Nonlinear data assimilation:

Particle filters from a Bayesian perspective

Femke C. Vossepoel

TU Delft

Based on the book available from
https://github.com/geirev/Data-Assimilation-Fundamentals.git

Who am I?

- PhD in Aerospace Engineering, TU Delft
- 7 years in Oceanography and Climate Sciences, several institutes
- 9 years in the oil and gas industry
- Since 2016: Associate Professor in Geoscience and Engineering, CEG, TU Delft
- Since 2024: Professor in Earth System Simulation, Geoscience and Engineering, CEG, TU Delft
- Mother of three children (2003, 2005, 2007)
- Marathon runner

Who are you?

Data assimilation minimises a cost function to fit a model to observations

Current data assimilation methods are computationally too expensive

Data assimilation provides information about our models and their uncertainties

Who loves linear algebra?

Agenda for my lectures this week

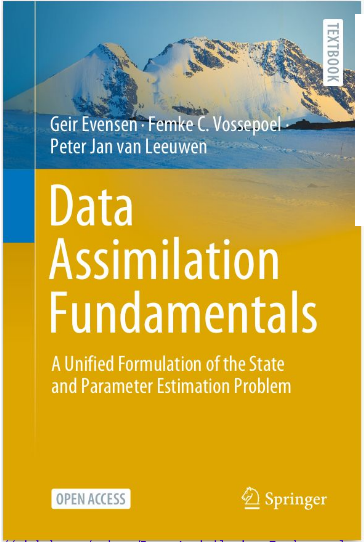
Today:

- What is data assimilation and how and why do we use it?
- Bayes' theorem as the basis of data assimilation
- Basics of the Particle Filter

Tomorrow:

- Practical particle filtering
- Examples of applications

Full details in:



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

Slides and script are/will be uploaded to:

https://github.com/femkevossepoel/DA_SummerSchool_Brasov

Notation

	1	2	3	4	5
model state	$\mathbf{x}$	$\mathbf{x}$	$\mathbf{x}$	$\mathbf{x}$	$\mathbf{g}(\mathbf{m})$
parameters			$\mathbf{p}$	$\boldsymbol{\theta}$	$\mathbf{m}$
state vector	$\mathbf{x}$	$\mathbf{x}$	$\mathbf{x}$	$\mathbf{z}$	$\mathbf{m}$
observations	$\mathbf{y}$	$\mathbf{z}$	$\mathbf{y}$	$\mathbf{d}$	$\mathbf{d}$
observation operator	$\mathbf{H}$	$\mathbf{H}$	$\mathbf{H}$	$\mathbf{H}$	$\mathbf{g}$
model error covariance		$\mathbf{Q}$		$\mathbf{C}_{zz}$	$\mathbf{C}_M$
initial/prior model covariance	$\mathbf{P}$ or $\mathbf{B}$	$\mathbf{P}$	$\mathbf{P}$	$\mathbf{C}_{xx}$	
measurement error covariance	$\mathbf{R}$	$\mathbf{R}$	$\mathbf{R}$	$\mathbf{C}_{dd}$	$\mathbf{C}_D$
system noise/model error	η	w	w	η	
measurement error	ϵ	v	v	ϵ	ϵ

1. Ide et al (1997), Raanes, Stordal, Janjic
2. Arnold Heemink
3. Martin Verlaan
4. Femke Vossepoel/ Evensen et al (2022)
5. Oliver et al (2011)

We start from Bayes' theorem

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

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

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

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

$$P(x \in [a, b]) = \int_{a}^{b} f(x)dx \tag{2}$$

- And two special cases

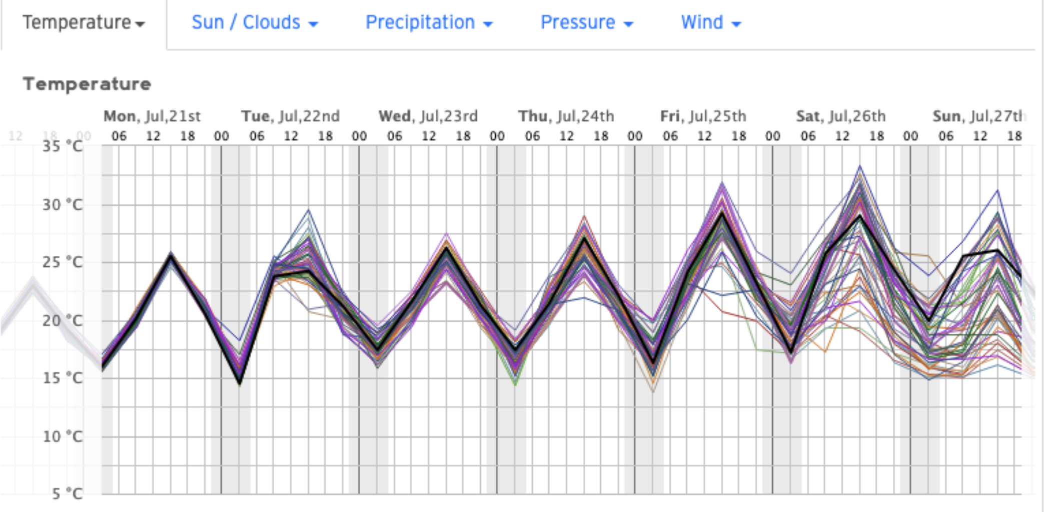
$$P(x = c) = \int_{c}^{c} f(x)dx = 0 \qquad \text{and} \qquad \int_{-\infty}^{\infty} \delta(x)dx = 1 \tag{3}$$

Also, we have

- The joint probability
$$f(x, y) = f(x)f(y|x) = f(y)f(x|y) \tag{4}$$
- Solving for $f(x|y)$ gives Bayes' theorem
$$f(x|y) = \frac{f(x)f(y|x)}{f(y)} \tag{5}$$
- 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.

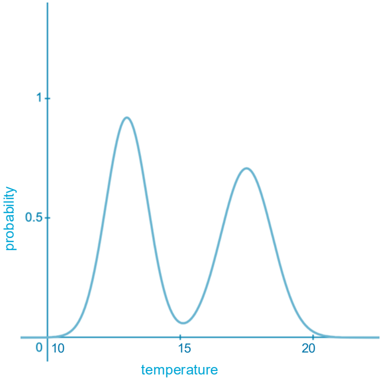
But why would we care?

Everyday use of data assimilation

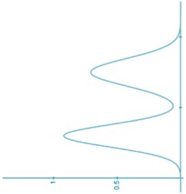


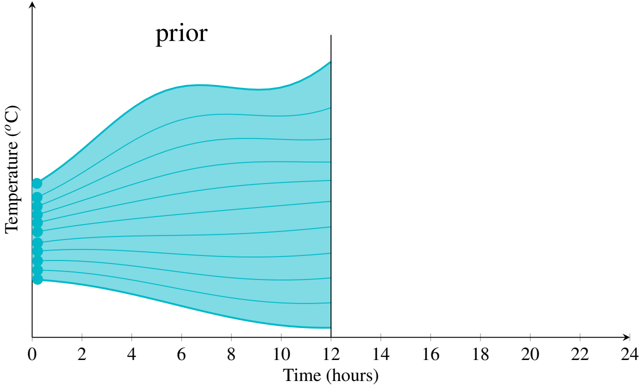
Available from: https://github.com/geirev/Data-Assimilation-Fundamentals

Consider a distribution

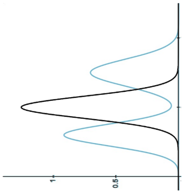


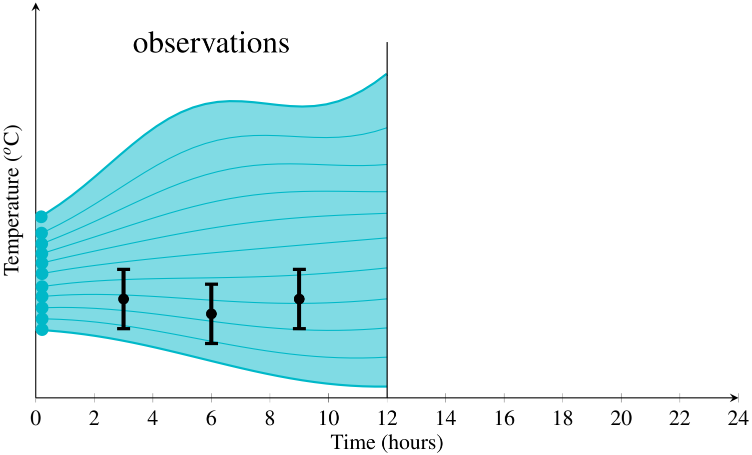
Generate a forecast - our prior



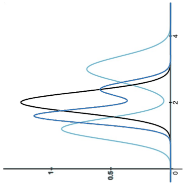


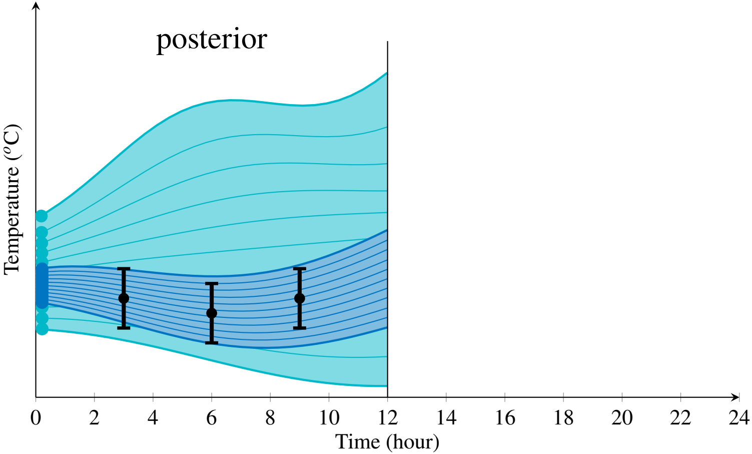
Collect observations





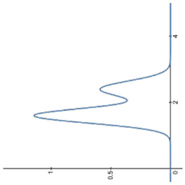
Combine forecast and observations - posterior

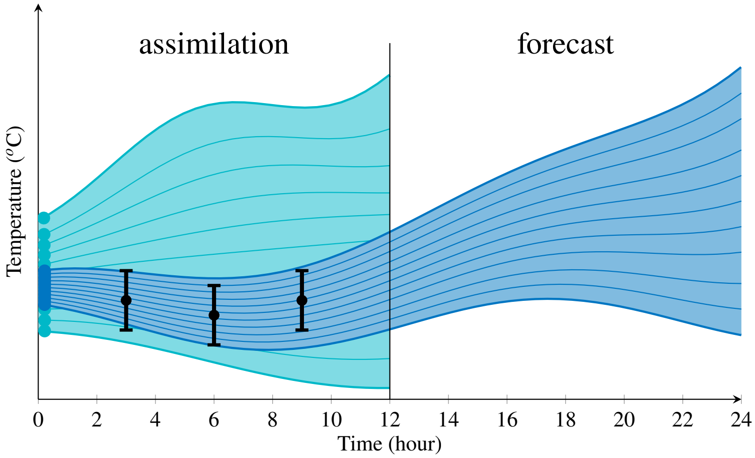




Available from: https://github.com/geirev/Data-Assimilation-Fundamentals

Produce forecast





Bayes' theorem

Given (now using again $\mathbf{z}$ and $\mathbf{d}$ for x and y):

- A state variable $\mathbf{z}$ and its prior pdf: $f(\mathbf{z})$ (note: $\mathbf{z}$ can be a state variable, a parameter, a control, or something else..)
- A vector of observations $\mathbf{d}$ and their likelihood: $f(\mathbf{d}|\mathbf{z})$
- Bayes' theorem defines the posterior pdf, $f(\mathbf{z}|\mathbf{d})$:

Bayes' theorem

$$f(\mathbf{z}|\mathbf{d}) = \frac{f(\mathbf{z})f(\mathbf{d}|\mathbf{z})}{f(\mathbf{d})} \tag{6}$$

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 parameter or state x.
- The likelihood is used after data are available to describe a plausibility of a parameter or state value x (or a vector $\mathbf{x}$.
- The likelihood does not have to integrate to one.

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

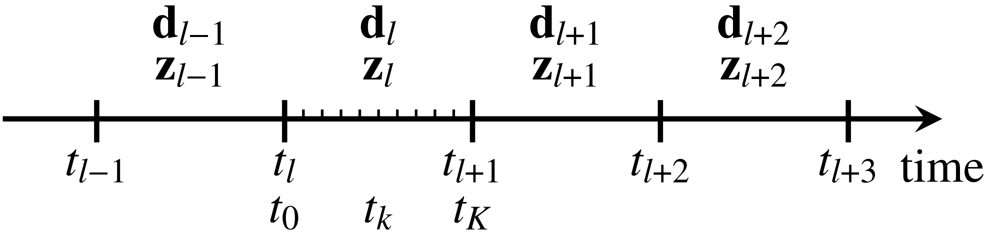
Probability is how likely are the chances of a certain data to occur if the model parameters are fixed and Likelihood is the chances of a particular model parameter explaining the given observed data.

We start from Bayes' theorem

$$f(\mathcal{Z}|\mathcal{D}) = \frac{f(\mathcal{D}|\mathcal{Z})f(\mathcal{Z})}{f(\mathcal{D})}. \tag{7}$$

- $\mathcal{Z} = (\mathbf{z}_0, \mathbf{z}_1, \ldots, \mathbf{z}_L)$ 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 data assimilation 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.

$$f(\mathbf{z}_l|\mathbf{z}_{l-1},\mathbf{z}_{l-2},\ldots,\mathbf{z}_0) = f(\mathbf{z}_l|\mathbf{z}_{l-1}), \tag{8}$$

Independent measurements

Approximation 2 (Independent measurements)
We assume that measurements are independent between different assimilation windows.

Independent measurements have uncorrelated errors

$$f(\mathcal{D}|\mathcal{Z}) = \prod_{l=1}^{L} f(\mathbf{d}_l|\mathbf{z}_l). \tag{9}$$

Bayes becomes

$$f(\mathcal{Z}|\mathcal{D}) \propto \prod_{l=1}^{L} f(\mathbf{d}_l|\mathbf{z}_l) \prod_{l=1}^{L} f(\mathbf{z}_l|\mathbf{z}_{l-1}) f(\mathbf{z}_0). \tag{10}$$

Recursive form of Bayes

$$f(\mathbf{z}_1, \mathbf{z}_0|\mathbf{d}_1) = \frac{f(\mathbf{d}_1|\mathbf{z}_1)f(\mathbf{z}_1|\mathbf{z}_0)f(\mathbf{z}_0)}{f(\mathbf{d}_1)}, \tag{11}$$

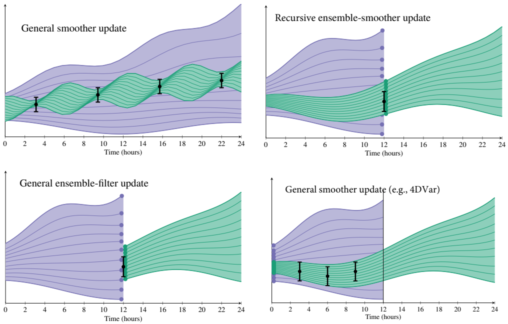
$$f(\mathbf{z}_2, \mathbf{z}_1, \mathbf{z}_0|\mathbf{d}_1, \mathbf{d}_2) = \frac{f(\mathbf{d}_2|\mathbf{z}_2)f(\mathbf{z}_2|\mathbf{z}_1)f(\mathbf{z}_1, \mathbf{z}_0|\mathbf{d}_1)}{f(\mathbf{d}_2)}, \tag{12}$$

$$\vdots \tag{13}$$

$$f(\mathcal{Z}|\mathcal{D}) = \frac{f(\mathbf{d}_L|\mathbf{z}_L)f(\mathbf{z}_L|\mathbf{z}_{L-1})f(\mathbf{z}_{L-1}, \ldots, \mathbf{z}_0|\mathbf{d}_{L-1}, \ldots, \mathbf{d}_1)}{f(\mathbf{d}_L)}. \tag{14}$$

Common classification

- Smoothers
  - (Ensemble) 4D-Var
  - (Ensemble) Randomized Maximum Likelihood (RML, EnRML)
  - Ensemble Smoothers: ES, ESMDA, IES
- Filters
  - Kalman Filter, Extended Kalman Filter
  - Ensemble Kalman Filter (EnKF)
  - Particle Filter, Particle Flow Filter



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

$$f(\mathbf{z}_1|\mathbf{d}_1) = \frac{f(\mathbf{d}_1|\mathbf{z}_1) \int f(\mathbf{z}_1|\mathbf{z}_0) f(\mathbf{z}_0)\, d\mathbf{z}_0}{f(\mathbf{d}_1)} = \frac{f(\mathbf{d}_1|\mathbf{z}_1) f(\mathbf{z}_1)}{f(\mathbf{d}_1)}, \tag{15}$$

$$f(\mathbf{z}_2|\mathbf{d}_1, \mathbf{d}_2) = \frac{f(\mathbf{d}_2|\mathbf{z}_2) \int f(\mathbf{z}_2|\mathbf{z}_1) f(\mathbf{z}_1|\mathbf{d}_1)\, d\mathbf{z}_1}{f(\mathbf{d}_2)} = \frac{f(\mathbf{d}_2|\mathbf{z}_2) f(\mathbf{z}_2|\mathbf{d}_1)}{f(\mathbf{d}_2)}, \tag{16}$$

$$\vdots$$

$$f(\mathbf{z}_L|\mathcal{D}) = \frac{f(\mathbf{d}_L|\mathbf{z}_L) \int f(\mathbf{z}_L|\mathbf{z}_{L-1}) f(\mathbf{z}_{L-1}|\mathbf{d}_{L-1}, \ldots, \mathbf{d}_1)\, d\mathbf{z}_{L-1}}{f(\mathbf{d}_L)} \tag{17}$$

$$= \frac{f(\mathbf{d}_L|\mathbf{z}_L) f(\mathbf{z}_L|\mathbf{d}_{L-1})}{f(\mathbf{d}_L)}. \tag{18}$$

Or just Bayes' for the assimilation window

$$f(\mathbf{z}|\mathbf{d}) = \frac{f(\mathbf{d}|\mathbf{z})\, f(\mathbf{z})}{f(\mathbf{d})}, \tag{19}$$

Discrete model with uncertain inputs

$$\mathbf{x}_k = \mathbf{m}(\mathbf{x}_{k-1}, \boldsymbol{\theta}, \mathbf{u}_k, \mathbf{q}_k). \tag{20}$$

- $\mathbf{x}_k$ 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.

Error propagation by Fokker-Planck equation

Stochastic model

$$d\mathbf{x} = \mathbf{m}(\mathbf{x})\,dt + d\mathbf{q}. \tag{21}$$

Fokker-Planck is an advection-diffusion equation in the state-space

$$\frac{\partial f(\mathbf{x})}{\partial t} + \sum_i \frac{\partial \big(m_i(\mathbf{x}) f(\mathbf{x})\big)}{\partial x_i} = \frac{1}{2}\mathbf{C}_{qq} \sum_{i,j} \frac{\partial^2 f(\mathbf{x})}{\partial x_i\,\partial x_j}. \tag{22}$$

Error propagation by covariance evolution

Comparing the evolution of the true model state with that of our estimated model state

$$\mathbf{x}^{\mathrm{t}}_{k+1} = \mathbf{m}(\mathbf{x}^{\mathrm{t}}_k) + \mathbf{q}_k \approx \mathbf{m}(\mathbf{x}_k) + \mathbf{M}_k(\mathbf{x}^{\mathrm{t}}_k - \mathbf{x}_k) + \mathbf{q}_k, \tag{23}$$

$$\mathbf{x}_{k+1} = \mathbf{m}(\mathbf{x}_k), \tag{24}$$

Subtract Eq. (24) from Eq. (23), square the result, and take the expectation,

$$\mathbf{C}_{xx,k+1} \approx \mathbf{M}_k \mathbf{C}_{xx,k} \mathbf{M}_k^{\mathrm{T}} + \mathbf{C}_{qq,k}. \tag{25}$$

- $\mathbf{M}_k$ is the model's tangent-linear operator evaluated at $\mathbf{x}_k$.
- $\mathbf{C}_{qq}$ is the model error covariance matrix.

Error propagation by ensemble predictions

- Represent uncertainty by an ensemble of samples

$$\mathbf{x}_{j,0} \sim f(\mathbf{x}) \quad \text{and} \quad \mathbf{q}_{j,k} \sim f(\mathbf{x}_{k+1}|\mathbf{x}_k) \tag{26}$$

- Nonlinear propagation of uncertainty by ensemble integrations using the dynamical model.

$$\mathbf{x}_{j,k+1} = \mathbf{m}(\mathbf{x}_{j,k}, \mathbf{q}_{j,k}). \tag{27}$$

We can then compute statistics like mean and covariance, e.g.,

$$\mathbb{E}[\mathbf{x}] \approx \overline{\mathbf{x}} = \frac{1}{N}\sum_{j=1}^{N}\mathbf{x}_j \tag{28}$$

$$\mathbf{C}_{xx} \approx \overline{(\mathbf{x}-\overline{\mathbf{x}})(\mathbf{x}-\overline{\mathbf{x}})^{\mathrm{T}}} = \frac{1}{N-1}\sum_{j=1}^{N}(\mathbf{x}_j-\overline{\mathbf{x}})(\mathbf{x}_j-\overline{\mathbf{x}})^{\mathrm{T}} \tag{29}$$

General smoother formulation

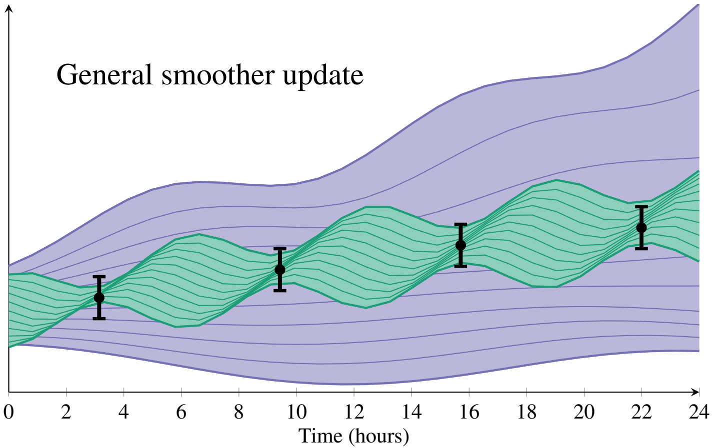
- Solve for model solution over an assimilation window $\mathbf{x}$.
- Condition on measurements distributed over the assimilation window.

Predicted measurements $\mathbf{y}$

$$\mathbf{y} = \mathbf{g}(\mathbf{z}) = \mathbf{h}(\mathbf{x}) = \mathbf{h}\big(\mathbf{m}(\mathbf{x}_0, \mathbf{q})\big) \tag{30}$$

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

General smoother formulation



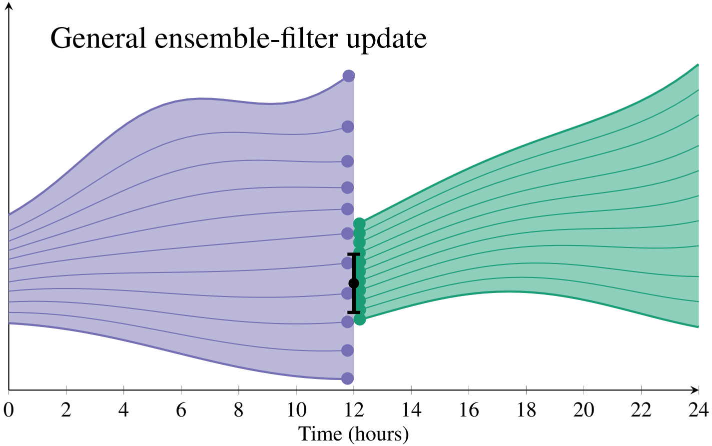
General filter formulation

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

$$\mathbf{y} = \mathbf{g}(\mathbf{z}) = \mathbf{h}(\mathbf{x}_K). \tag{31}$$

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

General filter formulation



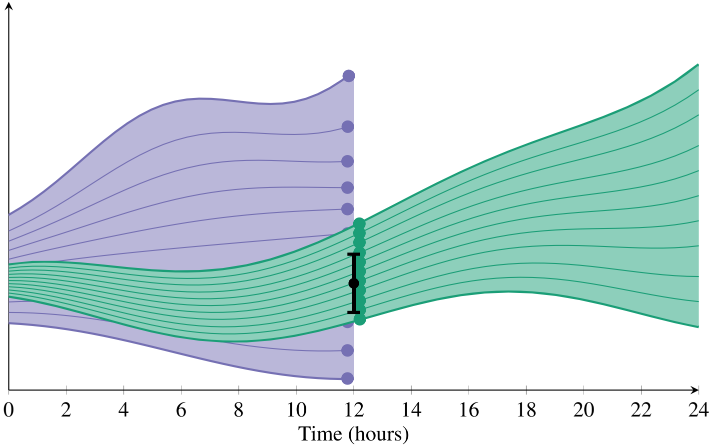
Recursive smoother formulation

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

$$\mathbf{y} = \mathbf{g}(\mathbf{z}) = \mathbf{h}(\mathbf{x}_K), \tag{32}$$

- Ensemble Kalman Smoother (EnKS)

Recursive smoother formulation



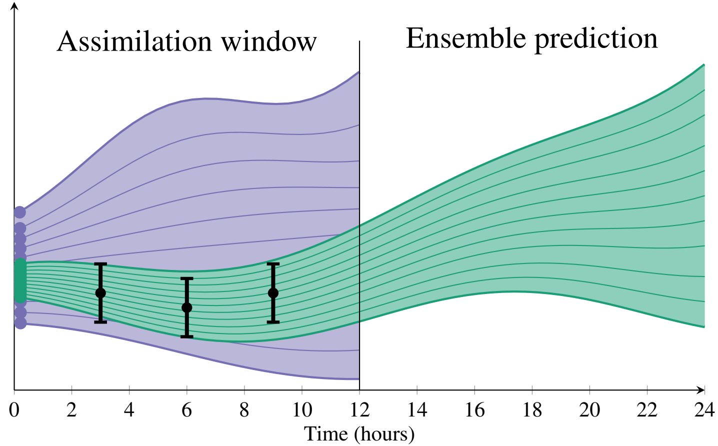
Smoother for parameter estimation

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

$$\mathbf{y} = \mathbf{g}(\mathbf{z}) = \mathbf{h}\big(\mathbf{m}(\boldsymbol{\theta})\big). \tag{33}$$

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

Smoother for initial conditions or input parameters (perfect models)



Deriving the marginal posterior pdf

Nonlinear "perfect" model and measurements

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

Bayesian formulation

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

Likelihood contains model operator

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

Marginal pdf

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

Bayes' theorem related to the predicted measurements

We introduce nonlinearity through the likelihood

$$f(\mathbf{z}|\mathbf{d}) = \frac{f\big(\mathbf{d}|\mathbf{g}(\mathbf{z})\big)\, f(\mathbf{z})}{f(\mathbf{d})}. \tag{34}$$

Summary

- Bayes theorem forms the basis of most data assimilation schemes
- We discriminate smoothers and filters, most methods use ensembles for the estimation of the posterior distribution of a state or parameter
- In data assimilation methods, we typically assume the following:
  - The model is a 1st-order Markov process
  - The measurements are independent
  - In case of a filter: we do not project the measurement's information backward in time
- The state vector can contain the state variables, the parameters, model errors, and/or controls
- Nonlinearity in the model 'enters' the data assimilation via the model operator $\mathbf{g}(\mathbf{z})$

Questions?