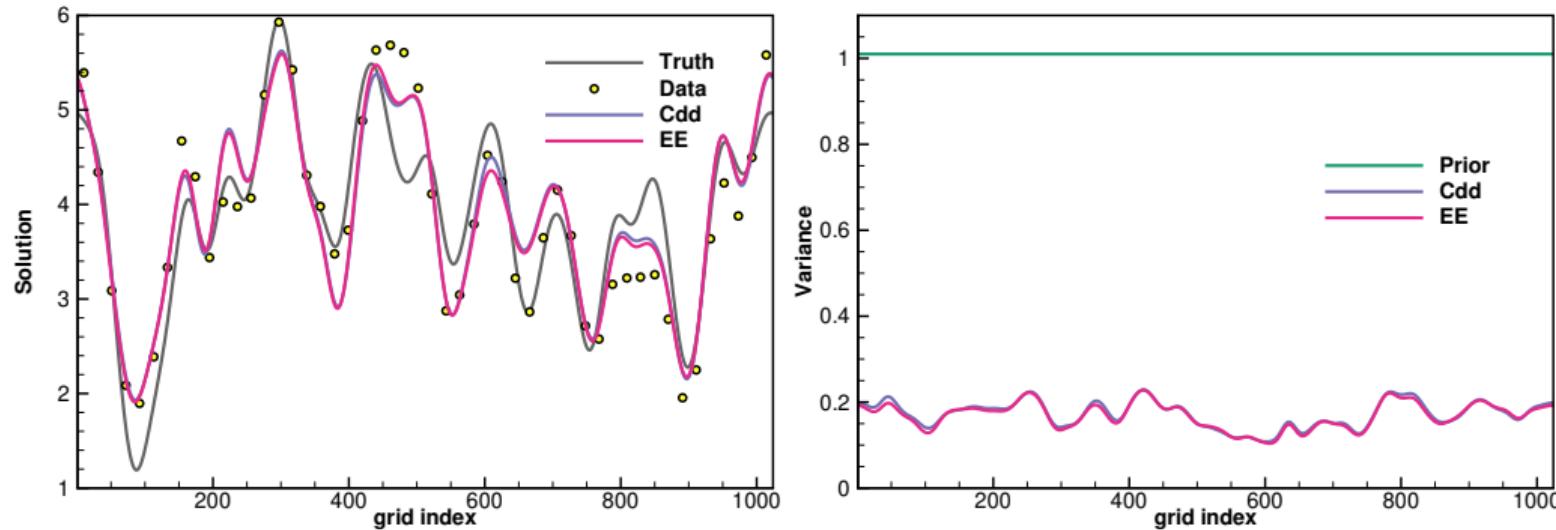
- Ensemble Randomized Maximum Likelihood EnRML (Chen and Oliver, 2013).
- Ensemble DA with multiple updates ESMDA (Emerick and Reynolds, 2013).
- Analysis of iterative ensemble smoothers (Evensen, 2018).
- IES with model errors (Evensen, 2019).
- Ensemble subspace RML (Evensen et al., 2019).

Linear EnKF update examples

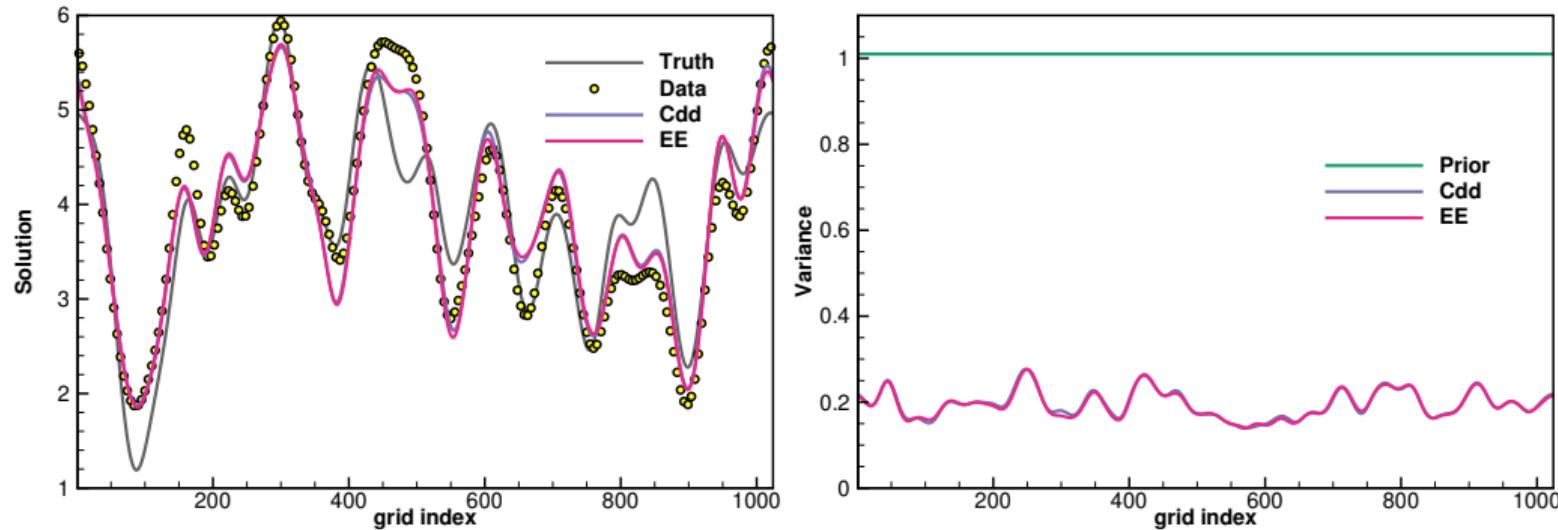
EnKF update with large ensemble size $N = 2000$ and $m = 50$



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

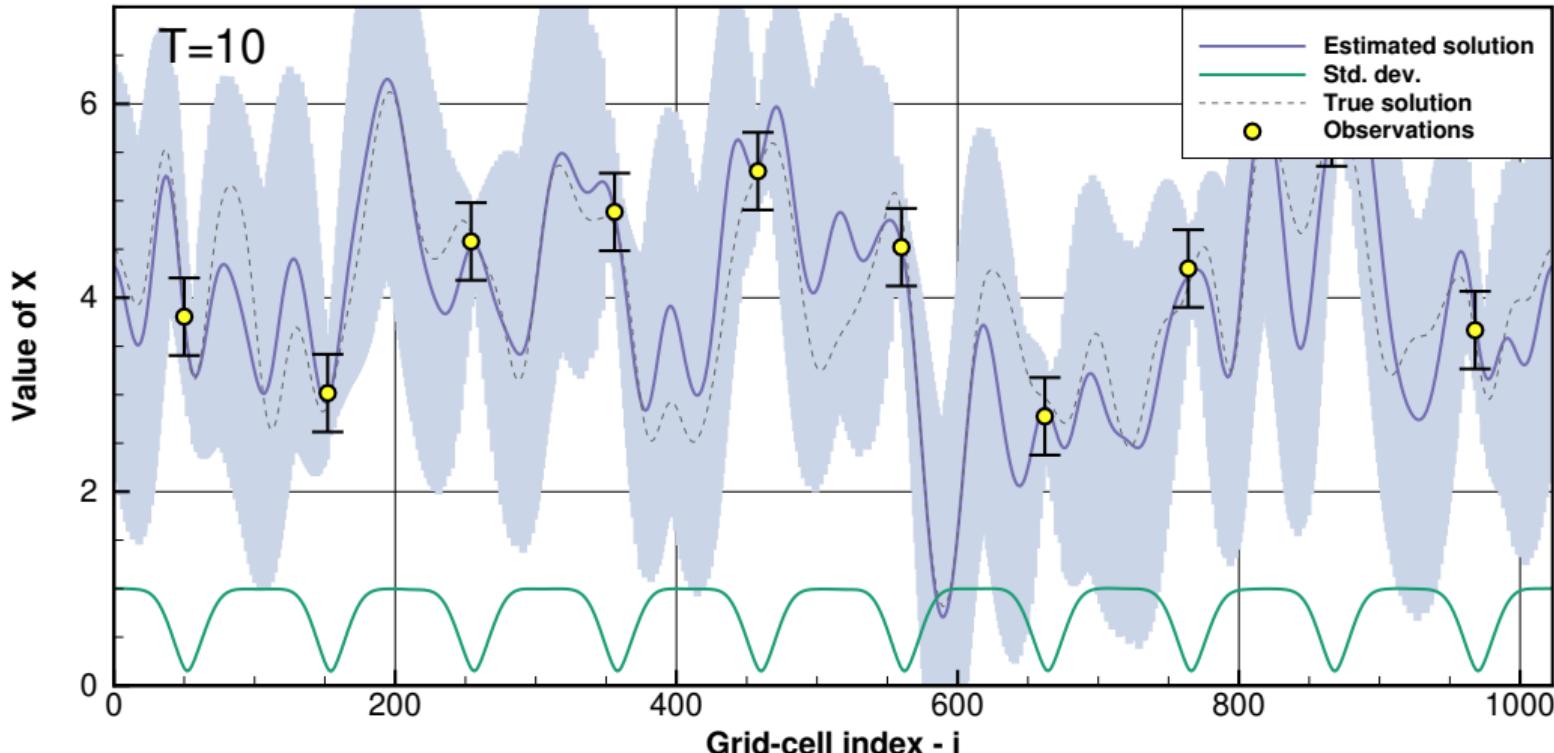


EnKF update with $N = 100$ and many measurements $m = 200$

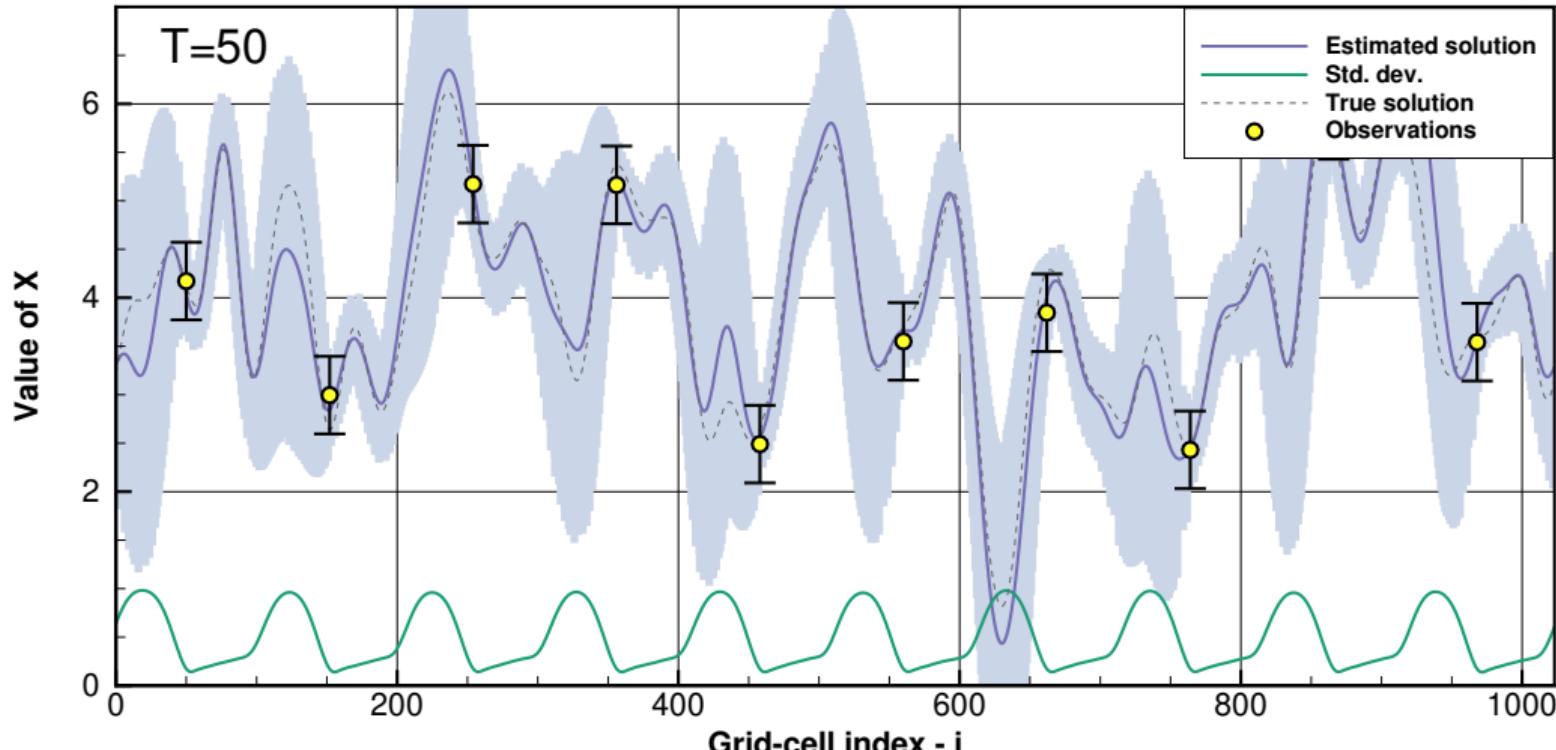


EnKF for an advection equation

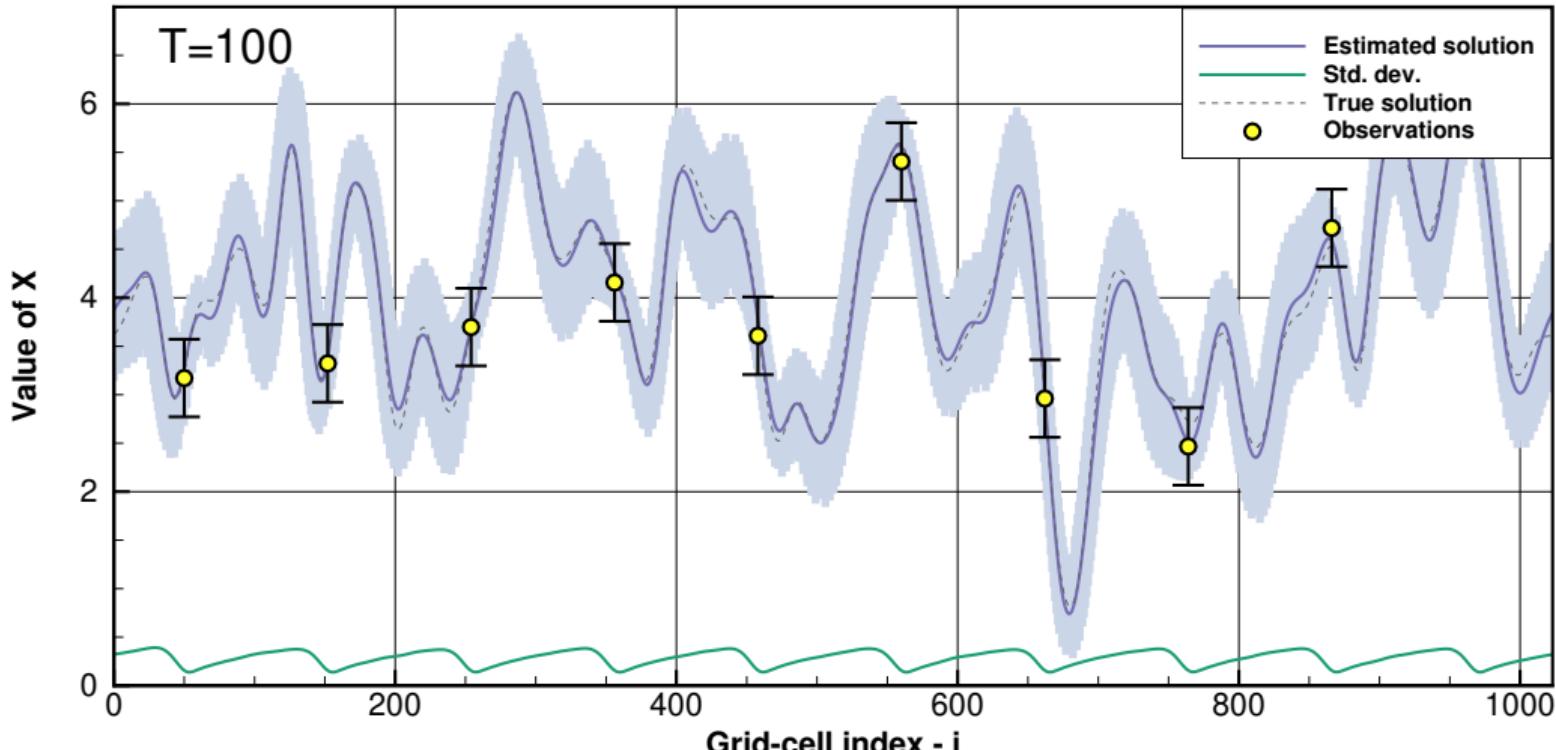
EnKF with the advection equation after two updates



EnKF with the advection equation after ten updates



EnKF with the advection equation after twenty updates

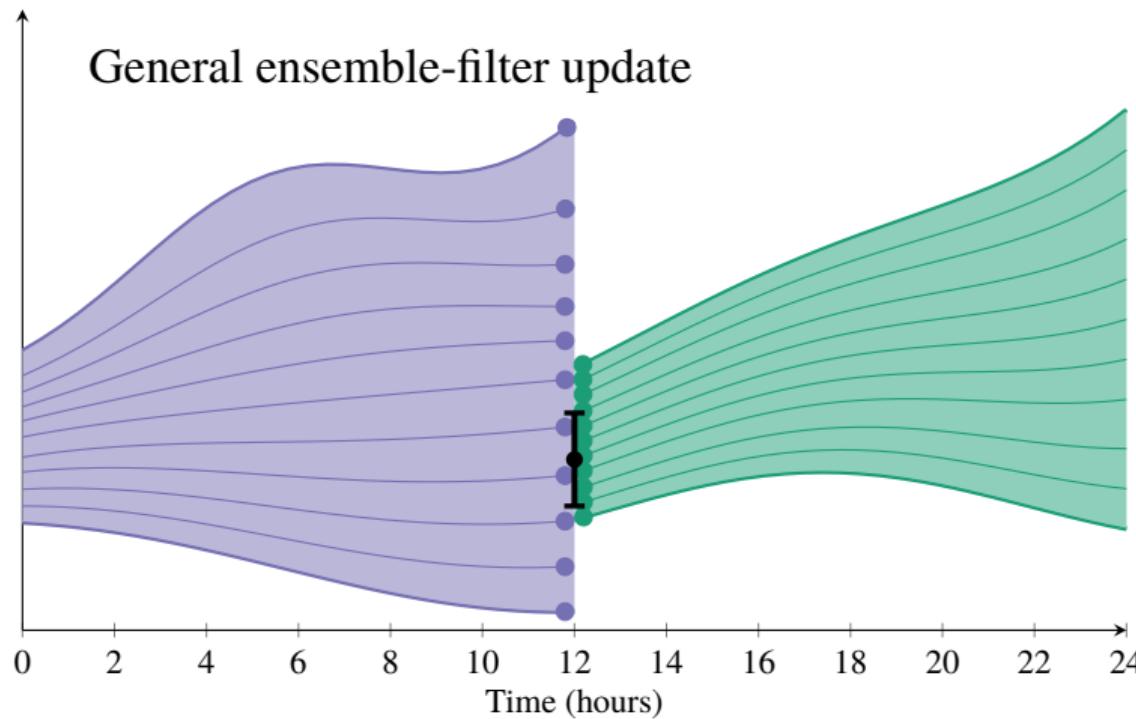


EnKF with the advection equation: Animation

Animation

EnKF with the Lorenz equations

General filter formulation



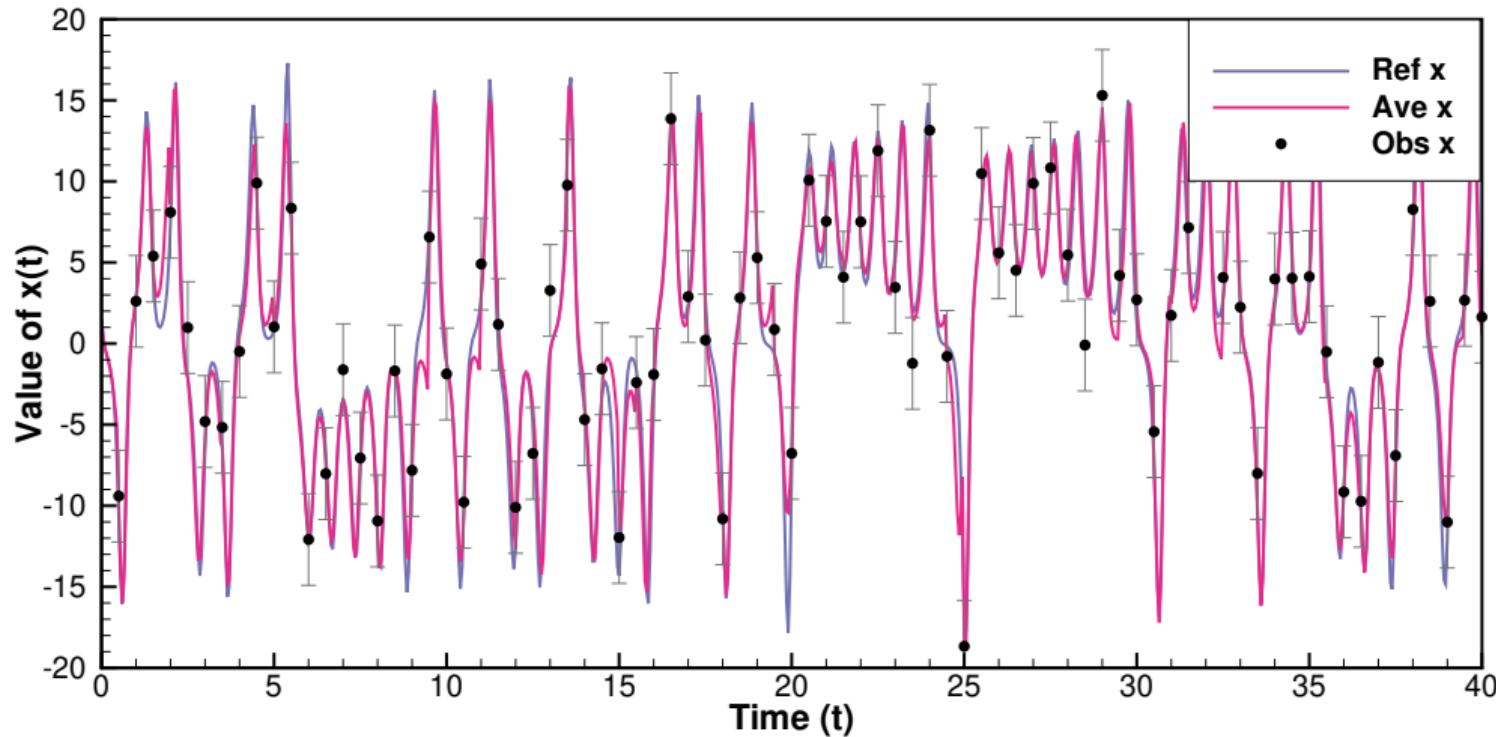
EnKF with the Lorenz model

$$\frac{\partial x}{\partial t} = \sigma(y - x), \quad (169)$$

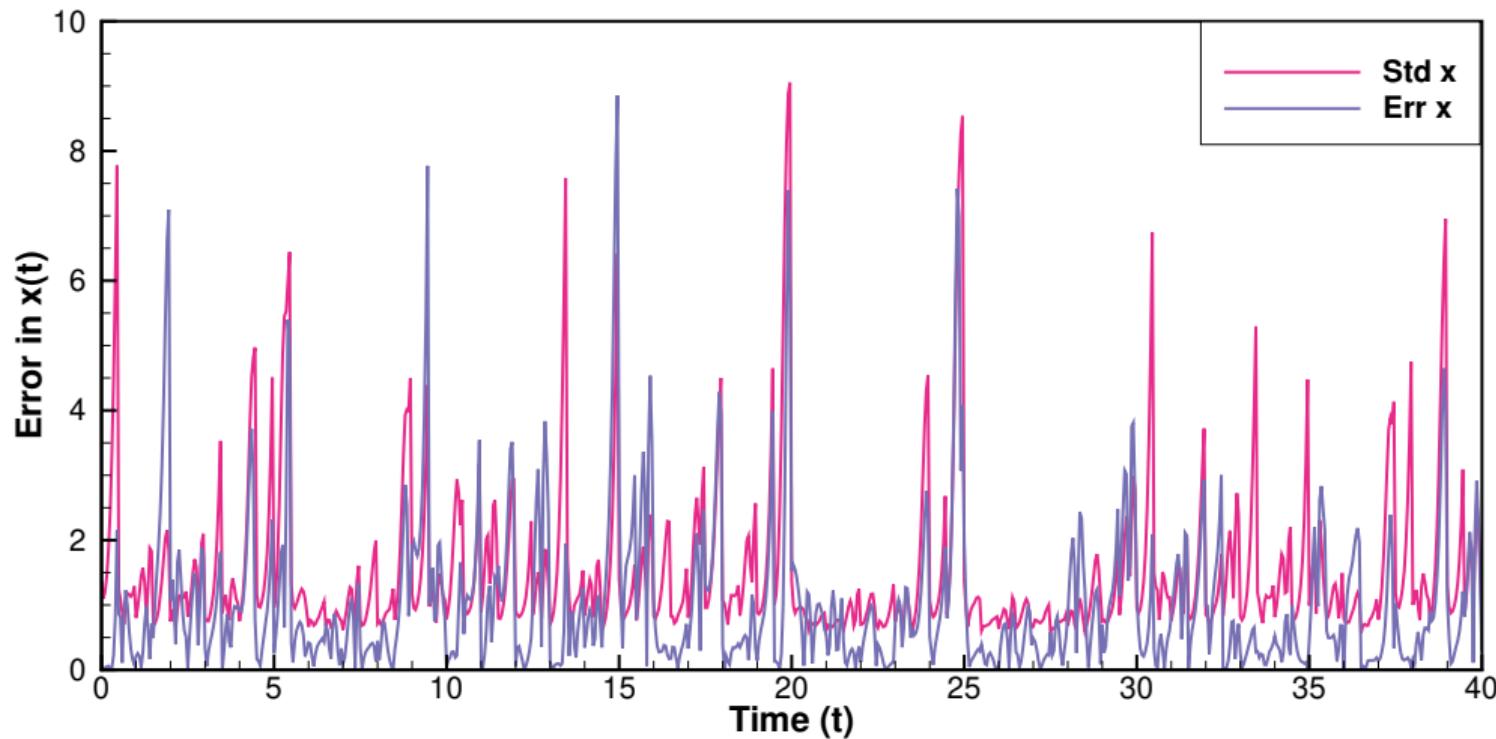
$$\frac{\partial y}{\partial t} = \rho x - y - xz, \quad (170)$$

$$\frac{\partial z}{\partial t} = xy - \beta z. \quad (171)$$

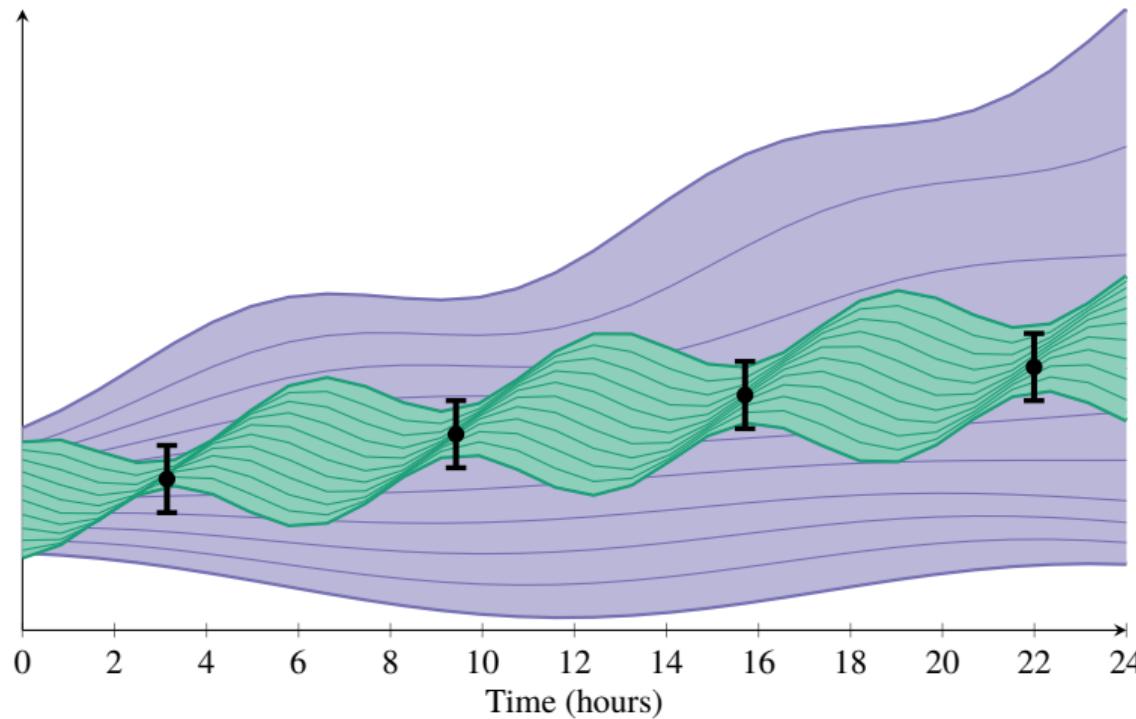
EnKF with the Lorenz model: estimate



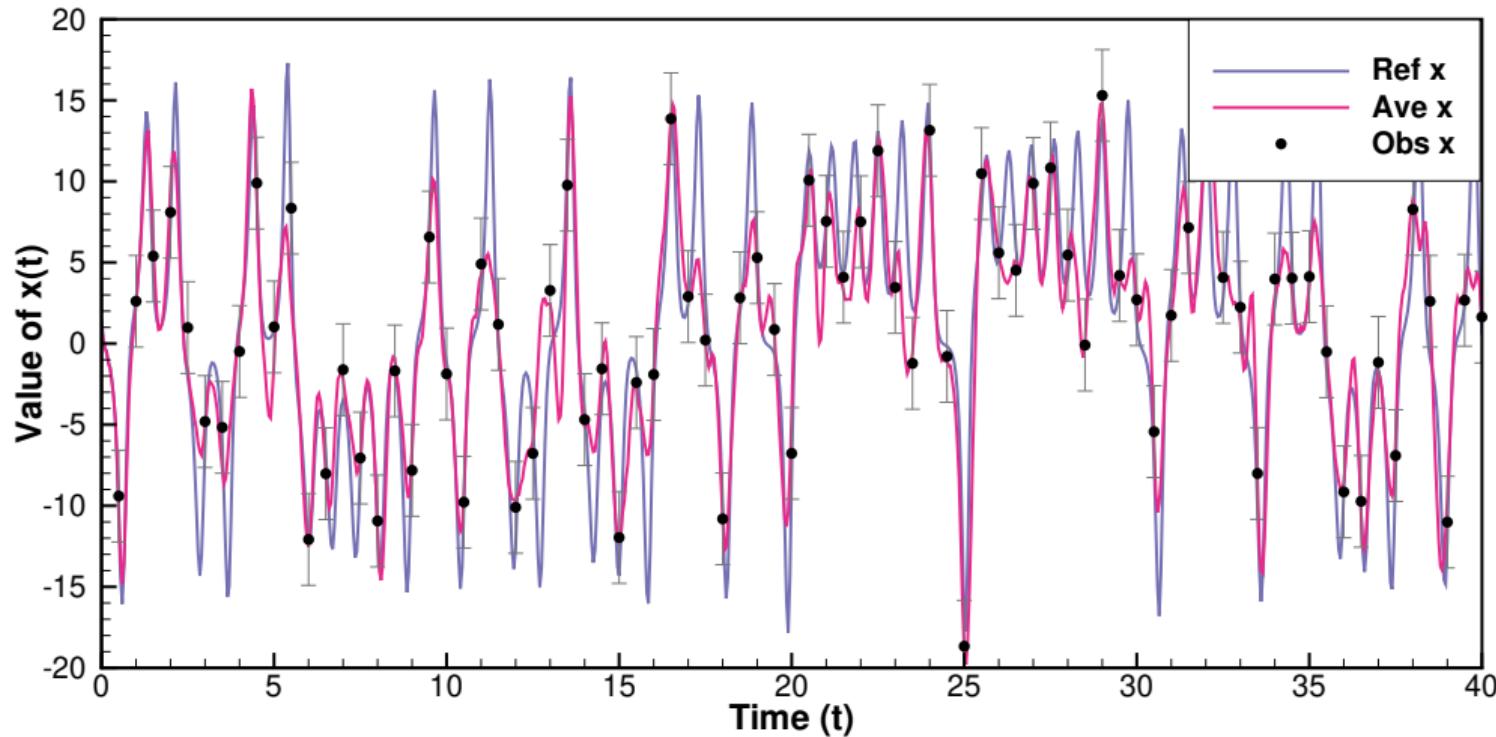
EnKF with the Lorenz model: error estimate



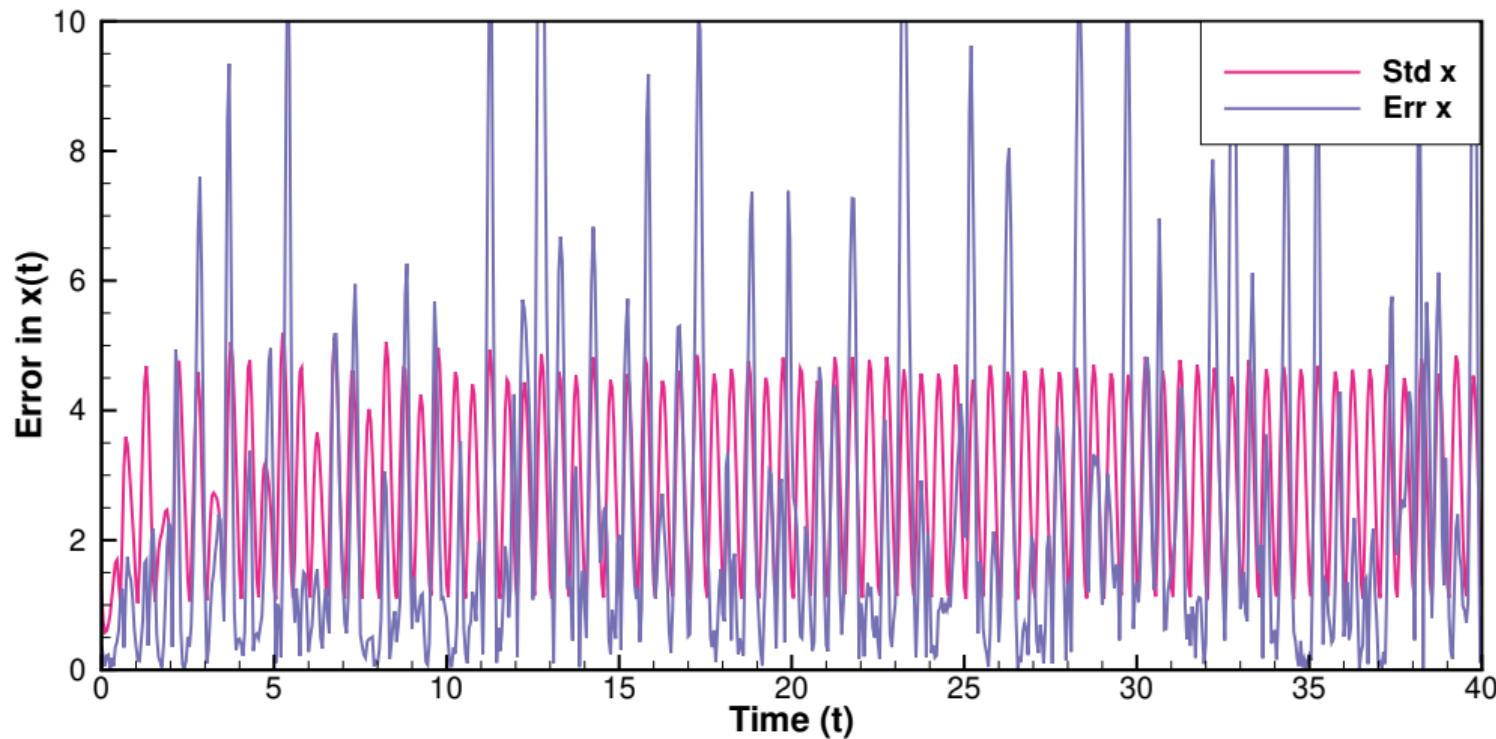
General smoother formulation



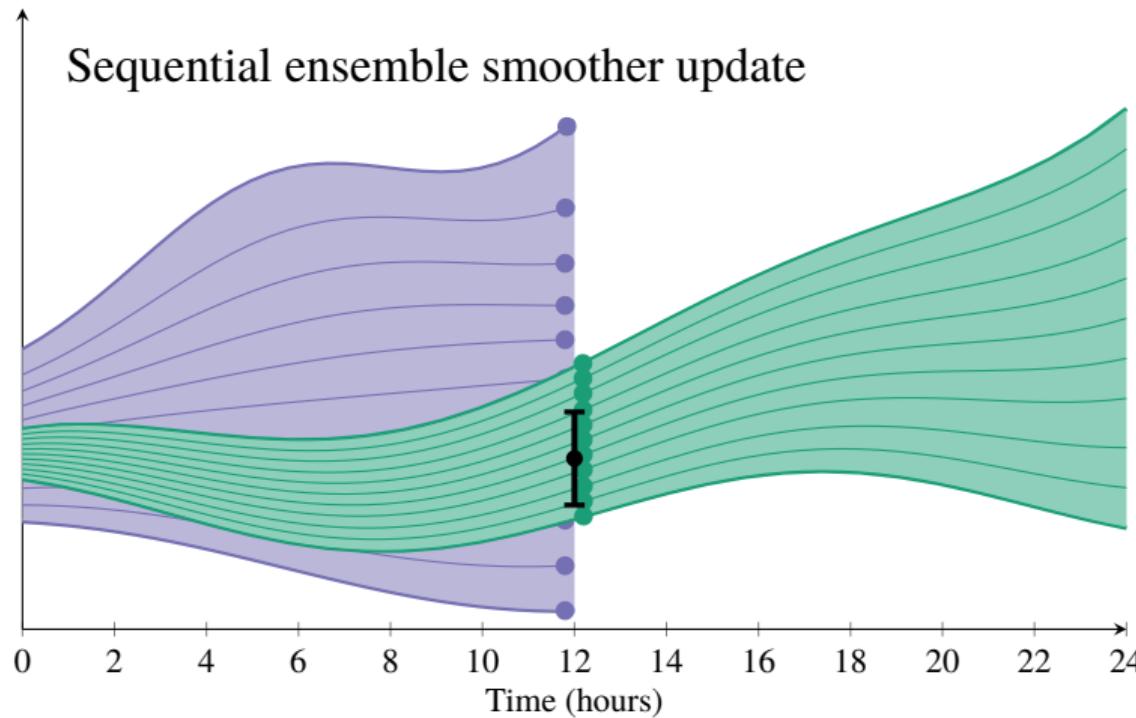
ES with the Lorenz model: estimate



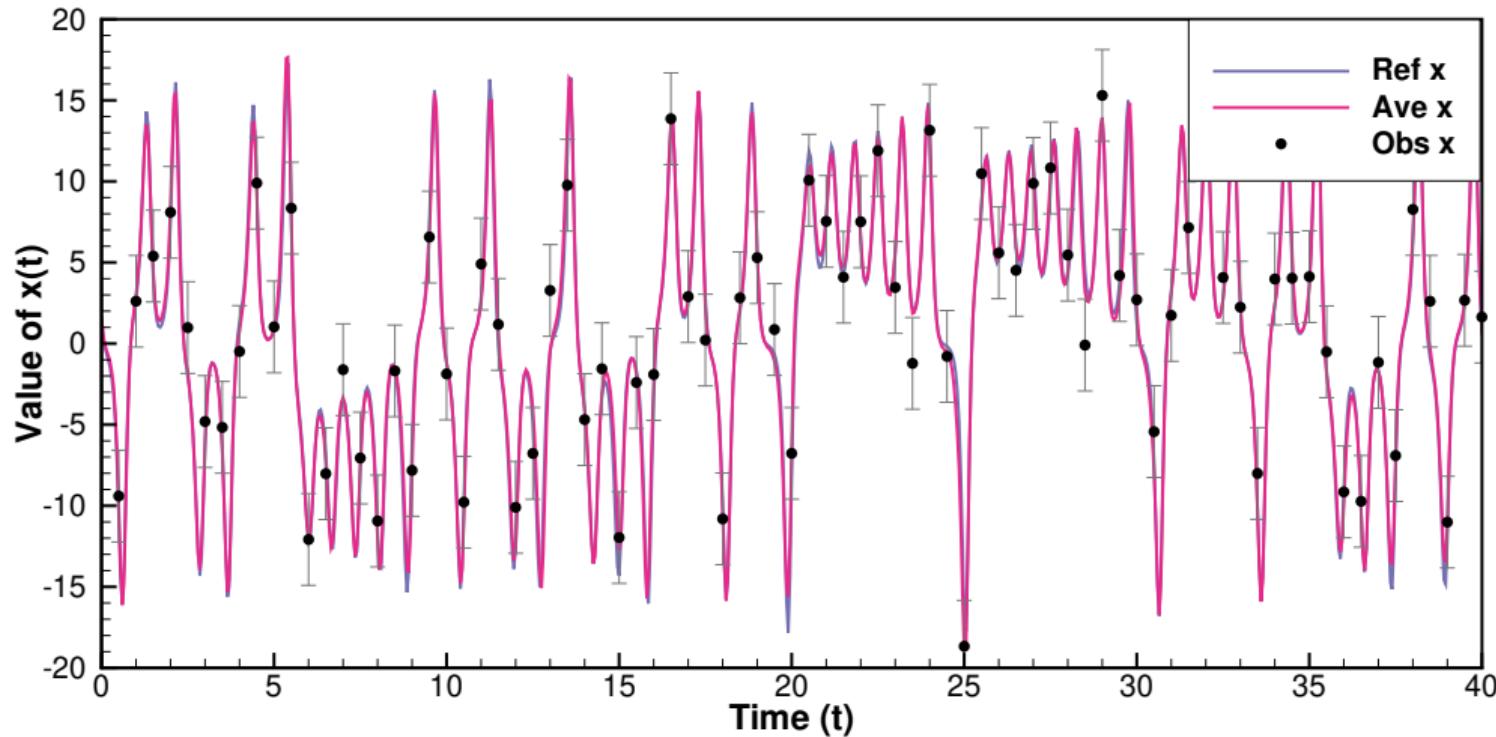
ES with the Lorenz model: error estimate



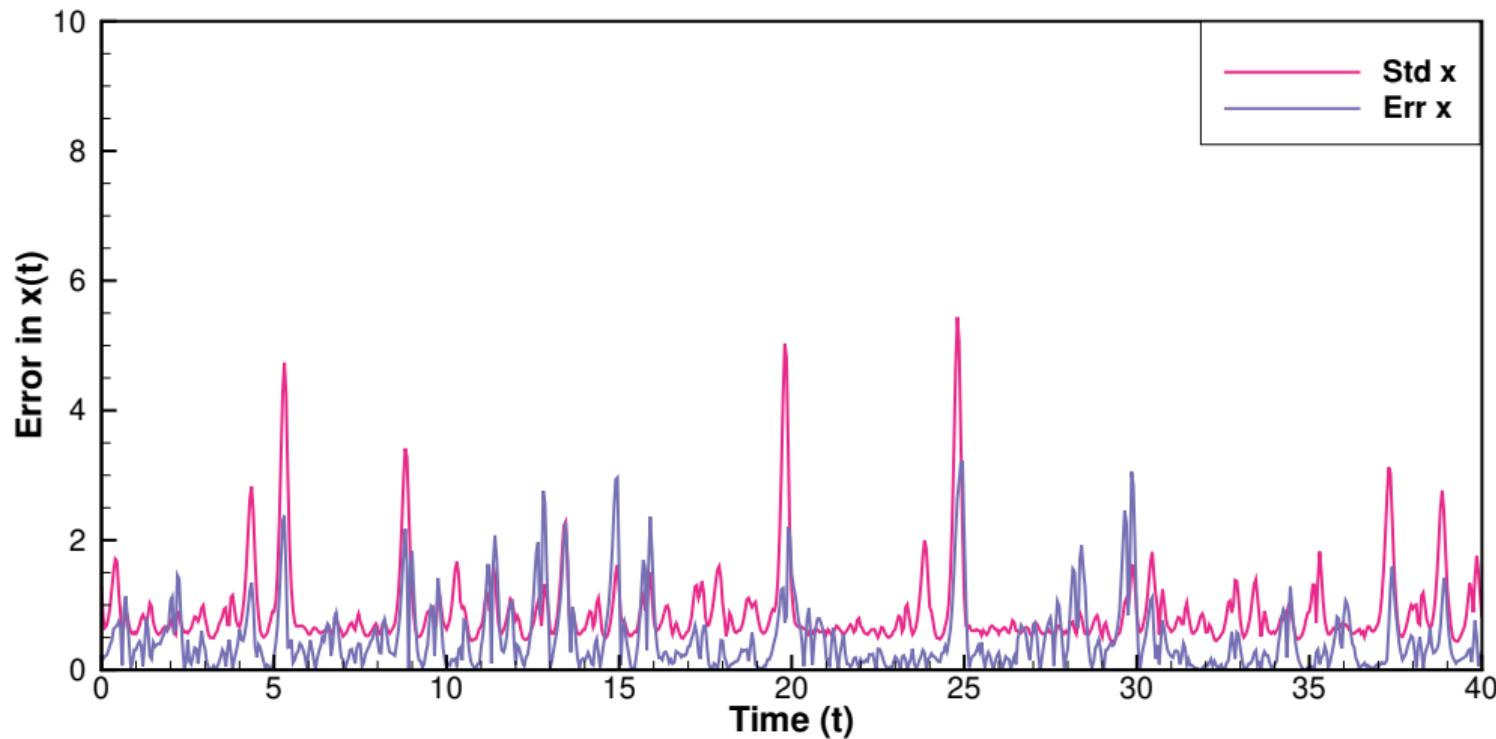
Recursive smoother formulation



EnKS with the Lorenz model: estimate



EnKS with the Lorenz model: error estimate



EnKF algorithms

Subspace EnKF update

```

1: subroutine EnKF_update( $Z, D, Y$ )
2: Input:  $Z \in \mathbb{R}^{n \times N}$ 
3: Input:  $D \in \mathbb{R}^{m \times N}$ 
4: Input:  $Y \in \mathbb{R}^{m \times N}$ 
5:  $\Pi = \left( I - \frac{1}{N} \mathbf{1} \mathbf{1}^T \right) / \sqrt{N-1}$ 
6:  $E = D \Pi$ 
7:  $Y = Y \Pi$ 
8: if  $n < N - 1$  then
9:    $Y = Y A^\dagger A$ 
10: end if
11:  $S = Y$ 
12:  $W = S^T (S S^T + E E^T)^{-1} (D - Y)$ 
13:  $Z \leftarrow Z \left( I + W / \sqrt{N-1} \right)$ 

```

- ▷ Prior state-vector ensemble
- ▷ Perturbed measurements
- ▷ Predicted measurements
- ▷ $\Pi \in \mathbb{R}^{N \times N}$
- ▷ $E \in \mathbb{R}^{m \times N}$
- ▷ $Y \in \mathbb{R}^{m \times N}$
- ▷ $S \in \mathbb{R}^{m \times N}$
- ▷ $W \in \mathbb{R}^{N \times N}$
- ▷ Update returned in Z

Standard EnKF application

```

1: Input:  $\mathbf{Z} \in \mathbb{R}^{n \times N}$                                 ▷ Initial model state-vector ensemble
2: Input:  $\mathbf{D}_l \in \mathbb{R}^{m \times N}$                     ▷ Perturbed measurements for each assimilation window
3: for  $l = 1, \dots$  do                                ▷ Loop over assimilation windows
4:    $\mathbf{X}_0 = \mathbf{Z}$ 
5:   for  $k = 1, K$  do                                ▷ Ensemble integration
6:      $\mathbf{X}_k = \mathbf{m}(\mathbf{X}_{k-1})$ 
7:   end for
8:    $\mathbf{X} = [\mathbf{X}_0, \dots, \mathbf{X}_K]$ 
9:    $\mathbf{Y} = \mathbf{h}(\mathbf{X})$                                 ▷ Predicted measurements
10:  call EnKF_update( $\mathbf{X}_K, \mathbf{D}_l, \mathbf{Y}$ )
11:   $\mathbf{Z} = \mathbf{X}_K$                                 ▷ Define state vector for next assimilation window
12: end for

```

EnKF updating \mathbf{X}_0

```

1: Input:  $\mathbf{Z} \in \Re^{n \times N}$                                 ▷ Initial model state-vector ensemble
2: Input:  $\mathbf{D}_l \in \Re^{m \times N}$                       ▷ Perturbed measurements for each assimilation window
3: for  $l = 1, \dots$  do                                ▷ Loop over assimilation windows
4:    $\mathbf{X}_0 = \mathbf{Z}$                                      ▷ Ensemble integration
5:   for  $k = 1, K$  do                                ▷ Ensemble integration
6:      $\mathbf{X}_k = \mathbf{m}(\mathbf{X}_{k-1})$                   ▷ Predicted measurments
7:   end for
8:    $\mathbf{X} = [\mathbf{X}_0, \dots, \mathbf{X}_K]$ 
9:    $\mathbf{Y} = \mathbf{h}(\mathbf{X})$                                 ▷ Predicted measurments
10:  call EnKF_update( $\mathbf{X}_0, \mathbf{D}_l, \mathbf{Y}$ )           ▷ Rerun ensemble integration
11:  for  $k = 1, K$  do                                ▷ Rerun ensemble integration
12:     $\mathbf{X}_k = \mathbf{m}(\mathbf{X}_{k-1})$ 
13:  end for
14:   $\mathbf{Z} = \mathbf{X}_K$                                 ▷ Define state vector for next assimilation window
15: end for

```

ES updating \mathbf{X}

```

1: Input:  $\mathbf{Z} \in \Re^{n \times N}$                                 ▷ Initial model state-vector ensemble
2: Input:  $\mathbf{D}_l \in \Re^{m \times N}$                       ▷ Perturbed measurements for each assimilation window
3: for  $l = 1, \dots$  do                                ▷ Loop over assimilation windows
4:    $\mathbf{X}_0 = \mathbf{Z}$ 
5:   for  $k = 1, K$  do                                ▷ Ensemble integration
6:      $\mathbf{X}_k = \mathbf{m}(\mathbf{X}_{k-1})$ 
7:   end for
8:    $\mathbf{X} = [\mathbf{X}_0, \dots, \mathbf{X}_K]$ 
9:    $\mathbf{Y} = \mathbf{h}(\mathbf{X})$                                 ▷ Predicted measurements
10:  call EnKF_update( $\mathbf{X}, \mathbf{D}_l, \mathbf{Y}$ )
11:   $\mathbf{Z} = \mathbf{X}_K$                                 ▷ Define state vector for next assimilation window
12: end for

```

EnKS updating $[\mathbf{X}_l, \dots, \mathbf{X}_0]$

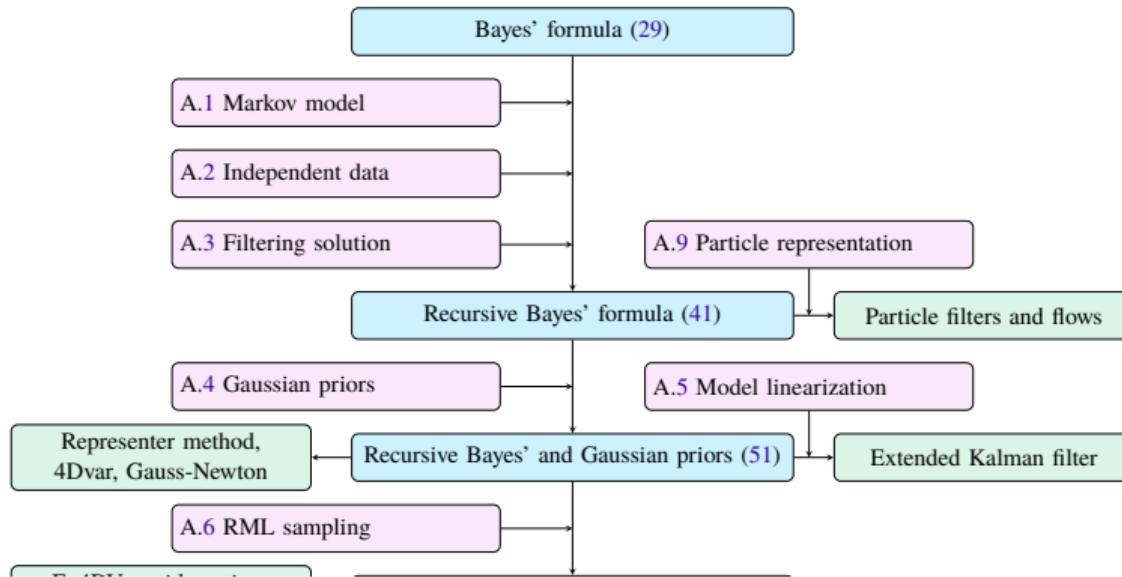
```

1: Input:  $\mathbf{Z} \in \Re^{n \times N}$ 
2: Input:  $\mathbf{D}_l \in \Re^{m \times N}$ 
3: for  $l = 1, \dots$  do
4:    $\mathbf{X}_0 = \mathbf{Z}$ 
5:   for  $k = 1, K$  do
6:      $\mathbf{X}_k = \mathbf{m}(\mathbf{X}_{k-1})$ 
7:   end for
8:    $\mathbf{X} = [\mathbf{X}_0, \dots, \mathbf{X}_K]$ 
9:    $\mathbf{Y} = \mathbf{h}(\mathbf{X})$ 
10:   $\mathcal{X}_l = [\mathcal{X}_{l-1}, \mathbf{X}]$ 
11:  call EnKF_update( $\mathcal{X}_l, \mathbf{D}_l, \mathbf{Y}$ )
12:   $\mathbf{X} = \mathcal{X}_l(l)$ 
13:   $\mathbf{Z} = \mathbf{X}_K$ 
14: end for

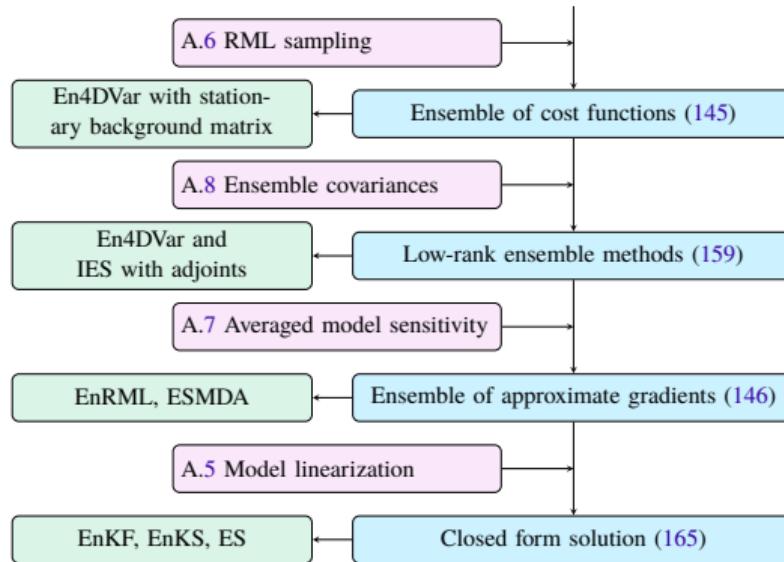
```

▷ Initial model state-vector ensemble
 ▷ Perturbed measurements for each assimilation window
 ▷ Loop over assimilation windows
 ▷ Ensemble integration
 ▷ Predicted measurements
 ▷ Define state vector for next assimilation window

Graphic overview



Graphic overview



Particle filters and flows

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

$$f(\mathbf{z}) \approx \sum_{j=1}^N \frac{1}{N} \delta(\mathbf{z} - \mathbf{z}_j), \quad (172)$$

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

Comparison of methods on a scalar model

Scalar models

$$y = g(x, q) = x + 0.3x^3 + q, \quad (173)$$

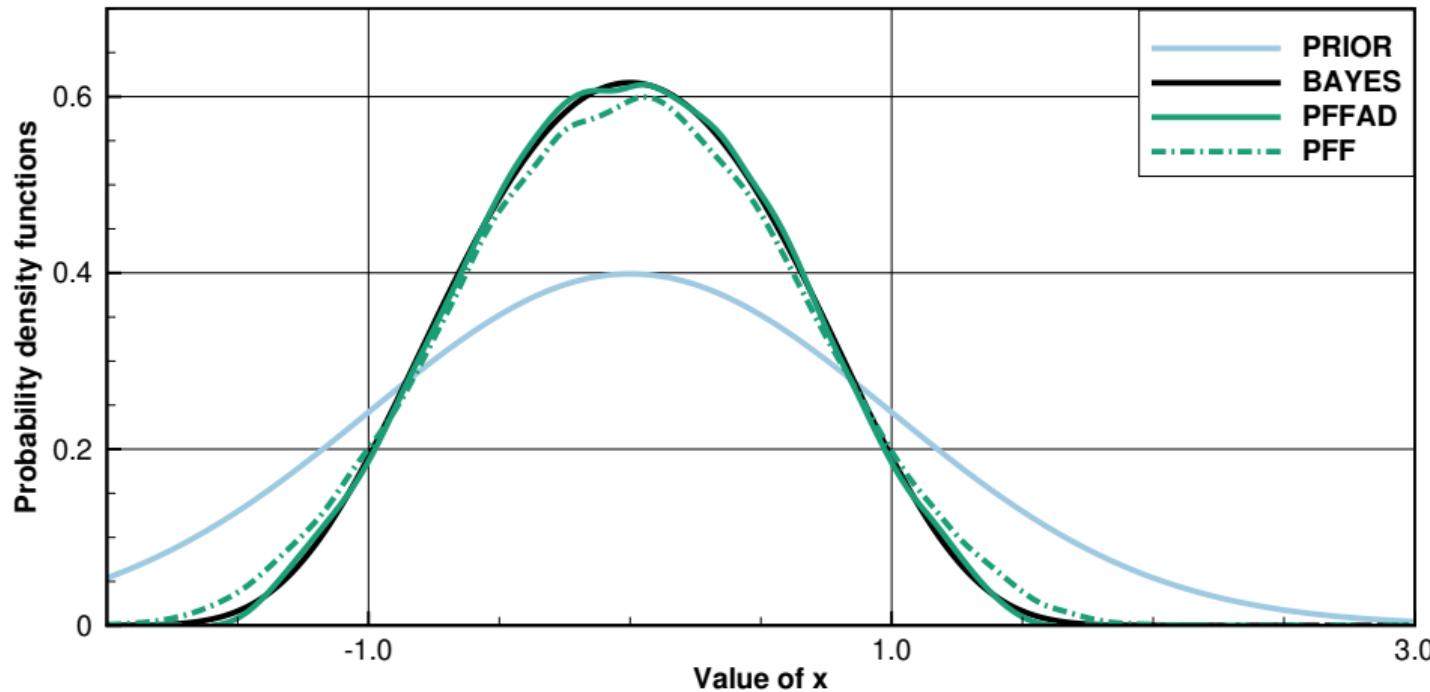
$$y = g(x, q) = 1 + \sin(x) + q, \quad (174)$$

Scalar examples

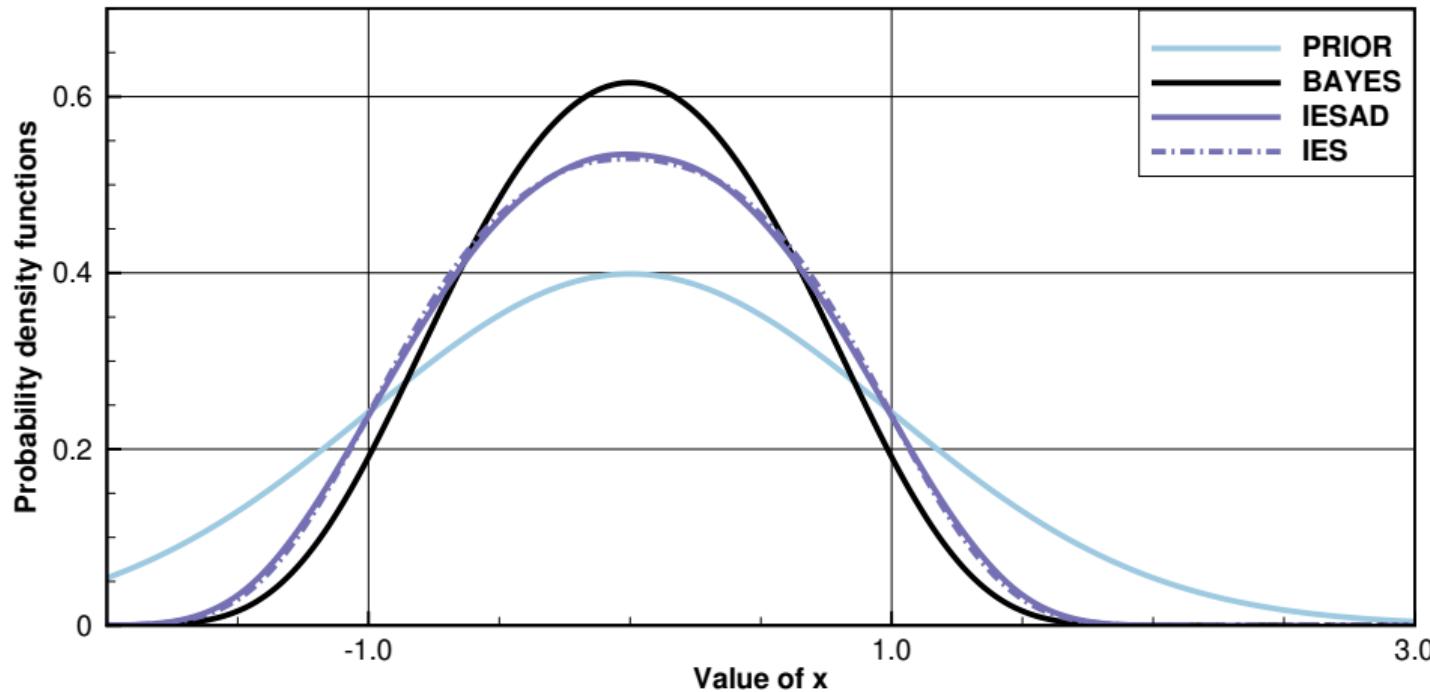
1. Model Eq. (173), $x_j^f = \mathcal{N}(x^f = 0.0, C_{xx} = 1.0)$, $d_j = \mathcal{N}(0.0, 1.0)$, and $q = \mathcal{N}(0, C_{qq} = 0.25)$.
2. Model Eq. (173), $x_j^f = \mathcal{N}(x^f = 1.0, C_{xx} = 1.0)$, $d_j = \mathcal{N}(0.0, -1.0)$, and $q = \mathcal{N}(0, C_{qq} = 0.25)$.
3. Model Eq. (174), $x_j^f = \mathcal{N}(x^f = 1.0, C_{xx} = 1.0)$, $d_j = \mathcal{N}(0.0, 1.0)$, and $q = \mathcal{N}(0, C_{qq} = 0.25)$.

Ensemble size $N = 10^7$.

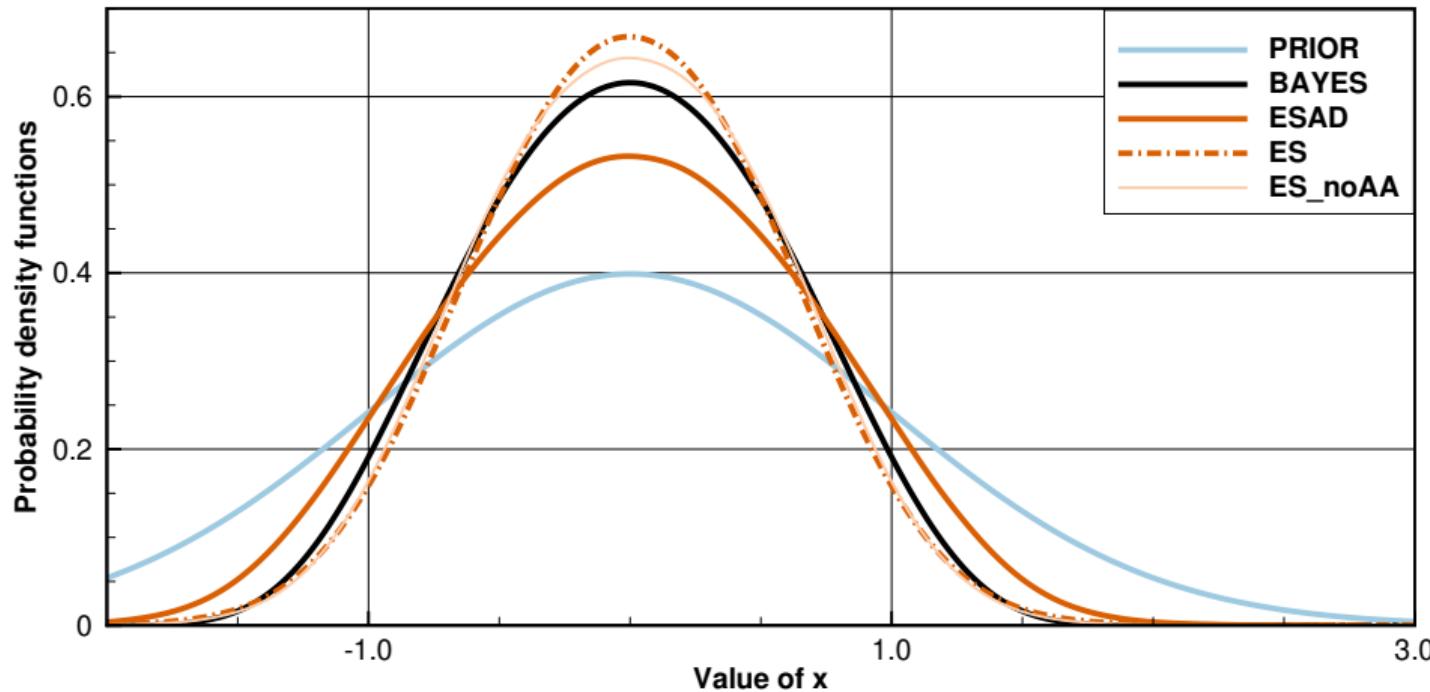
Case 1: Particle flow



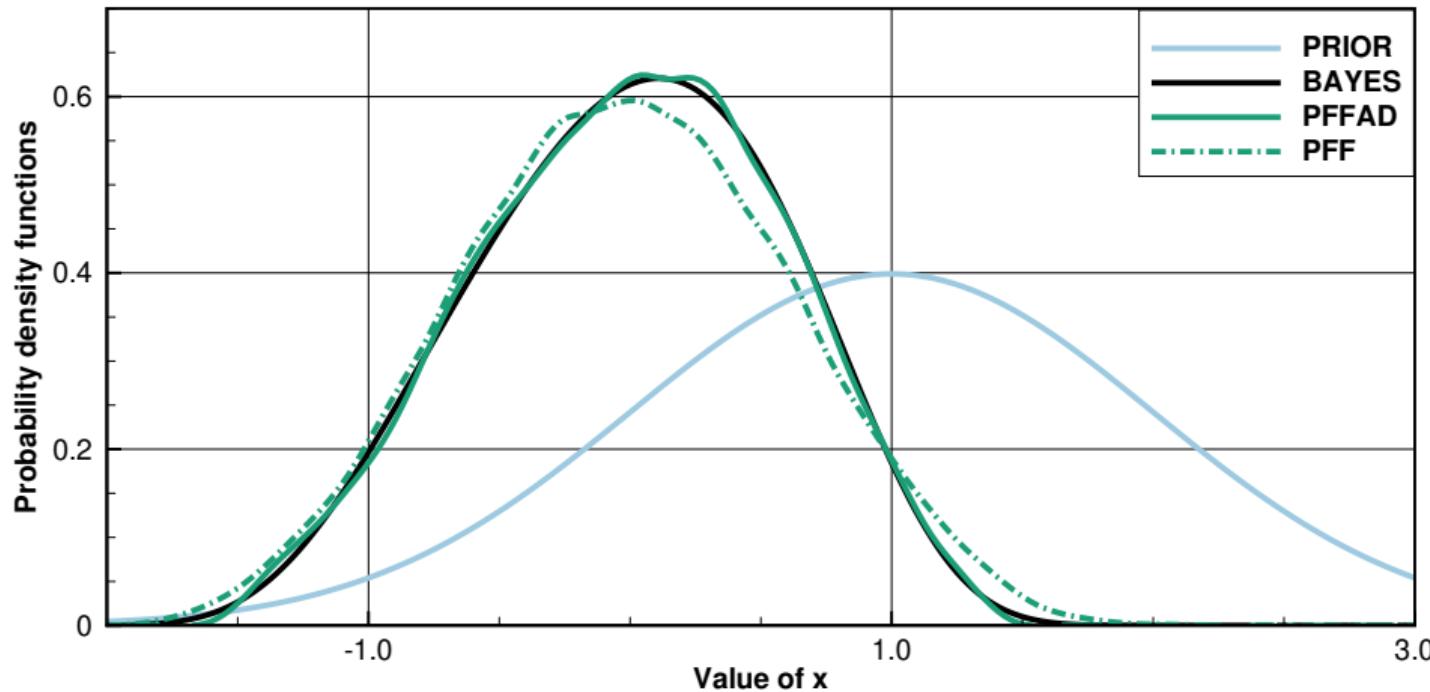
Case 1: IES



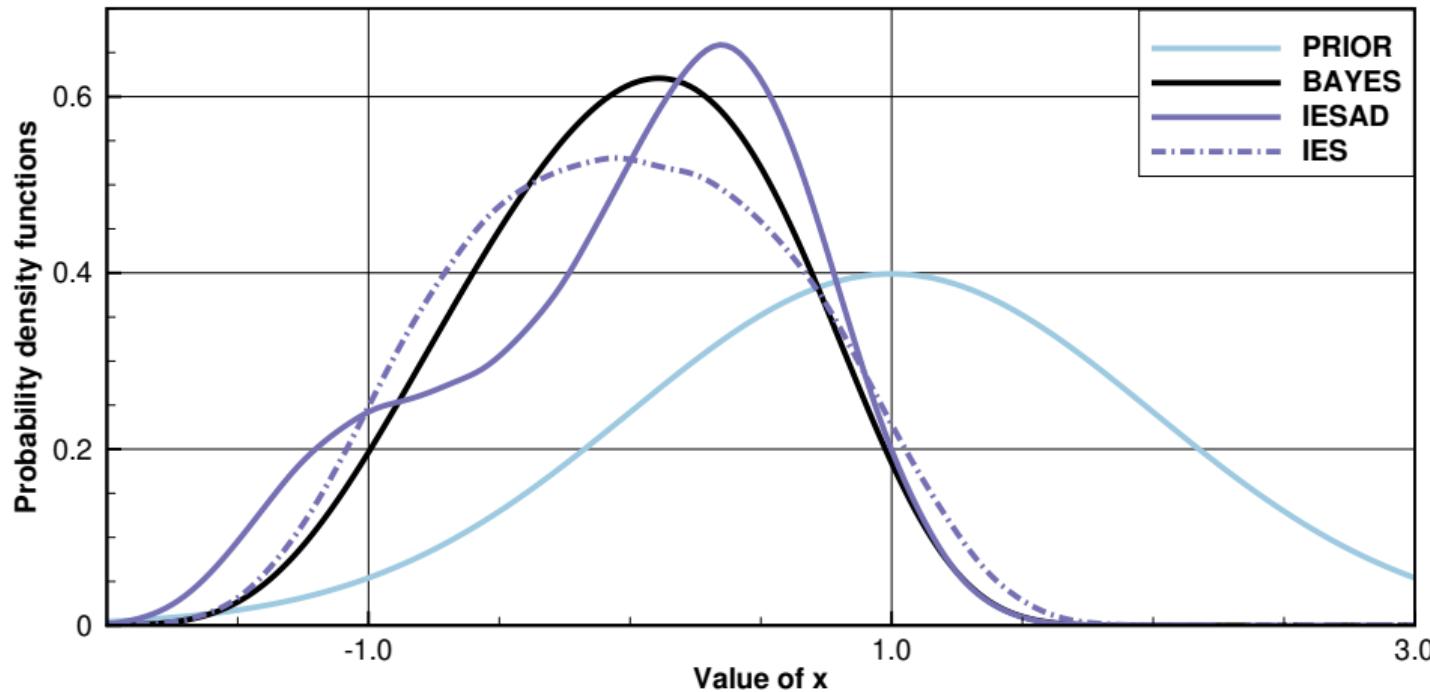
Case 1: EnKF



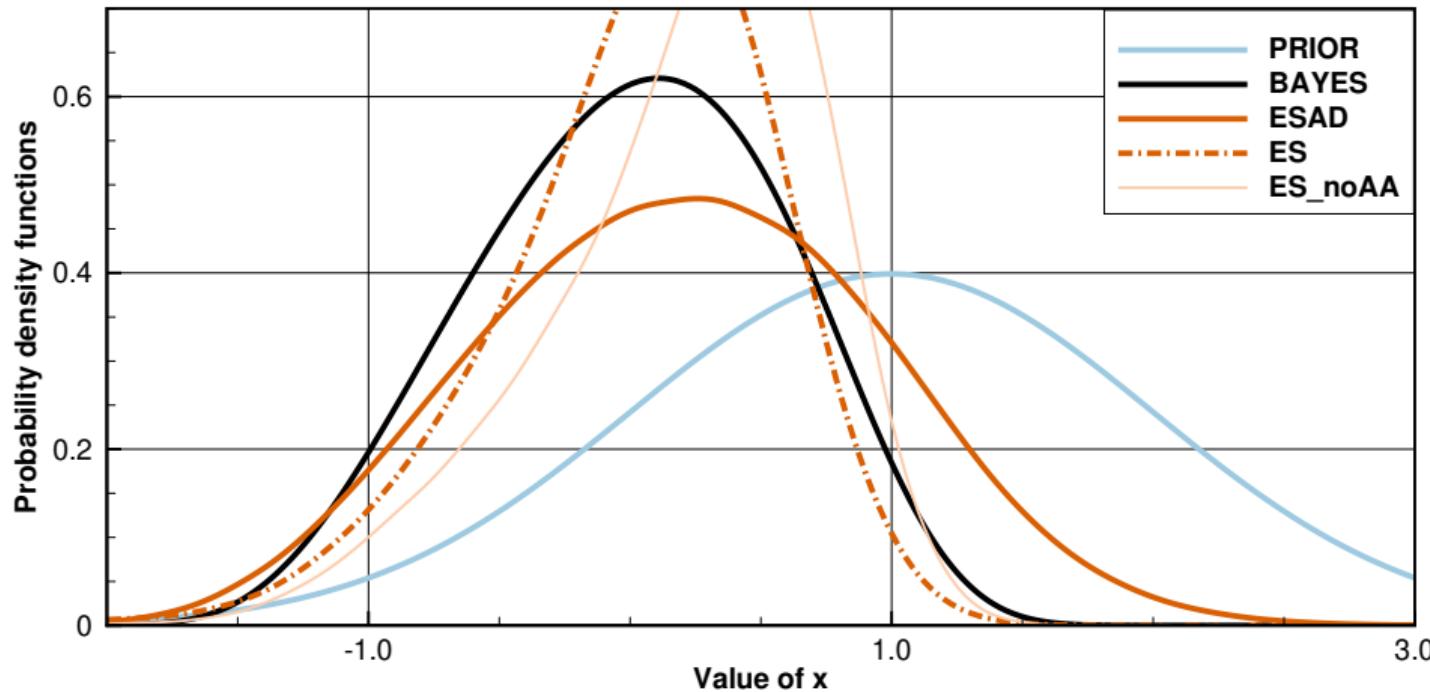
Case 2: Particle flow



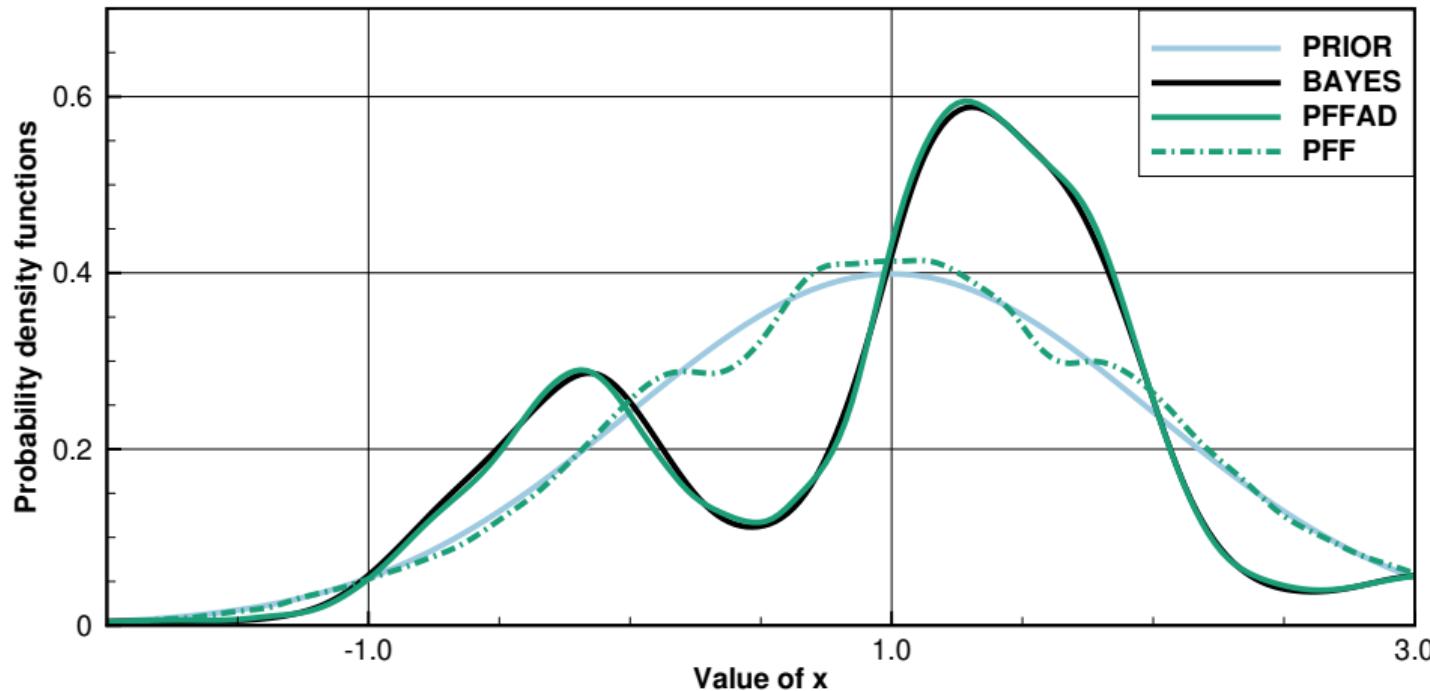
Case 2: IES



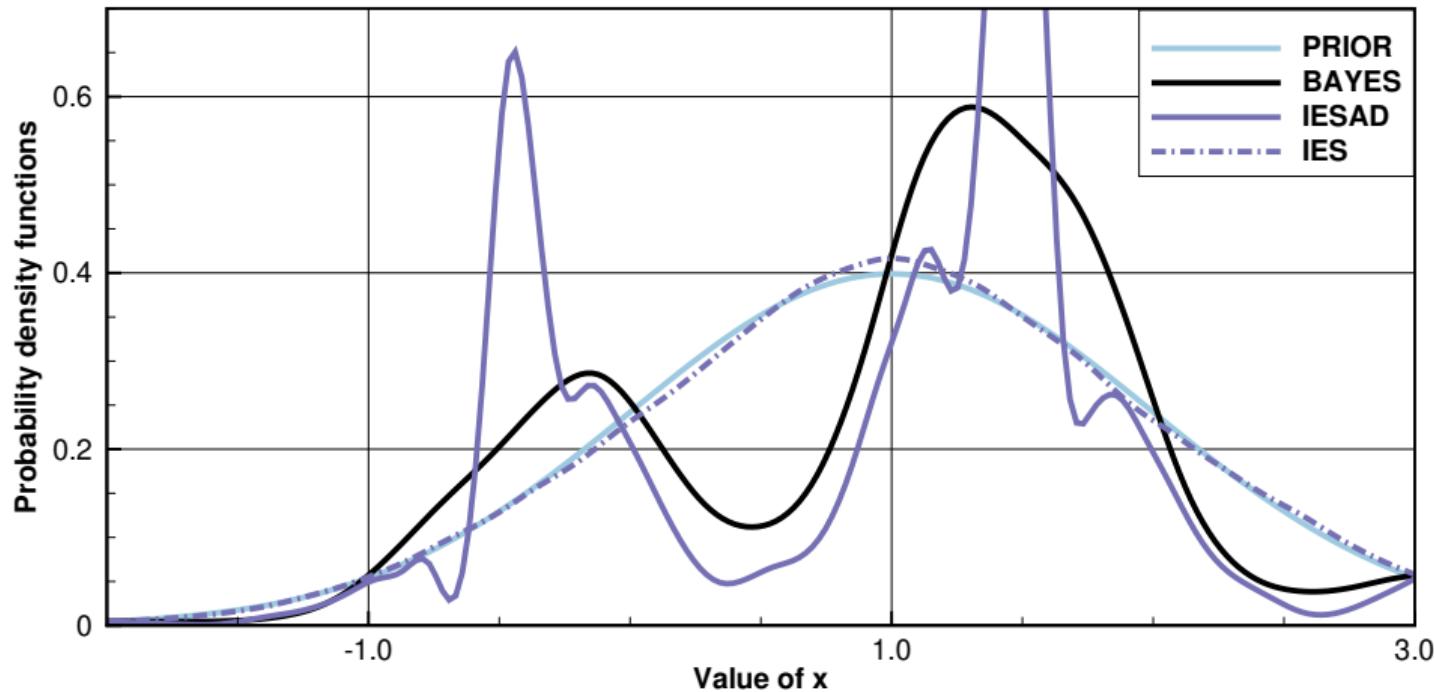
Case 2: EnKF



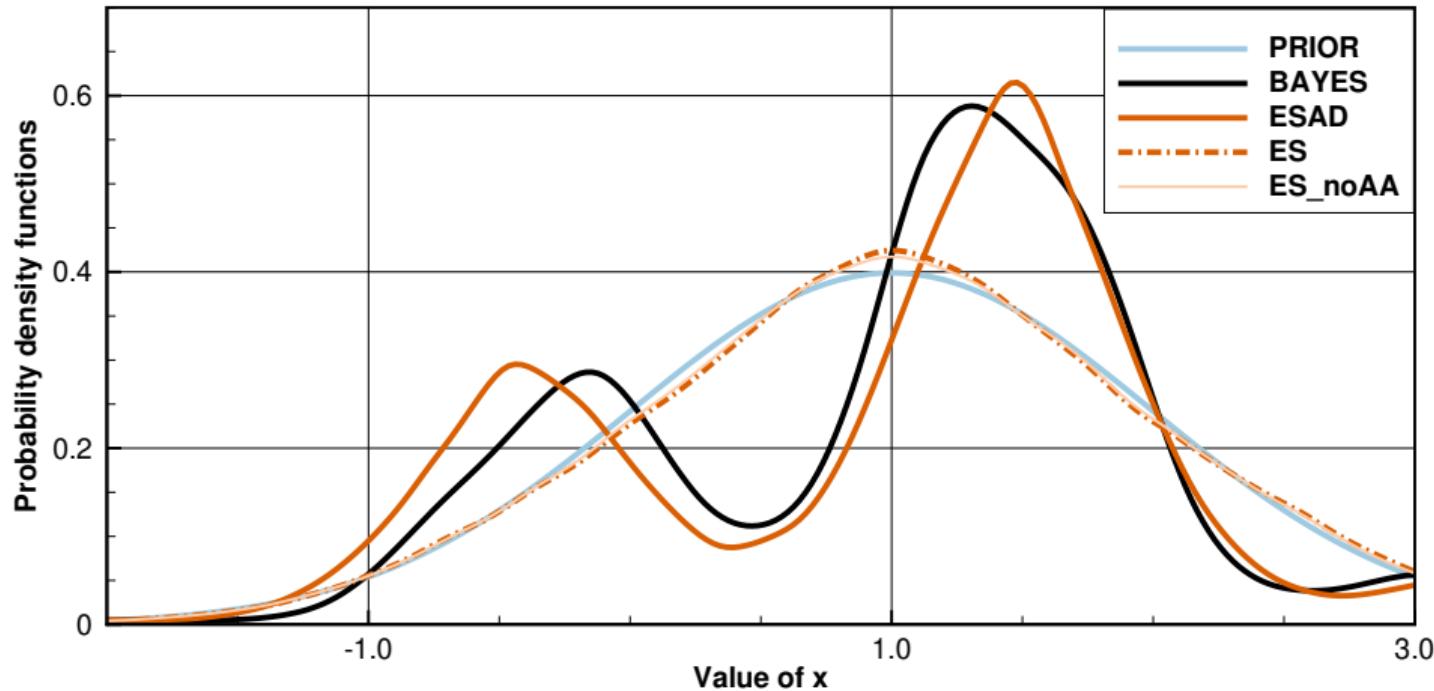
Case 3: Particle flow



Case 3: IES



Case 3: EnKF



ESMDA with a SARS-COV-2 pandemic nmodel

Available observations

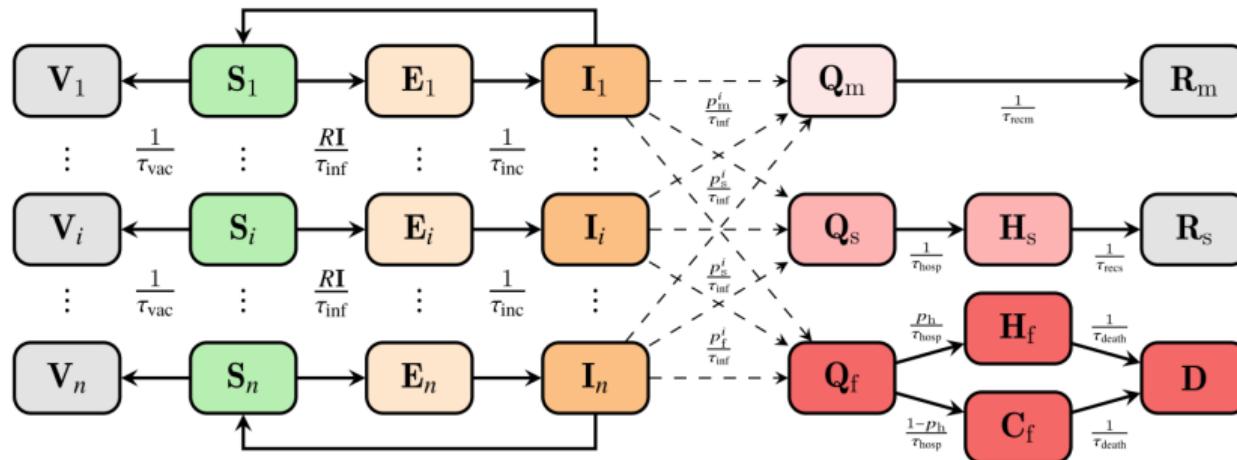
Hospitalized: Within different agegroups and gender.

Deaths: At hospitals or care homes.

Cases: Positive tests (highly underreported).

1. Data uncertainty and availability varies in different countries.
2. The SEIR model doesn't predict deaths and hospitalizations.

Extended SEIR model



- We add age classes to model age-specific infection and death rates.
- We differentiate between mild, severe, and fatal symptoms.
- We model those with fatal symptoms who die in care homes.

Extended SEIR model

$$\frac{\partial \mathbf{S}_i}{\partial t} = - \left(\sum_{j=1}^n \frac{R_{ij}(t) \mathbf{I}_j}{\tau_{\text{inf}}} \right) \mathbf{S}_i \quad (175)$$

$$\frac{\partial \mathbf{E}_i}{\partial t} = \left(\sum_{j=1}^n \frac{R_{ij}(t) \mathbf{I}_j}{\tau_{\text{inf}}} \right) \mathbf{S}_i - \frac{1}{\tau_{\text{inc}}} \mathbf{E}_i \quad (176)$$

$$\frac{\partial \mathbf{I}_i}{\partial t} = \frac{1}{\tau_{\text{inc}}} \mathbf{E}_i - \frac{1}{\tau_{\text{inf}}} \mathbf{I}_i \quad (177)$$

$$\frac{\partial \mathbf{Q}_m}{\partial t} = \sum_{i=1}^n \frac{p_m^i}{\tau_{\text{inf}}} \mathbf{I}_i - (1/\tau_{\text{recm}}) \mathbf{Q}_m \quad (178)$$

$$\frac{\partial \mathbf{Q}_s}{\partial t} = \sum_{i=1}^n \frac{p_s^i}{\tau_{\text{inf}}} \mathbf{I}_i - (1/\tau_{\text{hosp}}) \mathbf{Q}_s \quad (179)$$

$$\frac{\partial \mathbf{Q}_f}{\partial t} = \sum_{i=1}^n \frac{p_f^i}{\tau_{\text{inf}}} \mathbf{I}_i - (1/\tau_{\text{hosp}}) \mathbf{Q}_f \quad (180)$$

$$\frac{\partial \mathbf{H}_s}{\partial t} = \frac{1}{\tau_{\text{hosp}}} \mathbf{Q}_s - \frac{1}{\tau_{\text{recs}}} \mathbf{H}_s \quad (181)$$

$$\frac{\partial \mathbf{H}_f}{\partial t} = \frac{p_h}{\tau_{\text{hosp}}} \mathbf{Q}_f - \frac{1}{\tau_{\text{death}}} \mathbf{H}_f \quad (182)$$

$$\frac{\partial \mathbf{C}_f}{\partial t} = \frac{(1-p_h)}{\tau_{\text{hosp}}} \mathbf{Q}_f - \frac{1}{\tau_{\text{death}}} \mathbf{C}_f \quad (183)$$

$$\frac{\partial \mathbf{R}_m}{\partial t} = \frac{1}{\tau_{\text{recm}}} \mathbf{Q}_m \quad (184)$$

$$\frac{\partial \mathbf{R}_s}{\partial t} = \frac{1}{\tau_{\text{recs}}} \mathbf{H}_s \quad (185)$$

$$\frac{\partial \mathbf{D}}{\partial t} = \frac{1}{\tau_{\text{death}}} \mathbf{H}_f + \frac{1}{\tau_{\text{death}}} \mathbf{C}_f \quad (186)$$

Validity of the SEIR model

- Aggregated variables (statistical significance).
- Neglects import of cases (ok during lockdown).
- SEIR type models tend to successfully model epidemics.
- The simplicity is a huge advantage.

More complex models involve additional parameters.

Constant model parameters

1. Relative fractions p_m^i, p_s^i, p_f^i per age group.
2. Fractions dying in a Hospital p_h versus in a Care home $1 - p_h$.

Age group	1	2	3	4	5	6	7	8	9	10	11
Age range	0–5	6–12	13–19	20–29	30–39	40–49	50–59	60–69	70–79	80–89	90–105
p-mild	1.00	1.00	0.99	0.99	0.97	0.96	0.93	0.90	0.84	0.81	0.81
p-severe	0.00	0.00	0.00	0.00	0.02	0.02	0.05	0.08	0.11	0.11	0.11
p-fatal	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.03	0.06	0.06

Model parameters estimated by DA

Parameter	First guess	Description
τ_{inc}	5.5	Incubation period
τ_{inf}	3.8	Infection time
τ_{recm}	14.0	Recovery time mild cases
τ_{recs}	5.0	Recovery time severe cases
τ_{hosp}	6.0	Time until hospitalization
τ_{death}	16.0	Time until death
p_f	0.009	Case fatality rate
p_s	0.039	Hospitalization rate (severe cases)
I_0		Initial number of infectious
E_0		Initial number of exposed
$R(t)$		Effective reproductive number

Effective reproductive number

$$\mathbf{R}(t) = R(t)\hat{\mathbf{R}}$$

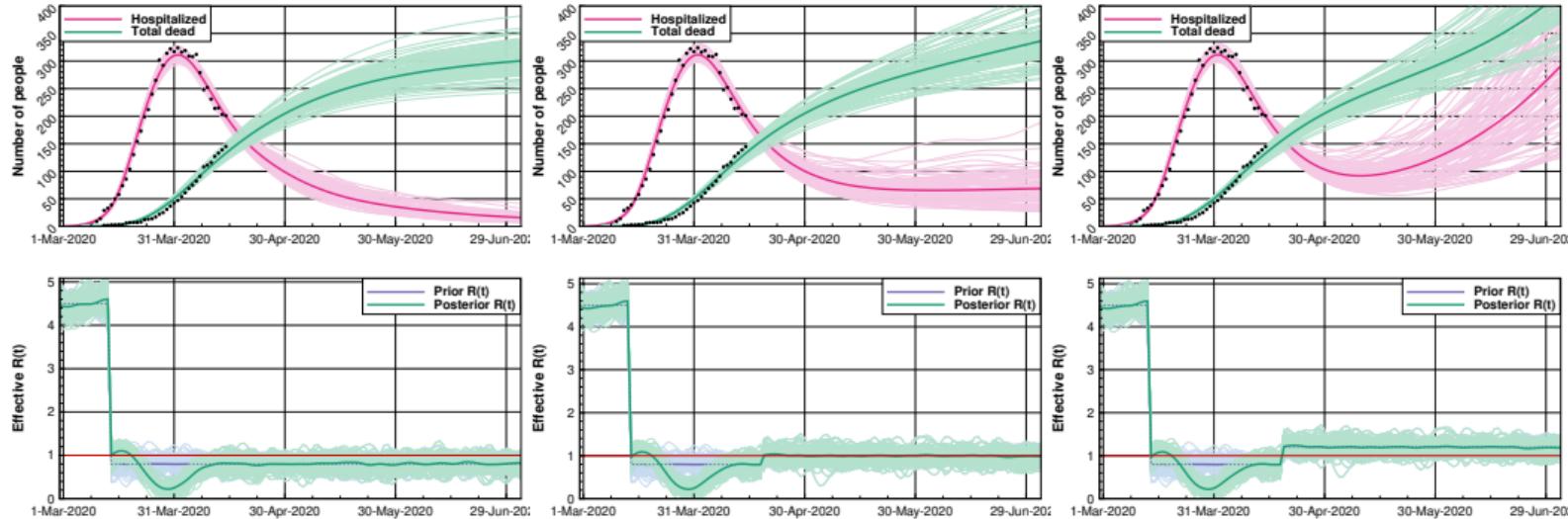
$\mathbf{R}(t)$ is a function of time (steered by how people isolate or interact).

- $R(t)$ is a scalar function of time.
- $\hat{\mathbf{R}}$ a constant matrix of transmissions between age classes..
- Behavior two weeks ago determines today's deaths and hospitalizations.
- We can estimate $R(t)$ for the past.
- We assume the value $R(t)$ for the future.

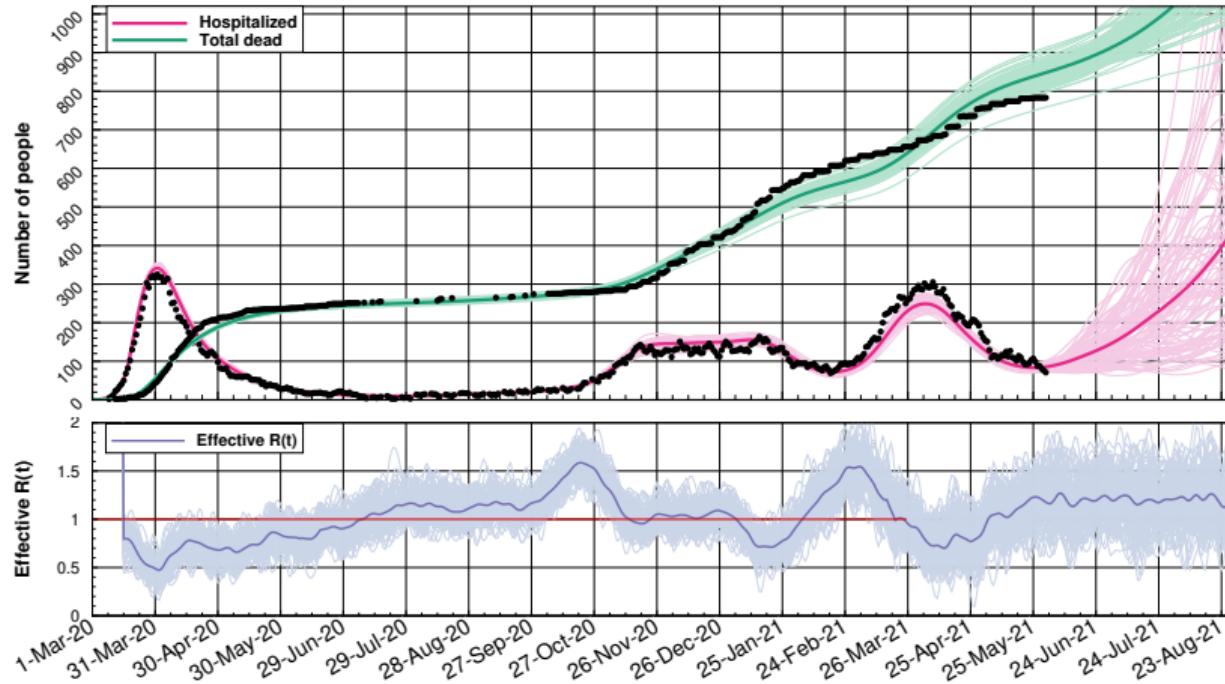
We used ESMDA

- Simple implementation and use.
- Efficient for large ensemble sizes.
- 5000 realizations and 32 ESMDA steps.

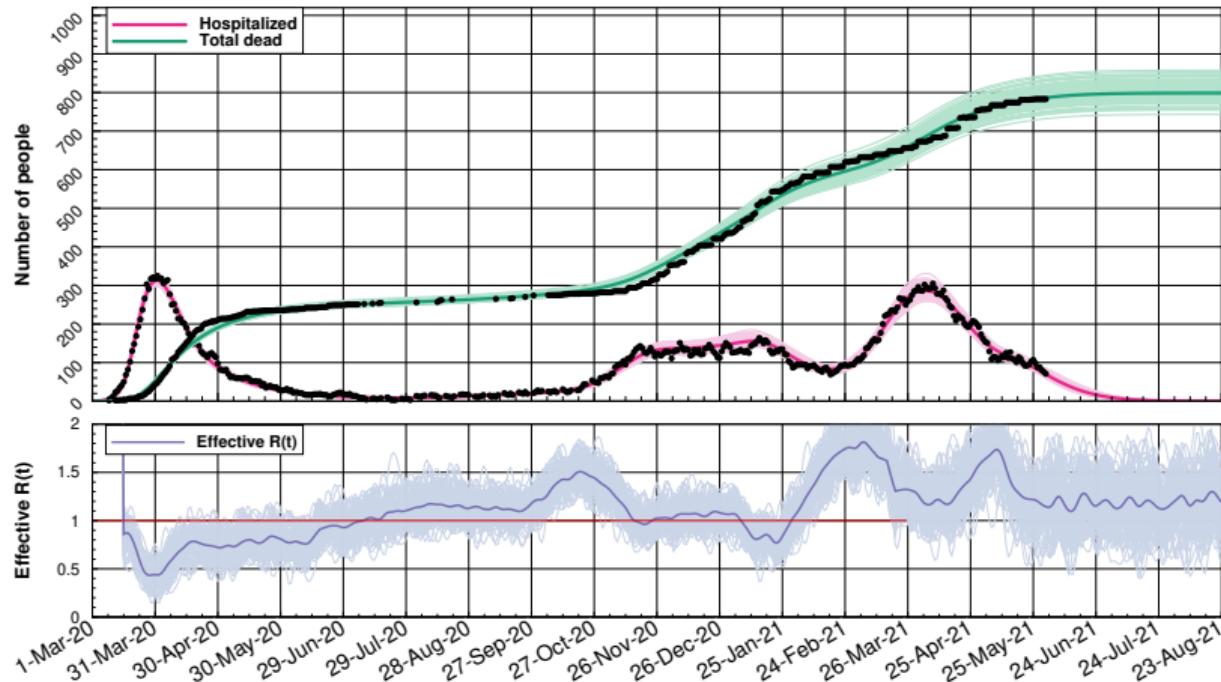
Back-to-school scenarios for Norway



Norway: prediction without vaccinations



Norway: prediction with vaccinations



Summary EnKF_seir

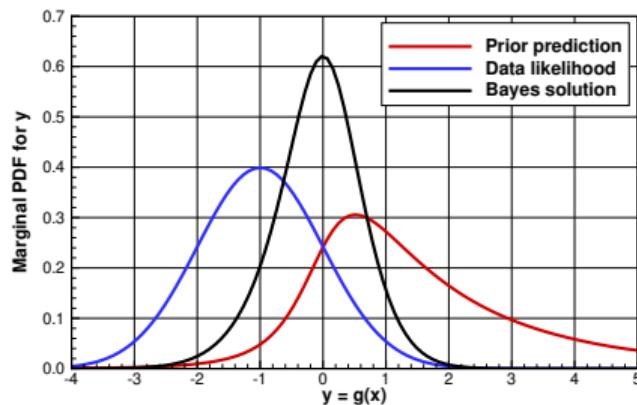
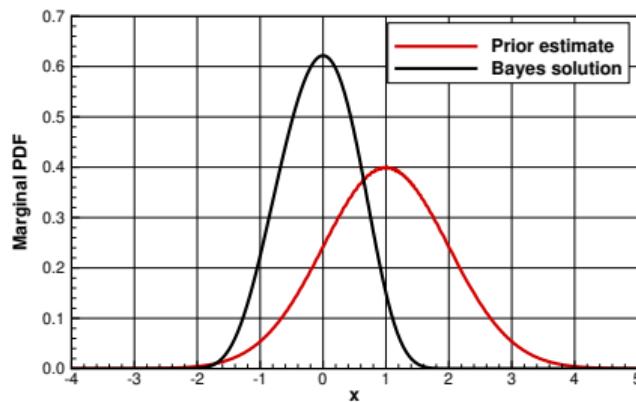
- The DA system tracks the epidemic accurately.
- We can estimate past $R(t)$.
- It is possible to quantify the impact of interventions.
- Short-term forecasting using R -persistence works well.
- Long-term scenario forecasting with specified future R .
- Code: https://github.com/geirev/EnKF_seir
- Paper: (Evensen et al., 2020)
<http://www.aimsceances.org/article/doi/10.3934/fods.2021001>

EnRML for History Matching Petroleum Models

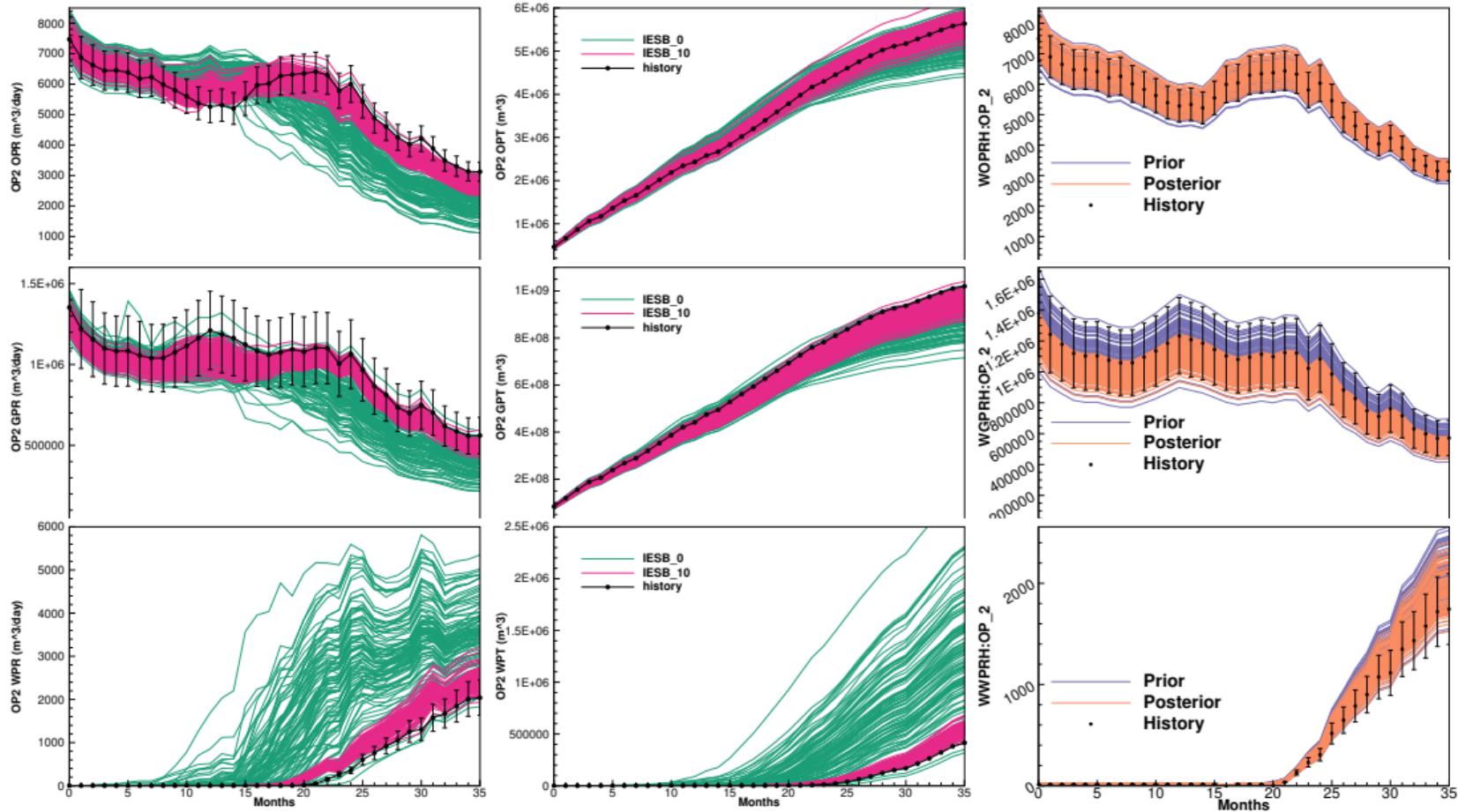
Illustration of the parameter estimation problem

Bayes theorem gives posterior probability function for parameters \mathbf{x}

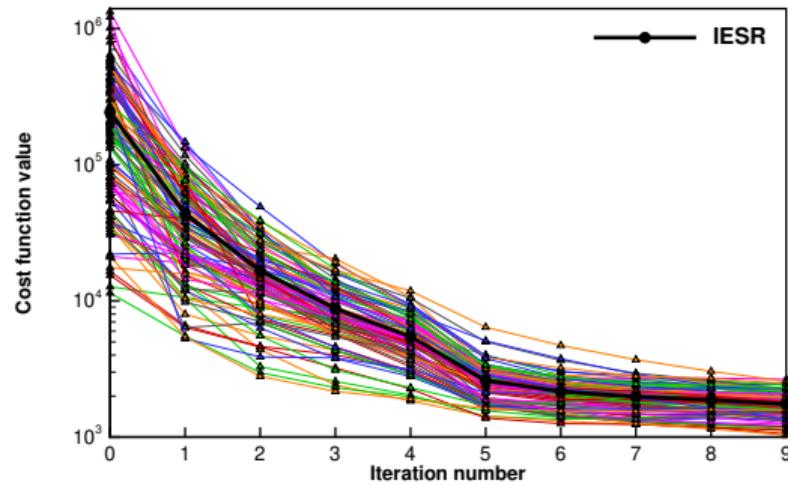
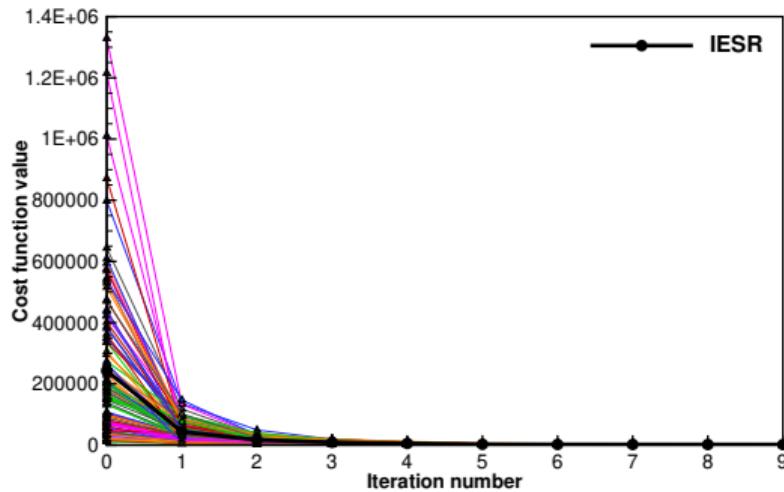
$$f(\mathbf{x}|\mathbf{d}) \propto f(\mathbf{d} | g(\mathbf{x})) f(\mathbf{x})$$



- \mathbf{x} represents model input parameters like, porosity, permeability, fault multipliers
- \mathbf{y} could represent predicted production of oil, gas, and water
- Prior pdf represents uncertainty of \mathbf{x} .
- Prior prediction pdf represents uncertainty of $\mathbf{y} = g(\mathbf{x})$.
- Data likelihood represents uncertainty of measurement \mathbf{d} .



Ensemble of cost functions



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/>

- 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](https://doi.org/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](https://doi.org/10.1016/j.cageo.2012.03.011).
- Evensen, G. *Data Assimilation: The Ensemble Kalman Filter*. Springer, 2nd edition, 2009. doi:[10.1007/978-3-642-03711-5](https://doi.org/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](https://doi.org/10.1007/s10596-018-9731-y).
- Evensen, G. Accounting for model errors in iterative ensemble smoothers. *Computat Geosci*, 23(4):761–775, 2019. doi:[10.1007/s10596-019-9819-z](https://doi.org/10.1007/s10596-019-9819-z).
- Evensen, G., P. N. Raanes, A. S. Stordal, and J. Hove. Efficient implementation of an iterative ensemble smoother for data assimilation and reservoir history matching. *Frontiers in Applied Mathematics and Statistics*, 5:47, 2019. doi:[10.3389/fams.2019.00047](https://doi.org/10.3389/fams.2019.00047).
- Evensen, G., J. Amezcuia, M. Bocquet, A. Carrassi, A. Farchi, A. Fowler, P. L. Houtekamer, C. K. Jones, R. J. de Moraes, M. Pulido, C. Sampson, and F. C. Vossepoel. An international initiative of predicting the sars-cov-2 pandemic using ensemble data assimilation. *Foundations of Data Science*, page 65, 2020. doi:[10.3934/fods.2021001](https://doi.org/10.3934/fods.2021001).
- Neal, R. M. Sampling from multimodal distributions using tempered transitions. *Statistics and Computing*, 6(4):353–366, 1996. doi:[10.1007/BF00143556](https://doi.org/10.1007/BF00143556).