Nonlinear data assimilation:

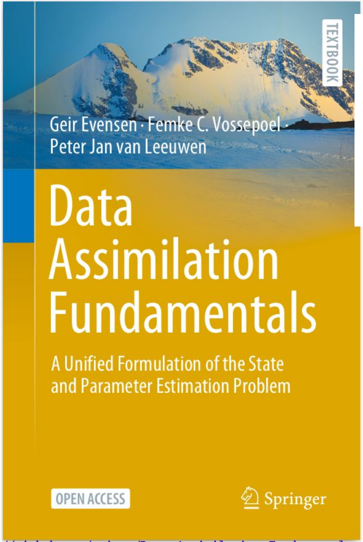
Particle filters from a Bayesian perspective

Femke C. Vossepoel, based on the book of Geir Evensen, Femke C. Vossepoel and Peter Jan van Leeuwen

Book available from
https://github.com/geirev/Data-Assimilation-Fundamentals.git

NORCE TUDelft COLORADO STATE UNIVERSITY

Full details in:

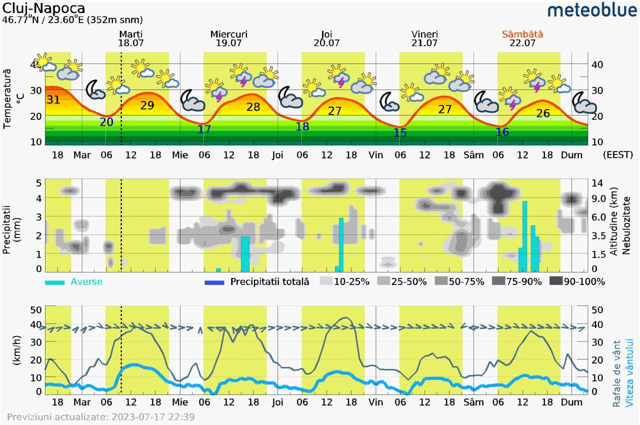


Available from: https://github.com/geirev/Data-Assimilation-Fundamentals

Slides and script are/will be uploaded to:

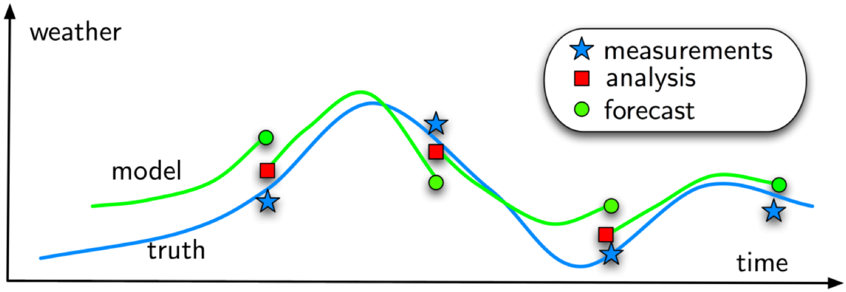
https://github.com/femkevossepoel/DA_SummerSchool_Cluj

Let's start with the weather...



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

Data assimilation concept (sequential)



Univariate example

We assume that state variables of state $\mathcal{Z}$ (*e.g.*, temperature) and imperfect observations thereof $\mathcal{D}$ are normally distributed:

$$f(z) = \mathcal{N}(\mu, \tau^2) \tag{1}$$

$$f(\mathbf{d} \mid z) = \mathcal{N}(z, \sigma^2) \tag{2}$$

Note:

- $\mathbf{d}$ is a vector
- In this case, we are considering the state as a scalar, z
- We assume to have n observations
- τ is the standard deviation of z
- σ is the standard deviation of $\mathbf{d}$

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

Univariate example

$$f(\mathbf{d}|z) = \prod_{i=1}^{n} (1/\sqrt{2\pi\sigma^2}) exp\{-0.5(d_i - z)^2/\sigma^2\} \tag{3}$$

$$\propto exp\{-0.5 \sum_{i=1}^{n} (d_i - z)^2/\sigma^2\}. \tag{4}$$

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

Univariate example

Using

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

and

$$f(z) = \mathcal{N}(\mu, \tau^2) \tag{5}$$

$$f(\mathbf{d} \mid z) = \mathcal{N}(z, \sigma^2) \tag{6}$$

gives

$$f(z|\mathbf{d}) \propto ezp\{-0.5[\sum_{i=1}^{n}(d_i - z)^2/\sigma^2 + (z-\mu)^2/\tau^2]\} \tag{7}$$

$$\propto exp\{-0.5[z^2(n/\sigma^2 + 1/\tau^2) - 2\Big(\sum d_i/\sigma^2 + \mu^2/\tau^2\Big)z]\} \tag{8}$$

$$\tag{9}$$

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

Univariate example

This gives for state $\mathcal{Z}$ ($\sim$ meaning "is distributed as")

$$\mathcal{Z}|\mathbf{d} \sim \mathcal{N}((n/\sigma^2 + 1/\tau^2)^{-1} + (\sum d_i/\sigma^2 + \mu/\tau^2), \\ (n/\sigma^2 + 1/\tau^2)^{-1}) \tag{10}$$

We can write the posterior mean as

$$E(\mathcal{Z}|\mathbf{d}) = \frac{\sigma^2\tau^2}{\sigma^2 + n\tau^2}(n\bar{d}/\sigma^2 + \mu/\tau^2) \tag{11}$$

$$= w_d\bar{d} + w_\mu\mu. \tag{12}$$

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

Univariate example

This can then also be written as

$$E(\mathcal{Z}|\mathbf{d}) = \mu + (\frac{n\tau^2}{\sigma^2 + n\tau^2})(\bar{\mathbf{d}} - \mu) \tag{13}$$

$$= \mu + K(1 - K)\tau^2, \tag{14}$$

the posterior as an update of the prior, and where K is referred to as the gain.

Exercise

Consider the example:

- state $\mathcal{Z} \sim \mathcal{N}(20, 3)$
- observations $\mathcal{D}_i | z \sim \mathcal{N}(z, 1)$
- two observations $\mathbf{d} = [19, 23]^{\mathrm{T}}$

Questions:

- what is the posterior mean?
- what is the posterior variance?

Exercise

Same questions for

- observations $\mathcal{D}_i|z \sim \mathcal{N}(z, 10)$

- what is the posterior mean?
- what is the posterior variance?

Univariate example

- observations $\mathcal{D}_i|z \sim \mathcal{N}(z, 1)$
- observations $\mathcal{D}_i|z \sim \mathcal{N}(z, 10)$

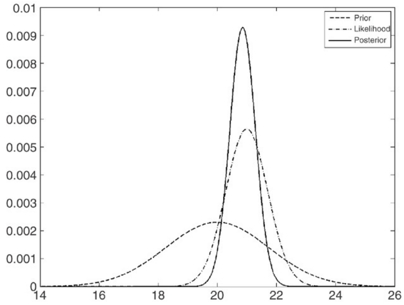


Fig. 1. Posterior distribution with normal prior and normal likelihood; relatively precise data.

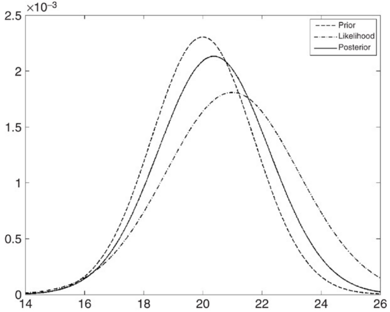


Fig. 2. Posterior distribution with normal prior and normal likelihood; relatively uncertain data.

Example from Wikle & Berliner (Physica D, 2007)

Bayes' theorem

Given:

- A state variable $\mathbf{z}$ and its prior pdf: $f(\mathbf{z})$
- 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

$$f(\mathbf{z}|\mathbf{d}) = \frac{f(\mathbf{z})f(\mathbf{d}|\mathbf{z})}{f(\mathbf{d})} \tag{15}$$

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

$$P(x) = \int_{-\infty}^{\infty} f(x)dx = 1 \tag{16}$$

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

$$P(x \in [a, b]) = \int_{a}^{b} f(x)dx \tag{17}$$

- And two special cases

$$P(x = c) = \int_{c}^{c} f(x)dx = 0 \qquad \text{and} \qquad \int_{-\infty}^{\infty} \delta(x)dx = 1 \tag{18}$$

Also, we have

- The joint probability
$$f(x,y) = f(x)f(y|x) = f(y)f(x|y) \tag{19}$$
- Solving for $f(x|y)$ gives Bayes' theorem
$$f(x|y) = \frac{f(x)f(y|x)}{f(y)} \tag{20}$$
- 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 (now using again $\mathbf{z}$ and $\mathbf{d}$ for x and y):

- A state variable $\mathbf{z}$ and its prior pdf: $f(\mathbf{z})$
- 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

$$f(\mathbf{z}|\mathbf{d}) = \frac{f(\mathbf{z})f(\mathbf{d}|\mathbf{z})}{f(\mathbf{d})} \tag{21}$$

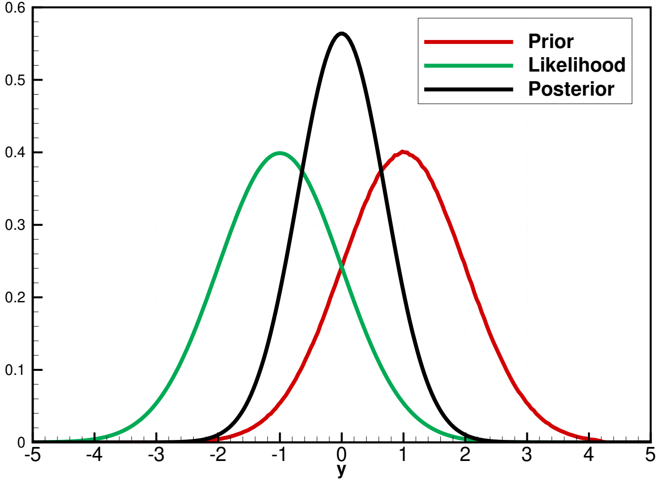
What is the likelihood function: $f(d|x)$

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

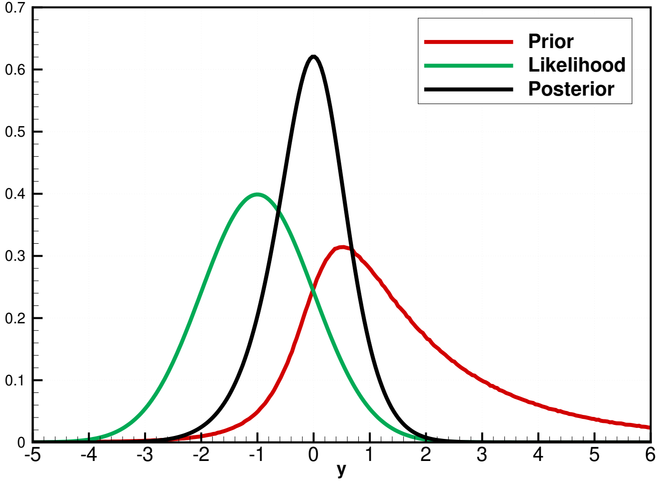
Likelihood is the plausability of a particular distribution explaining the given data. The higher the likelihood of a distribution, the more likely it is to explain the observed data.

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

Example of using Bayes' theorem



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.

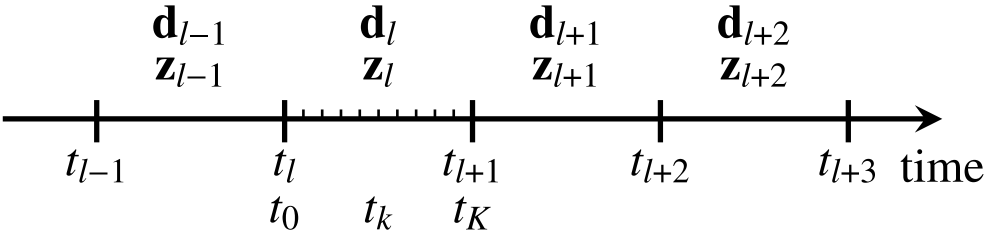
One of the strengths of Bayes' formula is that it does not try to solve the ill-defined problem of "inverting observations" but instead updates prior knowledge.

We start from Bayes' theorem

$$f(\mathcal{Z}|\mathcal{D}) = \frac{f(\mathcal{D}|\mathcal{Z})f(\mathcal{Z})}{f(\mathcal{D})}. \tag{22}$$

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

$$f(\mathbf{z}_l|\mathbf{z}_{l-1}, \mathbf{z}_{l-2}, \ldots, \mathbf{z}_0) = f(\mathbf{z}_l|\mathbf{z}_{l-1}), \tag{23}$$

Independent measurements

Approximation 2 (Independent measurements)

We assume that measurements are independent between different assimilation windows.

Independent measurements have uncorrelated errors

$$f(\mathcal{D}|\mathcal{Z}) = \prod_{l=1}^{L} f(\mathbf{d}_l|\mathbf{z}_l). \tag{24}$$

Bayes becomes

$$f(\mathcal{Z}|\mathcal{D}) \propto \prod_{l=1}^{L} f(\mathbf{d}_l|\mathbf{z}_l) \prod_{l=1}^{L} f(\mathbf{z}_l|\mathbf{z}_{l-1}) f(\mathbf{z}_0). \tag{25}$$

Recursive form of Bayes

$$f(\mathbf{z}_1, \mathbf{z}_0 | \mathbf{d}_1) = \frac{f(\mathbf{d}_1 | \mathbf{z}_1) f(\mathbf{z}_1 | \mathbf{z}_0) f(\mathbf{z}_0)}{f(\mathbf{d}_1)}, \tag{26}$$

$$f(\mathbf{z}_2, \mathbf{z}_1, \mathbf{z}_0 | \mathbf{d}_1, \mathbf{d}_2) = \frac{f(\mathbf{d}_2 | \mathbf{z}_2) f(\mathbf{z}_2 | \mathbf{z}_1) f(\mathbf{z}_1, \mathbf{z}_0 | \mathbf{d}_1)}{f(\mathbf{d}_2)}, \tag{27}$$

$$\vdots \tag{28}$$

$$f(\mathcal{Z} | \mathcal{D}) = \frac{f(\mathbf{d}_L | \mathbf{z}_L) f(\mathbf{z}_L | \mathbf{z}_{L-1}) f(\mathbf{z}_{L-1}, \dots, \mathbf{z}_0 | \mathbf{d}_{L-1}, \dots, \mathbf{d}_1)}{f(\mathbf{d}_L)}. \tag{29}$$

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

$$f(\mathbf{z}_1|\mathbf{d}_1) = \frac{f(\mathbf{d}_1|\mathbf{z}_1)\int f(\mathbf{z}_1|\mathbf{z}_0) f(\mathbf{z}_0)\,d\mathbf{z}_0}{f(\mathbf{d}_1)} = \frac{f(\mathbf{d}_1|\mathbf{z}_1) f(\mathbf{z}_1)}{f(\mathbf{d}_1)}, \tag{30}$$

$$f(\mathbf{z}_2|\mathbf{d}_1,\mathbf{d}_2) = \frac{f(\mathbf{d}_2|\mathbf{z}_2)\int f(\mathbf{z}_2|\mathbf{z}_1) f(\mathbf{z}_1|\mathbf{d}_1)\,d\mathbf{z}_1}{f(\mathbf{d}_2)} = \frac{f(\mathbf{d}_2|\mathbf{z}_2) f(\mathbf{z}_2|\mathbf{d}_1)}{f(\mathbf{d}_2)}, \tag{31}$$

$$\vdots$$

$$f(\mathbf{z}_L|\mathcal{D}) = \frac{f(\mathbf{d}_L|\mathbf{z}_L)\int f(\mathbf{z}_L|\mathbf{z}_{L-1}) f(\mathbf{z}_{L-1}|\mathbf{d}_{L-1},\ldots,\mathbf{d}_1)\,d\mathbf{z}_{L-1}}{f(\mathbf{d}_L)} \tag{32}$$

$$= \frac{f(\mathbf{d}_L|\mathbf{z}_L) f(\mathbf{z}_L|\mathbf{d}_{L-1})}{f(\mathbf{d}_L)}. \tag{33}$$

Or just Bayes' for the assimilation window

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

Discrete model with uncertain inputs

$$\mathbf{x}_k = \mathbf{m}(\mathbf{x}_{k-1}, \boldsymbol{\theta}, \mathbf{u}_k, \mathbf{q}_k). \tag{35}$$

- $\mathbf{x}_k$ 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.

Error propagation by Fokker-Planck equation

Stochastic model

$$d\mathbf{x} = \mathbf{m}(\mathbf{x})\, dt + d\mathbf{q}. \tag{36}$$

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

$$\frac{\partial f(\mathbf{x})}{\partial t} + \sum_i \frac{\partial \big(m_i(\mathbf{x}) f(\mathbf{x})\big)}{\partial x_i} = \frac{1}{2}\mathbf{C}_{qq} \sum_{i,j} \frac{\partial^2 f(\mathbf{x})}{\partial x_i\, \partial x_j}. \tag{37}$$

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

Error propagation by covariance evolution

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

$$\mathbf{x}^{\mathrm{t}}_{k+1} = \mathbf{m}(\mathbf{x}^{\mathrm{t}}_{k}) + \mathbf{q}_k \approx \mathbf{m}(\mathbf{x}_k) + \mathbf{M}_k(\mathbf{x}^{\mathrm{t}}_k - \mathbf{x}_k) + \mathbf{q}_k, \tag{38}$$

$$\mathbf{x}_{k+1} = \mathbf{m}(\mathbf{x}_k), \tag{39}$$

Subtract Eq. (39) from Eq. (38), square the result, and take the expectation,

$$\mathbf{C}_{xx,k+1} \approx \mathbf{M}_k \mathbf{C}_{xx,k} \mathbf{M}_k^{\mathrm{T}} + \mathbf{C}_{qq,k}. \tag{40}$$

- $\mathbf{M}_k$ 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

$$\mathbf{x}_{j,0} \sim f(\mathbf{x}) \quad \text{and} \quad \mathbf{q}_{j,k} \sim f(\mathbf{x}_{k+1}|\mathbf{x}_k) \tag{41}$$

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

$$\mathbf{x}_{j,k+1} = \mathbf{m}(\mathbf{x}_{j,k}, \mathbf{q}_{j,k}). \tag{42}$$

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

$$\mathbb{E}[\mathbf{x}] \approx \overline{\mathbf{x}} = \frac{1}{N}\sum_{j=1}^{N} \mathbf{x}_j \tag{43}$$

$$\mathbf{C}_{xx} \approx \overline{(\mathbf{x}-\overline{\mathbf{x}})(\mathbf{x}-\overline{\mathbf{x}})^{\mathrm{T}}} = \frac{1}{N-1}\sum_{j=1}^{N}(\mathbf{x}_j-\overline{\mathbf{x}})(\mathbf{x}_j-\overline{\mathbf{x}})^{\mathrm{T}} \tag{44}$$

General smoother formulation

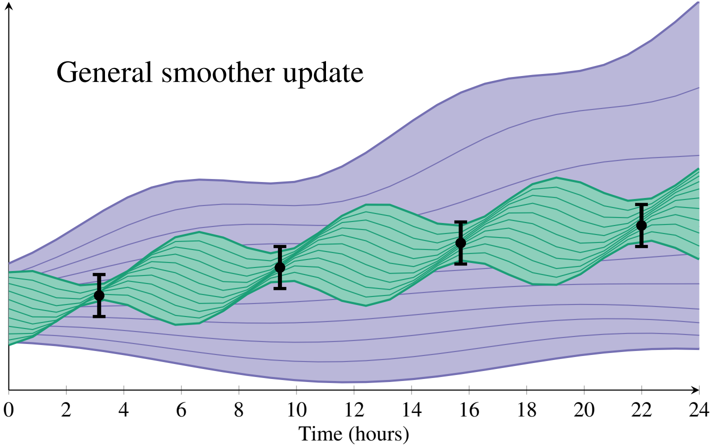
- Solve for model solution over an assimilation window $\mathbf{x}$.
- Condition on measurements distributed over the assimilation window.

Predicted measurements $\mathbf{y}$

$$\mathbf{y} = \mathbf{g}(\mathbf{z}) = \mathbf{h}(\mathbf{x}) = \mathbf{h}\big(\mathbf{m}(\mathbf{x}_0, \mathbf{q})\big) \tag{45}$$

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

General smoother formulation



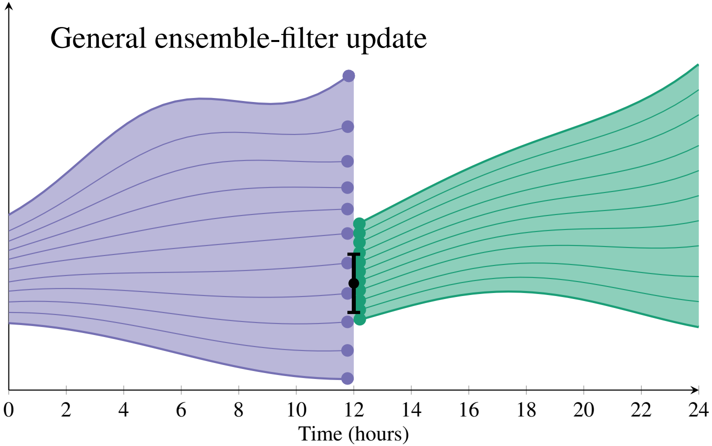
General filter formulation

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

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

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

General filter formulation



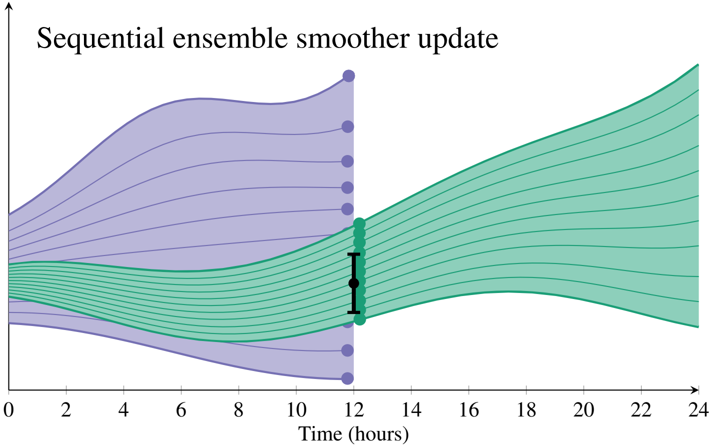
Recursive smoother formulation

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

$$\mathbf{y} = \mathbf{g}(\mathbf{z}) = \mathbf{h}(\mathbf{x}_K), \tag{47}$$

- Ensemble Kalman Smoother (EnKS)

Recursive smoother formulation



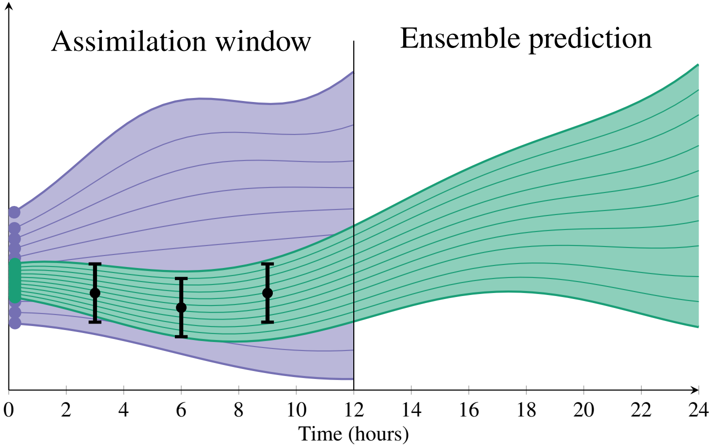
Parameter estimation

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

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

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

Smoother for perfect models



Deriving the marginal posterior pdf

Nonlinear “perfect” model and measurements

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

Bayesian formulation

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

Model pdf

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

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

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

BREAK