# Lecture 1:
# Principles of geophysical data assimilation


Marc Bocquet

With help from Alberto Carrassi, Alban Farchi

CEREA, École des Ponts ParisTech and EdF R&D, Île-De-France, France
Institut Pierre-Simon Laplace

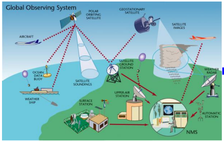
(marc.bocquet@enpc.fr)

# Synopsis of the course

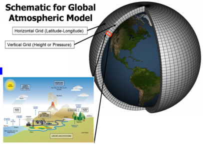
- Tuesday, June 6 16:00-17:30
  Lecture 1: Principles of geophysical data assimilation. The Bayesian standpoint. Classical methods of data assimilation: 3D-Var, the Kalman filter, 4D-Var, the ensemble Kalman filter.
- Wednesday, June 7, 09:00-10:30
  Lecture 2: Combining data assimilation and machine learning: Machine learning and the geosciences. Surrogate modelling, offline and online. Illustrations in the climate sciences.
- Thursday, June 8, 09:00-11:00
  Lecture 3: Training session: Learning dynamics and surrogate modelling.

# Outline

# Data assimilation (DA) in the geosciences




Data assimilation
best combines
observations and models




Expanded from numerical weather prediction to the climate science/geosciences:

- Oceanography
- Atmospheric chemistry
- Climate prediction and assessment
- Glaciology, sea-ice.
- Hydrology and hydraulics
- Geology
- Space weather
- and many other fields

# Data assimilation: an inference problem

- Inference is the process of taking a decision based on limited information.

- Information comes from
  - an approximate knowledge about the laws (if any) governing the time evolution of the dynamical system
  - imperfect (partial, noisy, indirect) observations of this system

- Sequential inference is the problem of updating our knowledge about the system each time a new batch of observations becomes available.

# First ingredient: the dynamical model

- We will assume that a model of the natural process of interest is available as a discrete stochastic dynamical system,

$$\mathbf{x}_k = \mathcal{M}_{k:k-1}(\mathbf{x}_{k-1}, \boldsymbol{\lambda}) + \boldsymbol{\eta}_k.$$

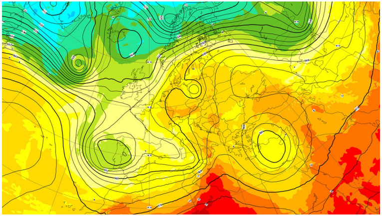
- $\mathbf{x}_k \in \mathbb{R}^{N_x}$ and $\boldsymbol{\lambda} \in \mathbb{R}^{N_\mathrm{p}}$ are the model state and parameter vectors respectively.

- $\mathcal{M}_{k:k-1} : \mathbb{R}^{N_x} \to \mathbb{R}^{N_x}$ is usually a nonlinear, possibly chaotic, map from $t_{k-1}$ to $t_k$.

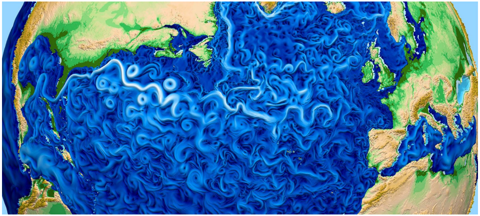
- $\boldsymbol{\eta}_k \in \mathbb{R}^{N_x}$ is the model error, represented as a stochastic additive term (more general representations are possible).

# First ingredient: the dynamical model

- In the geosciences:
  - The state space dimension is huge (up to $10^9$ for operational systems, up to $10^7$ for research systems). A big data problem with costly models to integrate.
  - Numerical models (i.e. implementation of $\mathcal{M}$) are often computationally very costly.
  - The unstable dynamics of chaotic geofluids has implicit consequences on the design of DA algorithms: One key reason why we use sequential inference.



ECMWF IFS: Geopotential at 500hPa and temperature at 850hPa



E3SM Earth system model

# Second ingredient: the observations

- Noisy observations, $\mathbf{y}_k \in \mathbb{R}^{N_y}$, are available at discrete times and are related to the model state vector through

$$\mathbf{y}_k = \mathcal{H}_k(\mathbf{x}_k) + \boldsymbol{\epsilon}_k,$$

with $\mathcal{H} : \mathbb{R}^{N_x} \rightarrow \mathbb{R}^{N_y}$ being the (generally nonlinear) observation operator mapping from the model to the observational space.

- The observation error, $\boldsymbol{\epsilon}_k$, is represented as a stochastic term. It accounts for the instrumental error, for deficiencies in the formulation of $\mathcal{H}$, and for the representation error.

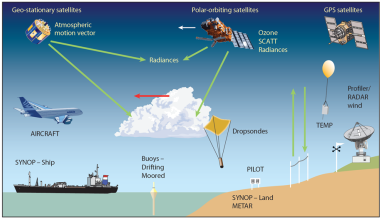
- The representation error arises from the presence of unresolved scales and represents their effect on the resolved scales – it is ubiquitous in physical science and inherent to the discretisation procedure [Janjić et al. 2018].

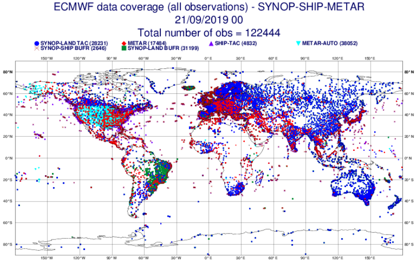
- We assume that the observation dimension is constant, so that $N_y(k) \equiv N_y$ (the generalisation is simple). Remark: often $N_y \ll N_x$, i.e. the amount of available data is insufficient to fully describe the system.

# Second ingredient: the observations

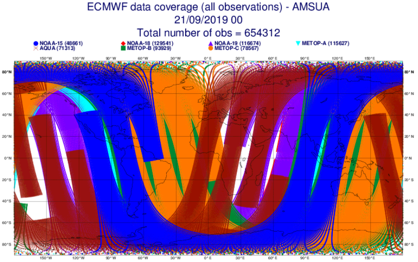
▶ In the geosciences: The observation space dimension is huge (up to $10^7$ for operational systems, up to $10^6$ for research systems). A big data problem.

▶ The Earth observations gather measurements of many sources: conventional and space-borne.



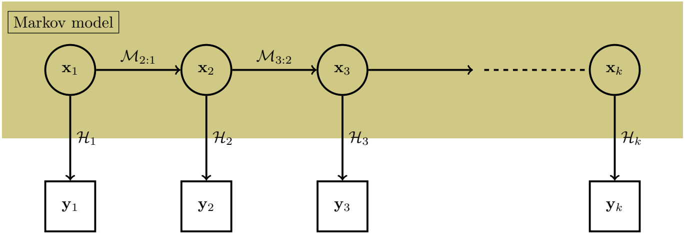

Conventional observations coverage used at ECMWF




AMSUA observations used at ECMWF

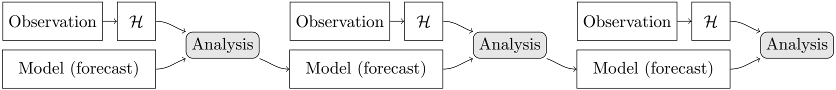
# Hidden Markov model

- Considering the states and observations as random variables, the dynamical model, together with the observation model, define a Hidden Markov model:




- This is an inverse problem: Estimate the state $\mathbf{x}$ given the observation $\mathbf{y}$.

- Data assimilation for forecasting chaotic geofluids: sequential schemes



# Bayesian inference

- When making inference we have to decide how much we trust the uncertain information. We need to quantify the uncertainty.

- Given the random nature of the problem,

  uncertainty quantification is achieved using probabilities.

- The Bayesian approach offers a natural mathematical framework to understand and formalise this problem.

- In particular, the goal of Bayesian inference is to estimate the uncertainty in $\mathbf{x}$ given $\mathbf{y}$, i.e compute the conditional probability density function (pdf) $p(\mathbf{x}|\mathbf{y})$.

# Bayesian inference

- Bayes/Laplace's rule:

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})}$$

with $p(\mathbf{y}|\mathbf{x})$ the likelihood of the observations, $p(\mathbf{x})$ the prior/background on the system's state, and $p(\mathbf{y})$ the evidence. The evidence is a normalisation factor that does not depend on $\mathbf{x}$:

$$p(\mathbf{y}) = \int \mathrm{d}\mathbf{x}\, p(\mathbf{y}|\mathbf{x})p(\mathbf{x})\,.$$

- This is a probabilistic approach. It quantifies the uncertainty/the information. It does not provide a deterministic estimator. This would require to make a choice on top of Bayes'rule.

- The Bayesian approach is very satisfactorily [Jaynes 2003]. Most DA methods can be derived or comply with Bayes'rule.

# Sequential Bayesian estimation

- Recall our HMM given by the dynamical model and observation model:

$$\mathbf{x}_k = \mathcal{M}_{k:k-1}(\mathbf{x}_{k-1}, \boldsymbol{\lambda}) + \boldsymbol{\eta}_k, \qquad \mathbf{y}_k = \mathcal{H}_k(\mathbf{x}_k) + \boldsymbol{\epsilon}_k.$$

- The model and the observational errors, $\{\boldsymbol{\eta}_k\}_{k=1,\dots,K}$, $\{\boldsymbol{\epsilon}_k\}_{k=0,\dots,K}$ are assumed to be uncorrelated in time, mutually independent, and distributed according to the pdfs $p_{\boldsymbol{\eta}}$ and $p_{\boldsymbol{\epsilon}}$.

- Let us define the sequences of system states and observations within the interval $[t_0, \cdots, t_K]$ as $\mathbf{x}_{K:0} = \{\mathbf{x}_K, \mathbf{x}_{K-1}, \cdots, \mathbf{x}_0\}$ and $\mathbf{y}_{K:0} = \{\mathbf{y}_K, \mathbf{y}_{K-1}, \cdots, \mathbf{y}_0\}$ respectively.

We wish to estimate the posterior $p(\mathbf{x}_{K:0}|\mathbf{y}_{K:0})$ for increasing $K$. Using Bayes'rule:

$$p(\mathbf{x}_{K:0}|\mathbf{y}_{K:0}) \propto p(\mathbf{y}_{K:0}|\mathbf{x}_{K:0})p(\mathbf{x}_{K:0}).$$

# Sequential Bayesian estimation

▶ Since the observational errors are assumed to be uncorrelated in time we have $p(\mathbf{y}_k|\mathbf{x}_{K:0}) = p(\mathbf{y}_k|\mathbf{x}_k)$ and we can split the global likelihood:

$$p(\mathbf{y}_{K:0}|\mathbf{x}_{K:0}) = \prod_{k=0}^{K} p(\mathbf{y}_k|\mathbf{x}_k) = \prod_{k=0}^{K} p_{\boldsymbol{\epsilon}}\left(\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k)\right).$$

▶ Also, in virtue of the Markov property we have $p(\mathbf{x}_{k+1}|\mathbf{x}_{k:0}) = p(\mathbf{x}_{k+1}|\mathbf{x}_k)$ (prediction at $t_{k+1}$ only depends on the state at $t_k$), and we can split the global prior as

$$p(\mathbf{x}_{K:0}) = p(\mathbf{x}_0) \prod_{k=1}^{K} p(\mathbf{x}_k|\mathbf{x}_{k-1}) = p(\mathbf{x}_1) \prod_{k=0}^{K} p_{\boldsymbol{\eta}}\left(\mathbf{x}_k - \mathcal{M}_{k:k-1}(\mathbf{x}_{k-1})\right).$$

# Sequential Bayesian estimation

- By combining these equations using Bayes'rule we get the posterior distribution

$$p(\mathbf{x}_{K:0}|\mathbf{y}_{K:0}) \propto p(\mathbf{x}_0)p(\mathbf{y}_0|\mathbf{x}_0)\prod_{k=1}^{K} p(\mathbf{y}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{x}_{k-1})$$

$$\propto p(\mathbf{x}_0)p_{\boldsymbol{\epsilon}}\left(\mathbf{y}_0 - \mathcal{H}_0(\mathbf{x}_0)\right)\prod_{k=1}^{K} p_{\boldsymbol{\epsilon}}\left(\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k)\right)p_{\boldsymbol{\eta}}\left(\mathbf{x}_k - \mathcal{M}_{k:k-1}(\mathbf{x}_{k-1})\right).$$

- This equation is of central importance: it states that a new update can be obtained as soon as new observations are available.

- Sequential inference can be obtained by recursively estimating $p(\mathbf{y}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{x}_{k-1})$.

- The Bayesian formalism has all the qualities we wish for except that it does not lend to a closed form, analytically tractable solution.

# Sequential Bayesian estimation

- Thanks to the main result on the HMM:

$$p(\mathbf{x}_{K:0}|\mathbf{y}_{K:0}) \propto p(\mathbf{x}_0)p(\mathbf{y}_0|\mathbf{x}_0)\prod_{k=1}^{K} p(\mathbf{y}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{x}_{k-1})$$

we can define the following sequential algorithm to iteratively compute it:

$$p(\mathbf{x}_{k:0}|\mathbf{y}_{k:0}) \propto p(\mathbf{y}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{x}_{k-1})p(\mathbf{x}_{k-1:0}|\mathbf{y}_{k-1:0}).$$

- An analysis step, in which the conditional pdf $p(\mathbf{x}_k|\mathbf{y}_{k:0})$ is updated using the latest

observation vector, $\mathbf{y}_k$,

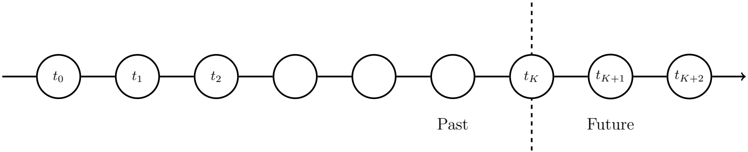
$$p(\mathbf{x}_k|\mathbf{y}_{k:0}) \propto p_{\boldsymbol{\eta}}\left(\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k)\right) p(\mathbf{x}_k|\mathbf{y}_{k-1:0}),$$

- which alternates with a forecast step that propagates this pdf, using the Chapman-Kolmogorov equation, forward in time until the new observation batch:

$$p(\mathbf{x}_{k+1}|\mathbf{y}_{k:0}) = \int \mathrm{d}\mathbf{x}\, p_{\boldsymbol{\eta}}\left(\mathbf{x}_k - \mathcal{M}_{k:k-1}(\mathbf{x}_{k-1})\right) p(\mathbf{x}_k|\mathbf{y}_{k:0})$$

to get $p(\mathbf{x}_{k+1}|\mathbf{y}_{k:0})$.

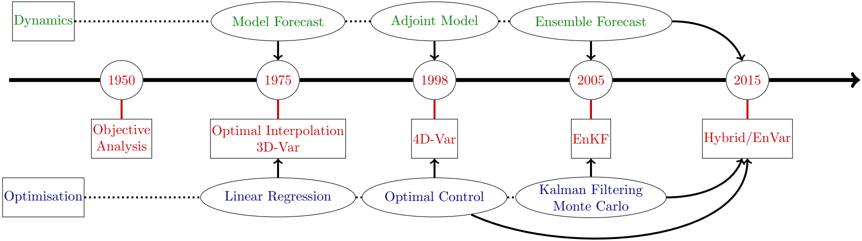
# Main goals of data assimilation




- Recall $\mathbf{x}_{K:0} = \{\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_K\}$, $\mathbf{y}_{K:0} = \{\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_K\}$:
  - Prediction: Estimate $\mathbf{x}_k$ for $k > K$, knowing $\mathbf{y}_{K:0}$,
  - Filtering: Estimate $\mathbf{x}_K$, knowing $\mathbf{y}_{K:0}$,
  - Smoothing: Estimate $\mathbf{x}_{K:0}$, knowing $\mathbf{y}_{K:0}$.
- Less formal names:
  - hindcasting, nowcasting and forecasting,
  - reanalysis,
  - parameter estimation.

# Mathematical methods in DA

▶ Introduction of mathematical methods in operational numerical weather prediction:




▶ Using increasingly complex mathematical methods and increasingly resolved high-dimensional models.

# Outline

# Gaussian approximation

▶ A key to obtain a (approximate) solution is to truncate the errors to second-order moments ∼ the Gaussian approximation. Most of DA methods are fully or partially based on this assumption.

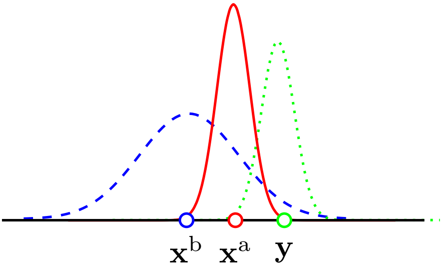
▶ The elementary building block of DA schemes is the statistical BLUE (Best Linear Unbiased Estimator) analysis. Time is considered fixed. $\mathbf{H}$ is assumed linear.

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \boldsymbol{\epsilon}^{\mathrm{o}}, \qquad \mathbf{x}^{\mathrm{b}} = \mathbf{x} + \boldsymbol{\epsilon}^{\mathrm{b}},$$

where $\boldsymbol{\epsilon}^{\mathrm{o}} \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$, and $\boldsymbol{\epsilon}^{\mathrm{b}} \sim \mathcal{N}(\mathbf{0}, \mathbf{B})$.

▶ Solution:

$$\begin{cases} \mathbf{x}^{\mathrm{a}} &= \mathbf{x}^{\mathrm{b}} + \mathbf{K}\left(\mathbf{y} - \mathbf{H}\mathbf{x}^{\mathrm{b}}\right) \\ \mathbf{K} &= \mathbf{B}\mathbf{H}^{\top}\left(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^{\top}\right)^{-1} \\ \mathbf{P}^{\mathrm{a}} &= \left(\mathbf{I} - \mathbf{K}\mathbf{H}\right)\mathbf{B}. \end{cases}$$



# Error statistics – Assumptions and definitions

- $\mathbf{x}^{\mathrm{t}}$ is defined as the true unknown state.

- Observation error statistics:

$$\boldsymbol{\epsilon}^{\mathrm{o}} = \mathbf{y} - \mathbf{H}\mathbf{x}^{\mathrm{t}} \qquad \text{with} \qquad \mathbb{E}[\boldsymbol{\epsilon}^{\mathrm{o}}] = \mathbf{0}, \qquad \mathbb{E}\left[\boldsymbol{\epsilon}^{\mathrm{o}}\boldsymbol{\epsilon}^{\mathrm{o}\top}\right] = \mathbf{R},$$

which is in particular satisfied if $\boldsymbol{\epsilon}^{\mathrm{o}} \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$.

- Background error statistics:

$$\boldsymbol{\epsilon}^{\mathrm{b}} = \mathbf{x}^{\mathrm{b}} - \mathbf{x}^{\mathrm{t}} \qquad \text{with} \qquad \mathbb{E}[\boldsymbol{\epsilon}^{\mathrm{b}}] = \mathbf{0}, \qquad \mathbb{E}\left[\boldsymbol{\epsilon}^{\mathrm{b}}\boldsymbol{\epsilon}^{\mathrm{b}\top}\right] = \mathbf{B}, \qquad \mathbb{E}\left[\boldsymbol{\epsilon}^{\mathrm{b}}\boldsymbol{\epsilon}^{\mathrm{o}\top}\right] = \mathbf{0}.$$

- Analysis error statistics:

$$\boldsymbol{\epsilon}^{\mathrm{a}} = \mathbf{x}^{\mathrm{a}} - \mathbf{x}^{\mathrm{t}} \qquad \text{with} \qquad \mathbb{E}[\boldsymbol{\epsilon}^{\mathrm{a}}] = \mathbf{0}, \qquad \mathbb{E}\left[\boldsymbol{\epsilon}^{\mathrm{a}}\boldsymbol{\epsilon}^{\mathrm{a}\top}\right] = \mathbf{P}^{\mathrm{a}}.$$

# Linear unbiased Ansatz for the estimate

- General Ansatz, linear in the observation and the first guess:

$$\mathbf{x}^{\mathrm{a}} = \mathbf{L}\mathbf{x}^{\mathrm{b}} + \mathbf{K}\mathbf{y}.$$

- Writing it in terms of errors:

$$\begin{aligned}
\mathbf{x}^{\mathrm{a}} - \mathbf{x}^{\mathrm{t}} &= \mathbf{L}\left(\mathbf{x}^{\mathrm{b}} - \mathbf{x}^{\mathrm{t}} + \mathbf{x}^{\mathrm{t}}\right) + \mathbf{K}\left(\mathbf{H}\mathbf{x}^{\mathrm{t}} + \boldsymbol{\epsilon}^{\mathrm{o}}\right) - \mathbf{x}^{\mathrm{t}}, \\
\boldsymbol{\epsilon}^{\mathrm{a}} &= \mathbf{L}\boldsymbol{\epsilon}^{\mathrm{b}} + \mathbf{K}\boldsymbol{\epsilon}^{\mathrm{o}} + \left(\mathbf{L} + \mathbf{K}\mathbf{H} - \mathbf{I}\right)\mathbf{x}^{\mathrm{t}}.
\end{aligned}$$

Then $\mathbb{E}[\boldsymbol{\epsilon}^{\mathrm{o}}] = \mathbf{0}$ and $\mathbb{E}[\boldsymbol{\epsilon}^{\mathrm{b}}] = \mathbf{0}$ imply $\mathbb{E}[\boldsymbol{\epsilon}^{\mathrm{a}}] = \left(\mathbf{L} + \mathbf{K}\mathbf{H} - \mathbf{I}\right)\mathbb{E}[\mathbf{x}^{\mathrm{t}}]$.
Hence, we wish to impose

$$\mathbf{L} = \mathbf{I} - \mathbf{K}\mathbf{H}.$$

- As a result, we obtain a linear and unbiased Ansatz:

$$\begin{aligned}
\mathbf{x}^{\mathrm{a}} &= (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{x}^{\mathrm{b}} + \mathbf{K}\mathbf{y}, \\
\mathbf{x}^{\mathrm{a}} &= \mathbf{x}^{\mathrm{b}} + \mathbf{K}\underbrace{\left(\mathbf{y} - \mathbf{H}\mathbf{x}^{\mathrm{b}}\right)}_{\text{innovation}}.
\end{aligned}$$

# Best linear unbiased estimator

- Posterior error:

$$\boldsymbol{\epsilon}^{\mathrm{a}} = \boldsymbol{\epsilon}^{\mathrm{b}} + \mathbf{K}(\boldsymbol{\epsilon}^{\mathrm{o}} - \mathbf{H}\boldsymbol{\epsilon}^{\mathrm{b}}),$$

so that

$$\begin{aligned}
\mathbf{P}^{\mathrm{a}} &= \mathbb{E}\left[(\boldsymbol{\epsilon}^{\mathrm{a}})(\boldsymbol{\epsilon}^{\mathrm{a}})^\top\right] = \mathbb{E}\left[\left(\boldsymbol{\epsilon}^{\mathrm{b}} + \mathbf{K}(\boldsymbol{\epsilon}^{\mathrm{o}} - \mathbf{H}\boldsymbol{\epsilon}^{\mathrm{b}})\right)\left(\boldsymbol{\epsilon}^{\mathrm{b}} + \mathbf{K}(\boldsymbol{\epsilon}^{\mathrm{o}} - \mathbf{H}\boldsymbol{\epsilon}^{\mathrm{b}})\right)^\top\right] \\
&= \mathbb{E}\left[\left(\mathbf{L}\boldsymbol{\epsilon}^{\mathrm{b}} + \mathbf{K}\boldsymbol{\epsilon}^{\mathrm{o}}\right)\left(\mathbf{L}\boldsymbol{\epsilon}^{\mathrm{b}} + \mathbf{K}\boldsymbol{\epsilon}^{\mathrm{o}}\right)^\top\right] = \mathbb{E}\left[\mathbf{L}\boldsymbol{\epsilon}^{\mathrm{b}}(\boldsymbol{\epsilon}^{\mathrm{b}})^\top\mathbf{L}^\top)\right] + \mathbb{E}\left[\mathbf{K}\boldsymbol{\epsilon}^{\mathrm{o}}(\boldsymbol{\epsilon}^{\mathrm{o}})^\top\mathbf{K}^\top\right] \\
&= \mathbf{L}\mathbf{B}\mathbf{L}^\top + \mathbf{K}\mathbf{R}\mathbf{K}^\top,
\end{aligned}$$

In summary:

$$\mathbf{P}^{\mathrm{a}} = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{B}(\mathbf{I} - \mathbf{K}\mathbf{H})^\top + \mathbf{K}\mathbf{R}\mathbf{K}^\top.$$

- We look for a metric as a global measure of the error. For instance $\mathrm{Tr}(\mathbf{P}^{\mathrm{a}})$. Let us find the optimal $\mathbf{K}$ that minimises this metric.

# Best linear unbiased estimator

- Variation of the metric with respect to a variation of $\mathbf{K}$, i.e. $\delta\mathbf{K}$:

$$\begin{aligned}\delta(\mathrm{Tr}(\mathbf{P}^{\mathrm{a}})) &= \mathrm{Tr}\left((-\delta\mathbf{K}\mathbf{H})\mathbf{B}\mathbf{L}^\top + \mathbf{L}\mathbf{B}(-\delta\mathbf{K}\mathbf{H})^\top + \delta\mathbf{K}\mathbf{R}\mathbf{K}^\top + \mathbf{K}\mathbf{R}\delta\mathbf{K}^\top\right)\\ &= \mathrm{Tr}\left((-\mathbf{L}\mathbf{B}^\top\mathbf{H}^\top - \mathbf{L}\mathbf{B}\mathbf{H}^\top + \mathbf{K}\mathbf{R}^\top + \mathbf{K}\mathbf{R})(\delta\mathbf{K})^\top\right)\\ &= 2\mathrm{Tr}\left((-\mathbf{L}\mathbf{B}\mathbf{H}^\top + \mathbf{K}\mathbf{R})(\delta\mathbf{K})^\top\right).\end{aligned}$$

- At optimality, one infers that $-(\mathbf{I}-\mathbf{K}^\star\mathbf{H})\mathbf{B}\mathbf{H}^\top + \mathbf{K}^\star\mathbf{R} = \mathbf{0}$, from which we obtain

$$\mathbf{K}^\star = \mathbf{B}\mathbf{H}^\top(\mathbf{R}+\mathbf{H}\mathbf{B}\mathbf{H}^\top)^{-1},$$

from which we get the BLUE solution:

$$\begin{cases}\mathbf{x}^{\mathrm{a}} = \mathbf{x}^{\mathrm{b}} + \mathbf{K}\left(\mathbf{y} - \mathbf{H}\mathbf{x}^{\mathrm{b}}\right)\\ \mathbf{K} = \mathbf{B}\mathbf{H}^\top\left(\mathbf{R}+\mathbf{H}\mathbf{B}\mathbf{H}^\top\right)^{-1}\\ \mathbf{P}^{\mathrm{a}} = \left(\mathbf{I}-\mathbf{K}\mathbf{H}\right)\mathbf{B}.\end{cases}$$

# Outline

# 3D-Var and BLUE in the linear case: derivation

- 3D-Var cost function:

$$J(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}-\mathbf{x}^{\mathrm{b}}\|^2_{\mathbf{B}^{-1}} + \frac{1}{2}\|\mathbf{y}-\mathbf{H}\mathbf{x}\|^2_{\mathbf{R}^{-1}}, \qquad \text{with} \quad \|\mathbf{x}\|^2_{\mathbf{A}} = \mathbf{x}^\top\mathbf{A}\mathbf{x}.$$

- Let us minimise $J$ and compute the variation of $J(\mathbf{x})$ with respect to a variation of $\mathbf{x}$:

$$\begin{aligned}
\delta J(\mathbf{x}) &= \frac{1}{2}\left(\delta\mathbf{x}\right)^\top\mathbf{B}^{-1}\left(\mathbf{x}-\mathbf{x}^{\mathrm{b}}\right) + \frac{1}{2}\left(\mathbf{x}-\mathbf{x}^{\mathrm{b}}\right)^\top\mathbf{B}^{-1}\delta\mathbf{x} \\
&\quad + \frac{1}{2}\left(-\mathbf{H}\delta\mathbf{x}\right)^\top\mathbf{R}^{-1}\left(\mathbf{y}-\mathbf{H}\mathbf{x}\right) + \frac{1}{2}\left(\mathbf{x}^{\mathrm{b}}-\mathbf{H}\mathbf{x}\right)\mathbf{R}^{-1}\left(-\mathbf{H}\delta\mathbf{x}\right) \\
&= \left(\delta\mathbf{x}\right)^\top\mathbf{B}^{-1}\left(\mathbf{x}-\mathbf{x}^{\mathrm{b}}\right) - \left(\delta\mathbf{x}\right)^\top\mathbf{H}^\top\mathbf{R}^{-1}\left(\mathbf{y}-\mathbf{H}\mathbf{x}\right) \\
&= \left(\delta\mathbf{x}\right)^\top\nabla J\,.
\end{aligned}$$

- The extremum condition is $\nabla J = \mathbf{B}^{-1}(\mathbf{x}^\star-\mathbf{x}^{\mathrm{b}}) - \mathbf{H}^\top\mathbf{R}^{-1}(\mathbf{y}-\mathbf{H}\mathbf{x}^\star) = \mathbf{0}$, which yields:

$$\mathbf{x}^\star = \mathbf{x}^{\mathrm{b}} + \underbrace{(\mathbf{B}^{-1}+\mathbf{H}^\top\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^\top\mathbf{R}^{-1}}_{\mathbf{K}^\star}(\mathbf{y}-\mathbf{H}\mathbf{x}^{\mathrm{b}})\,.$$

Thanks to the Sherman-Morrison-Woodbury identity,

$$\mathbf{K}^\star = (\mathbf{B}^{-1}+\mathbf{H}^\top\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^\top\mathbf{R}^{-1} = \mathbf{B}\mathbf{H}^\top\left(\mathbf{R}+\mathbf{H}\mathbf{B}\mathbf{H}^\top\right)^{-1}.$$

$\longrightarrow$ $\mathbf{x}^\star$ coincides with the BLUE optimal analysis $\mathbf{x}^{\mathrm{a}}$.

# 3D-Var and optimal interpolation

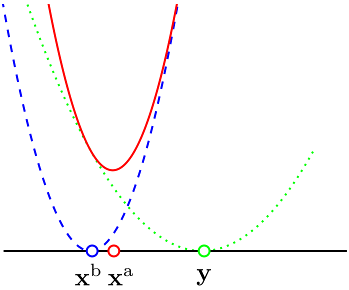
- Variational formulation of the same problem

$$J(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}-\mathbf{x}^{\mathrm{b}}\|^2_{\mathbf{B}^{-1}} + \frac{1}{2}\|\mathbf{y}-\mathbf{H}\mathbf{x}\|^2_{\mathbf{R}^{-1}},$$

which is equivalent to BLUE.

- Probabilistic/Bayesian interpretation:

$$p(\mathbf{x}|\mathbf{y}) \propto e^{-J(\mathbf{x})}$$



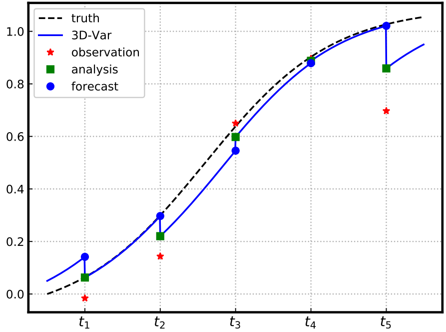

- Capable of handling a nonlinear observation operator using standard nonlinear optimisation methods:

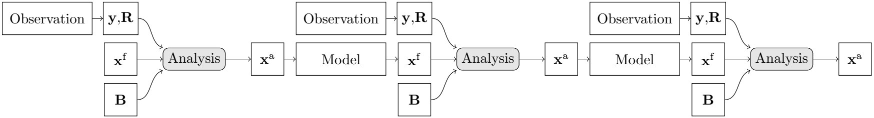
$$J(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}-\mathbf{x}^{\mathrm{b}}\|^2_{\mathbf{B}^{-1}} + \frac{1}{2}\|\mathbf{y}-\mathcal{H}(\mathbf{x})\|^2_{\mathbf{R}^{-1}}.$$

# Chaining the analyses in time

- Chaining the BLUE/3D-Var cycles:
  1. Analysis with a forecast at $t_k$: $\mathbf{x}^\mathrm{f}_k$ and with static information $\mathbf{B}$: $\mathbf{x}^\mathrm{a}_k$,
  2. Forecast to $t_{k+1}$: $\mathbf{x}^\mathrm{f}_{k+1} = \mathcal{M}_{k+1:k}(\mathbf{x}^\mathrm{a}_k)$.

- Also known as optimal interpolation (if the analysis step is BLUE).
- Relatively cheap. Used in oceanography, atmospheric chemistry. Requires a smart construction of $\mathbf{B}$.
- But the information about the errors is not propagated in time...



# The Kalman filter

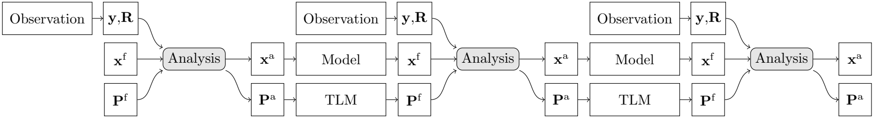
▶ Similar to optimal interpolation. But, now, we want to replace the static $\mathbf{B}$ with a dynamic $\mathbf{P}^\mathrm{f}$ which needs updating and propagating.

▶ Analysis step:

$$\begin{aligned}\mathbf{x}^\mathrm{a}_k &= \mathbf{x}^\mathrm{f}_k + \mathbf{K}_k\left(\mathbf{y}_k - \mathbf{H}_k\mathbf{x}^\mathrm{f}_k\right),\\ \mathbf{K}_k &= \mathbf{P}^\mathrm{f}_k\mathbf{H}_k^\top\left(\mathbf{R}_k + \mathbf{H}_k\mathbf{P}^\mathrm{f}\mathbf{H}_k^\top\right)^{-1},\\ \mathbf{P}^\mathrm{a}_k &= \left(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k\right)\mathbf{P}^\mathrm{f}_k.\end{aligned}$$

▶ Forecast step:

$$\begin{aligned}\mathbf{x}^\mathrm{f}_{k+1} &= \mathbf{M}_{k+1:k}\mathbf{x}^\mathrm{a}_k,\\ \mathbf{P}^\mathrm{f}_{k+1} &= \mathbf{M}_{k+1:k}\mathbf{P}^\mathrm{a}_k\mathbf{M}^\top_{k+1:k} + \mathbf{Q}_{k+1}.\end{aligned}$$



# The extended Kalman filter

▶ Optimal if the model and observation operators are linear and if all the initial and observations errors are Gaussian: it gives the exact Gaussian solution of Bayes' rule.

▶ Can be extended to nonlinear models with:

$$\begin{aligned}\mathbf{x}^{\mathrm{f}}_{k+1} &= \mathcal{M}_{k+1:k}(\mathbf{x}^{\mathrm{a}}_k),\\ \mathbf{P}^{\mathrm{f}}_{k+1} &= \mathbf{M}_{k+1:k}\mathbf{P}^{\mathrm{a}}_k\mathbf{M}^{\top}_{k+1:k} + \mathbf{Q}_{k+1},\end{aligned}$$

where $\mathbf{M}_{k+1:k}$ is the tangent linear model (linearisation at $\mathbf{x}^{\mathrm{a}}_k$) of $\mathcal{M}_{k+1:k}$.

▶ Extremely costly for large geophysical models: storage space (storage of $\mathbf{P}^{\mathrm{f}}$) and computations ($\mathbf{M}_{k+1:k}\mathbf{P}^{\mathrm{f}}_k\mathbf{M}^{\top}_{k+1:k}$ requires $2N_x$ integrations of the model).

▶ Solutions: The reduced-rank / ensemble Kalman filters (wait for the end of the lecture).

# The extended Kalman filter: numerical illustration

- Anharmonic oscillator:

$$\frac{\mathrm{d}^2 x}{\mathrm{d}\,t^2} - \Omega^2\, x + \Lambda^2\, x^3 = 0,$$

whose numerical implementation is

$$x_0 = 0\,, \quad x_1 = 1 \quad \text{and for } 1 \le k \le N: \quad x_{k+1} - 2x_k + x_{k-1} = \omega^2\, x_k - \lambda^2\, x_k^3.$$

$\longrightarrow$ Equations for a material dot in a double well potential $V(x) = -\frac{1}{2}\Omega^2 x^2 + \frac{1}{4}\Lambda^2 x^4$.

- Markovian dynamics with an augmented state vector:

$$\mathbf{u}_k = \begin{bmatrix} x_k \\ x_{k-1} \end{bmatrix},$$

with the augmented dynamics

$$\mathcal{M}_{k+1:k} = \begin{bmatrix} 2 + \omega^2 - \lambda^2 x_k^2 & -1 \\ 1 & 0 \end{bmatrix},$$

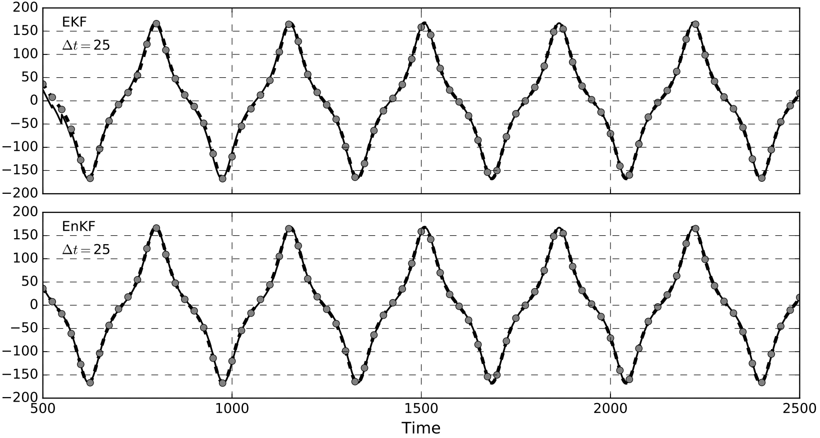
yields

$$\mathbf{u}_{k+1} = \mathcal{M}_{k+1:k}(\mathbf{u}_k).$$

- $\mathbf{H}_k = [1, 0]$. The observation equation is $y_k = \mathbf{H}_k \mathbf{u}_k + \epsilon_k$.

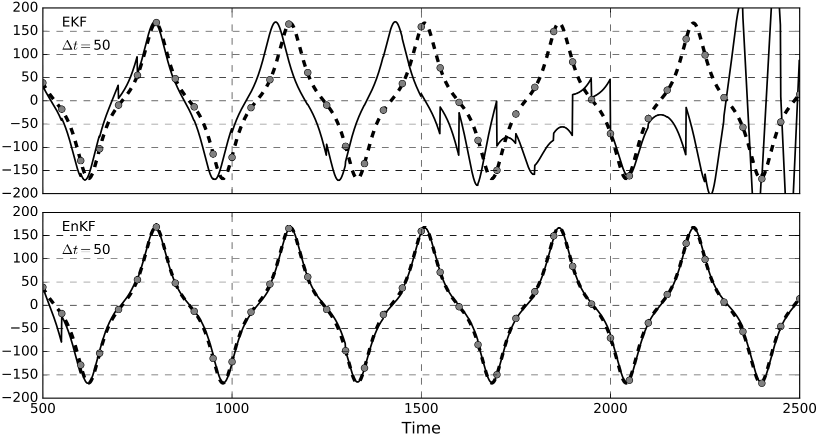
# The extended Kalman filter: numerical illustration

- Comparison with the EnKF that does not rely on the tangent linear approximation.



# The extended Kalman filter: numerical illustration

- Comparison with the EnKF that does not rely on the tangent linear approximation.



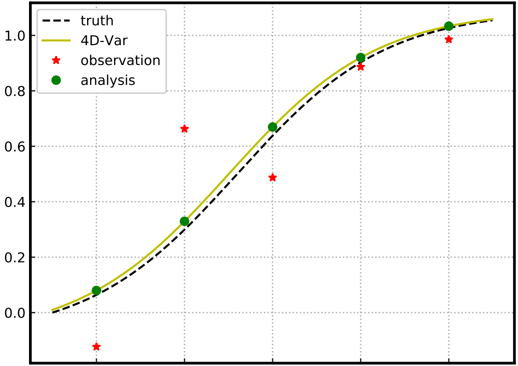
# 4D-Var

- Strongly constrained 4D-Var, i.e. *assuming the model* is perfect (no model error)

$$J(\mathbf{x}_0) = \frac{1}{2}\|\mathbf{x}_0 - \mathbf{x}_0^{\mathbf{b}}\|^2_{\mathbf{B}^{-1}} + \frac{1}{2}\sum_{k=0}^{K}\|\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k)\|^2_{\mathbf{R}_k^{-1}},$$

under the constraints that $\mathbf{x}_{k+1} = \mathcal{M}_{k+1:k}(\mathbf{x}_k)$ for $k = 0, \ldots, K-1$.

- Fits a model trajectory through the 4D data points.



# 4D-Var: algorithm

- Lagrangian for 4D-Var:

$$L(\mathbf{x}_{K:0}, \boldsymbol{\lambda}_{k:0}) = \frac{1}{2}\|\mathbf{x}_0 - \mathbf{x}_0^{\mathbf{b}}\|^2_{\mathbf{B}^{-1}} + \frac{1}{2}\sum_{k=0}^{K}\|\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k)\|^2_{\mathbf{R}_k^{-1}} + \sum_{k=1}^{K}\boldsymbol{\lambda}_k^\top\left(\mathbf{x}_k - \mathcal{M}_{k:k-1}(\mathbf{x}_{k-1})\right)$$

- Gradient of the Lagrangian with respect to $\mathbf{x}_{K:0}$:

$$\begin{aligned}
\nabla_{\mathbf{x}_0} L(\mathbf{x}_0) &= \mathbf{B}^{-1}\left(\mathbf{x}_0 - \mathbf{x}_0^{\mathbf{b}}\right) - \mathbf{H}_0^\top \mathbf{R}_0^{-1}\left(\mathbf{y}_0 - \mathbf{H}_0(\mathbf{x}_0)\right) - \mathbf{M}_{1:0}^\top \boldsymbol{\lambda}_1, \\
\nabla_{\mathbf{x}_k} L(\mathbf{x}_0) &= -\mathbf{H}_k^\top \mathbf{R}_k^{-1}\left(\mathbf{y}_k - \mathbf{H}_k(\mathbf{x}_k)\right) - \mathbf{M}_{k+1:k}^\top \boldsymbol{\lambda}_{k+1} + \boldsymbol{\lambda}_k, \\
\nabla_{\mathbf{x}_K} L(\mathbf{x}_0) &= -\mathbf{H}_K^\top \mathbf{R}_K^{-1}\left(\mathbf{y}_K - \mathbf{H}_K(\mathbf{x}_K)\right) + \boldsymbol{\lambda}_K.
\end{aligned}$$

- Requires the computation of the tangent linear and adjoint of $\mathcal{H}_k$ and $\mathcal{M}_{k+1:k}$.

- No perfect (general purpose) automatic differentiation tool: developing and maintaining the adjoint codes is computationally very costly!
$\longrightarrow$ *written 2019 – this has changed! – more on this tomorrow!*

# 4D-Var: algorithm

- Algorithm: one outer loop
  1. Given the initial condition $\mathbf{x}_0$, compute the trajectory $\mathbf{x}_{K:0}$ with the dynamical model $\mathcal{M}$.
  2. Compute the adjoint trajectory backwards in time:

$$\begin{aligned}\boldsymbol{\lambda}_K &= \mathbf{H}_K^\top \mathbf{R}_K^{-1}\left(\mathbf{y}_K - \mathbf{H}_K(\mathbf{x}_K)\right),\\ \boldsymbol{\lambda}_k &= \mathbf{H}_k^\top \mathbf{R}_k^{-1}\left(\mathbf{y}_k - \mathbf{H}_k(\mathbf{x}_k)\right) - \mathbf{M}_{k+1:k}^\top \boldsymbol{\lambda}_{k+1},\\ \boldsymbol{\lambda}_0 &= \mathbf{H}_0^\top \mathbf{R}_0^{-1}\left(\mathbf{y}_0 - \mathbf{H}_0(\mathbf{x}_0)\right) - \mathbf{M}_{1:0}^\top \boldsymbol{\lambda}_1.\end{aligned}$$

  3. This finally yields:

$$\nabla_{\mathbf{x}_0} J(\mathbf{x}_0) = \mathbf{B}^{-1}\left(\mathbf{x}_0 - \mathbf{x}_0^{\mathbf{b}}\right) - \boldsymbol{\lambda}_0.$$

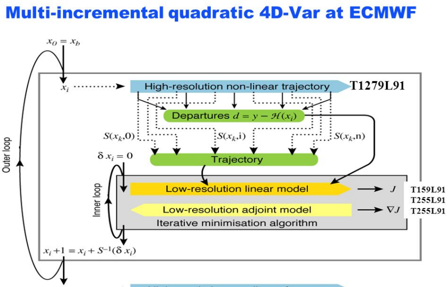
- Can be used to feed any gradient-based minimisation scheme (Newton, Gauss-Newton, L-BFGS, conjugate-gradient, Levenberg-Marquardt, trust region methods).

# 4D-Var: algorithm

▶ For high-dimensional systems: incremental strategy with outer/inner loops.
The inner-loop Lagrangian, which is quadratic in $\delta\mathbf{x}_{K:0}$, is

$$L^{(p)}(\delta\mathbf{x}_{K:0},\boldsymbol{\lambda}_{k:0}) = \frac{1}{2}\left\|\mathbf{x}_0^{(p)} - \mathbf{x}_0^{\mathrm{b}} + \delta\mathbf{x}_0\right\|^2_{\mathbf{B}^{-1}} + \frac{1}{2}\sum_{k=0}^{K}\left\|\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k^{(p)}) + \mathbf{H}^{(p)}(\delta\mathbf{x}_k)\right\|^2_{\mathbf{R}_k^{-1}}$$
$$+ \sum_{k=1}^{K}\boldsymbol{\lambda}_k^\top\left(\mathbf{x}_{k+1}^{(p)} - \mathcal{M}_{k+1:k}(\mathbf{x}_k^{(p)}) - \mathbf{M}_{k:k-1}^{(p)}(\delta\mathbf{x}_{k-1})\right).$$

It can efficiently be solved using a conjugate-gradient algorithm.



# 4D-Var: algorithm

- Let us assume Gaussian model error:

$$\mathbf{x}_k = \mathcal{M}_{k:k-1}(\mathbf{x}_{k-1}) + \boldsymbol{\eta}_k, \qquad \boldsymbol{\eta}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_k).$$

- Weakly constrained 4D-Var, i.e. assuming the model is imperfect [Trémolet 2006]

$$J(\mathbf{x}_{K:0}) = \frac{1}{2}\|\mathbf{x}_0 - \mathbf{x}_0^{\mathbf{b}}\|^2_{\mathbf{B}^{-1}} + \frac{1}{2}\sum_{k=0}^{K}\|\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k)\|^2_{\mathbf{R}_k^{-1}} + \frac{1}{2}\sum_{k=1}^{K}\|\mathbf{x}_k - \mathcal{M}_{k:k-1}(\mathbf{x}_{k-1})\|^2_{\mathbf{Q}_k^{-1}}.$$

- Adds much flexibility to trajectory fitting.

- Huge control variables ($K$ times bigger) for a very specific form of model error...
$\longrightarrow$ A simplified variant of weakly constrained 4D-Var has been implemented in the top layers of the IFS to correct a large bias [Laloyaux et al. 2020].

# The ensemble Kalman filter

▶ The idea [Evensen 1994; Houtekamer and Mitchell 1998] is to make the KF work in high dimensions and replace $\mathbf{P}$ ($\mathbf{P}^{\mathrm{a}}$ and $\mathbf{P}^{\mathrm{f}}$) with an ensemble of states $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{N_\mathrm{e}}$. The moments of the error could theoretically be approximated by the sample/empirical moments:

$$\overline{\mathbf{x}}^{\mathrm{f}} = \frac{1}{N_\mathrm{e}} \sum_{i=1}^{N_\mathrm{e}} \mathbf{x}_i^{\mathrm{f}}, \qquad \mathbf{P}^{\mathrm{f}} \approx \frac{1}{N_\mathrm{e}-1} \sum_{i=1}^{N_\mathrm{e}} \left(\mathbf{x}_i^{\mathrm{f}} - \overline{\mathbf{x}}^{\mathrm{f}}\right) \left(\mathbf{x}_i^{\mathrm{f}} - \overline{\mathbf{x}}^{\mathrm{f}}\right)^\top .$$

▶ Define the normalised anomaly or perturbation matrix $\in \mathbb{R}^{N_x \times N_\mathrm{e}}$

$$\left[\mathbf{X}_\mathrm{f}\right]_i = \frac{\mathbf{x}_i^{\mathrm{f}} - \overline{\mathbf{x}}^{\mathrm{f}}}{\sqrt{N_\mathrm{e}-1}} \qquad \Longrightarrow \qquad \mathbf{P}^{\mathrm{f}} \approx \mathbf{X}_\mathrm{f}\mathbf{X}_\mathrm{f}^\top .$$

Likewise

$$\overline{\mathbf{x}}^{\mathrm{a}} = \frac{1}{N_\mathrm{e}} \sum_{i=1}^{N_\mathrm{e}} \mathbf{x}_i^{\mathrm{a}}, \qquad \mathbf{P}^{\mathrm{a}} \approx \mathbf{X}_\mathrm{a}\mathbf{X}_\mathrm{a}^\top \quad \text{where} \quad \left[\mathbf{X}_\mathrm{a}\right]_i = \frac{\mathbf{x}_i^{\mathrm{a}} - \overline{\mathbf{x}}^{\mathrm{a}}}{\sqrt{N_\mathrm{e}-1}}.$$

# The ensemble Kalman filter: Ansatz and mean update

- An educated guess would suggest, for $i = 1 \dots N_\mathrm{e}$:

$$\mathbf{x}_i^\mathrm{a} = \mathbf{x}_i^\mathrm{f} + \mathbf{K}\left(\mathbf{y} - \mathbf{H}\mathbf{x}_i^\mathrm{f}\right).$$

but the correct answer is actually

$$\mathbf{x}_i^\mathrm{a} = \mathbf{x}_i^\mathrm{f} + \mathbf{K}\left(\mathbf{y} + \boldsymbol{\epsilon}_i - \mathbf{H}\mathbf{x}_i^\mathrm{f}\right).$$

where $\boldsymbol{\epsilon}_i$ is a stochastic noise sampled from $\mathcal{N}(\mathbf{0}, \mathbf{R})$, for each member.

- Checking the mean: on average, and summing over the ensemble members:

$$\overline{\mathbf{x}}^\mathrm{a} = \overline{\mathbf{x}}^\mathrm{f} + \mathbf{K}\left(\mathbf{y} - \mathbf{H}\overline{\mathbf{x}}^\mathrm{f}\right),$$

which is the same as the Kalman filter's mean update.

# The ensemble Kalman filter: perturbations update

▶ Checking the ensemble update: on average, does it mimic the Kalman filter? We define

$$\overline{\boldsymbol{\epsilon}} = \frac{1}{N_\mathrm{e}} \sum_{i=1}^{N_\mathrm{e}} \boldsymbol{\epsilon}_i, \qquad \boldsymbol{\Theta} = \frac{1}{\sqrt{N_\mathrm{e}-1}} \left[ \boldsymbol{\epsilon}_1 - \overline{\boldsymbol{\epsilon}} \quad \boldsymbol{\epsilon}_2 - \overline{\boldsymbol{\epsilon}} \quad \cdots \quad \boldsymbol{\epsilon}_{N_\mathrm{e}} - \overline{\boldsymbol{\epsilon}} \right].$$

The perturbations update then reads (ensemble minus the mean):

$$\mathbf{X}_\mathrm{a} = (\mathbf{I}_\mathrm{x} - \mathbf{K}\mathbf{H})\mathbf{X}_\mathrm{f} + \mathbf{K}\boldsymbol{\Theta},$$

which yields the empirical analysis error covariances:

$$\mathbf{P}^\mathrm{a} = (\mathbf{I}_\mathrm{x} - \mathbf{K}\mathbf{H})\mathbf{P}^\mathrm{f}(\mathbf{I}_\mathrm{x} - \mathbf{K}\mathbf{H})^\top + \mathbf{K}\boldsymbol{\Theta}\boldsymbol{\Theta}^\top\mathbf{K}^\top + (\mathbf{I}_\mathrm{x} - \mathbf{K}\mathbf{H})\mathbf{X}_\mathrm{f}\boldsymbol{\Theta}^\top\mathbf{K}^\top + \mathbf{K}\boldsymbol{\Theta}\mathbf{X}_\mathrm{f}^\top(\mathbf{I}_\mathrm{x} - \mathbf{K}\mathbf{H})^\top,$$

whose average on $\boldsymbol{\Theta}$ is

$$\mathbb{E}[\mathbf{P}^\mathrm{a}] = (\mathbf{I}_\mathrm{x} - \mathbf{K}\mathbf{H})\mathbf{P}^\mathrm{f}(\mathbf{I}_\mathrm{x} - \mathbf{K}\mathbf{H})^\top + \mathbf{K}\mathbf{R}\mathbf{K}^\top = (\mathbf{I}_\mathrm{x} - \mathbf{K}\mathbf{H})\mathbf{P}^\mathrm{f}.$$

The last identity is valid if $\mathbf{K}$ is the (optimal) Kalman gain.

▶ In the absence of the observation stochastic noise, the posterior error statistics would be incorrect!

# The ensemble Kalman filter: forecast

- Kalman gain representations:

Empirical: denoting $\mathbf{Y}_\mathrm{f} = \mathbf{H}\mathbf{X}_\mathrm{f} + \mathbf{\Theta}$, we have $\mathbf{K} = \mathbf{X}_\mathrm{f}\mathbf{Y}_\mathrm{f}^\top \left(\mathbf{Y}_\mathrm{f}\mathbf{Y}_\mathrm{f}^\top\right)^{-1}$

Deterministic: denoting $\mathbf{Y}_\mathrm{f} = \mathbf{H}\mathbf{X}_\mathrm{f}$, we have $\mathbf{K} = \mathbf{X}_\mathrm{f}\mathbf{Y}_\mathrm{f}^\top \left(\mathbf{R} + \mathbf{Y}_\mathrm{f}\mathbf{Y}_\mathrm{f}^\top\right)^{-1}$

- Forecast step: The ensemble is propagated using the full nonlinear model

$$\mathbf{x}^\mathrm{f}_{i,k+1} = \mathcal{M}_{k+1:k}\left(\mathbf{x}^\mathrm{a}_{i,k}\right),$$

whereas the extended Kalman filter uses the tangent linear model.

- Numerically costly ($N_\mathrm{e}$ propagations) but
  - the forecast scheme is embarrassingly parallel,
  - no need to derive the tangent linear model of the full model.

# The ensemble Kalman filter: surrogate for $\mathbf{H}$

▶ Instead of estimating $\mathbf{P}^{\mathrm{f}}\mathbf{H}^\top = \mathbf{X}_{\mathrm{f}}\mathbf{Y}_{\mathrm{f}}^\top$ and $\mathbf{H}\mathbf{P}^{\mathrm{f}}\mathbf{H}^\top = \mathbf{Y}_{\mathrm{f}}\mathbf{Y}_{\mathrm{f}}^\top$ in the Kalman gain, we can use the ensemble:

$$\overline{\mathbf{y}}^{\mathrm{f}} = \frac{1}{N_{\mathrm{e}}}\sum_{i=1}^{N_{\mathrm{e}}} \mathcal{H}(\mathbf{x}_i^{\mathrm{f}}),$$

$$\mathbf{P}^{\mathrm{f}}\mathbf{H}^\top = \frac{1}{N_{\mathrm{e}}-1}\sum_{i=1}^{N_{\mathrm{e}}} \left(\mathbf{x}_i^{\mathrm{f}} - \overline{\mathbf{x}}^{\mathrm{f}}\right)\left[\mathcal{H}(\mathbf{x}_i^{\mathrm{t}}) - \overline{\mathbf{y}}^{\mathrm{f}}\right]^\top,$$

$$\mathbf{H}\mathbf{P}^{\mathrm{f}}\mathbf{H}^\top = \frac{1}{N_{\mathrm{e}}-1}\sum_{i=1}^{N_{\mathrm{e}}} \left[\mathcal{H}(\mathbf{x}_i^{\mathrm{f}}) - \overline{\mathbf{y}}^{\mathrm{f}}\right]\left[\mathcal{H}(\mathbf{x}_i^{\mathrm{f}}) - \overline{\mathbf{y}}^{\mathrm{f}}\right]^\top.$$

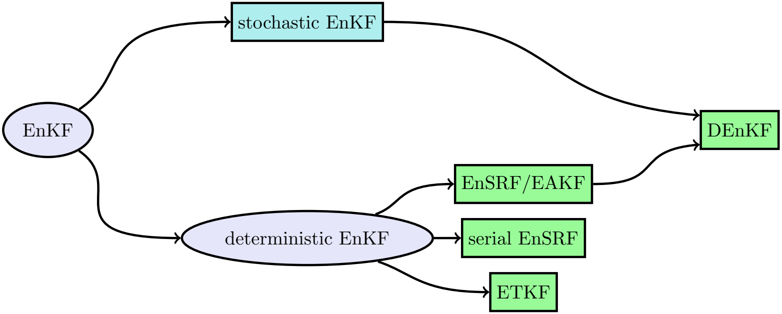
These approximations rely on the key assumption:

$$[\mathbf{Y}_{\mathrm{f}}]_i = \mathbf{H}\left(\mathbf{x}_i^{\mathrm{f}} - \overline{\mathbf{x}}^{\mathrm{f}}\right) \approx \mathcal{H}(\mathbf{x}_i^{\mathrm{f}}) - \overline{\mathbf{y}}^{\mathrm{f}}.$$

▶ This is sometimes called the secant method (alternative to finite-differences).

# The ensemble Kalman filter: a bunch of methods

- Two main flavors of EnKFs: stochastic and deterministic, but many variants.




- Several significant precursors and alternatives: reduced-rank square-root Kalman filter, SEEK, SEIK, unscented Kalman filter, etc.

# Localisation

▶ Covariance localisation seeks to regularise the sample covariance to mitigate the rank-deficiency of $\mathbf{P}^\mathrm{e}$ and the appearance of spurious correlations.

▶ Solution: compute the Schur product of $\mathbf{P}^\mathrm{e}$ with a well chosen smooth correlation matrix $\boldsymbol{\rho}$, that has exponentially vanishing correlations for distant parts.

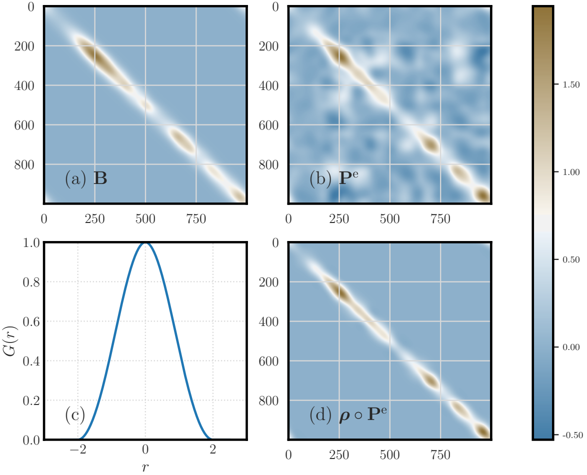
The Schur product of $\boldsymbol{\rho}$ and $\mathbf{B}$ is defined by (tapering of covariances)

$$[\boldsymbol{\rho} \circ \mathbf{P}^\mathrm{e}]_{ij} = [\boldsymbol{\rho}]_{ij} [\mathbf{P}^\mathrm{e}]_{ij} . \tag{1}$$

Applicable only if the long-range error correlations are negligible.

▶ The Schur product theorem ensures that this product is positive semi-definite, a proper covariance matrix. For sufficiently regular $\boldsymbol{\rho}$, $\boldsymbol{\rho} \circ \mathbf{P}^\mathrm{e}$ turns out to be full-rank.

# Covariance localisation with the Gaspari-Cohn function




Panel (a): True covariance matrix. Panel (b): Sample covariance matrix.
Panel (c): Gaspari-Cohn based correlation matrix used for covariance localisation.
Panel (d): Tapered covariance matrix.

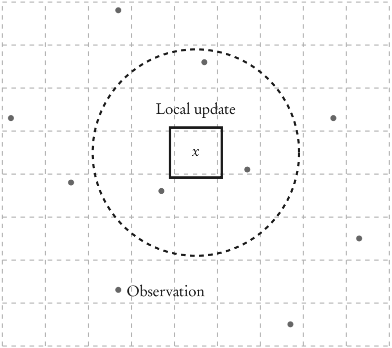
# Domain localisation

▶ Domain localisation: divide & conquer.

The DA analysis is performed in parallel in local domains. The outcomes of these analyses are later sewed together.

Applicable only if the long-range error correlations are negligible.

Elegant but nor suited for the assimilation of non-local observations such as radiances.




▶ Both localisation schemes have successfully been applied to the EnKF [Hamill et al. 2001; Houtekamer and Mitchell 2001; Evensen 2003; Hunt et al. 2007].

# Inflation

▶ Localisation addresses the rank-deficiency issue, but sampling errors are not entirely removed in the process: long EnKF runs may ultimately diverge!

▶ Ad hoc means to counteract sampling errors is to inflate the error covariance matrix by a multiplicative factor $\lambda^2 \geq 1$:

$$\mathbf{P}^{\mathrm{e}} \longrightarrow \lambda^2 \mathbf{P}^{\mathrm{e}}, \tag{2}$$

or, alternatively,

$$\mathbf{x}_{[n]} \longrightarrow \overline{\mathbf{x}} + \lambda \left( \mathbf{x}_{[n]} - \overline{\mathbf{x}} \right). \tag{3}$$

▶ Inflation can also come in an additive form: $\mathbf{x}_{[n]} \longrightarrow \mathbf{x}_{[n]} + \boldsymbol{\epsilon}_{[n]}$.

▶ Note that inflation is not only used to cure sampling errors, but is also often used to counteract model error impact.

▶ As a drawback, inflation often needs to be tuned, which is numerically costly. Hence, adaptive schemes have been developed to make the task more automatic [El Gharamti 2018; Raanes et al. 2019].

# References I


[1] B. D. O. Anderson and J. B. Moore. *Optimal Filtering*. Englewood Cliffs, New Jersey: Prentice-Hall, Inc, 1979, p. 357.

[2] J. L. Anderson. “An ensemble adjustment Kalman filter for data assimilation”. In: *Mon. Wea. Rev.* 129 (2001), pp. 2884–2903.

[3] A. Andrews. “A square root formulation of the Kalman covariance equations”. In: *AIAA J.* 6 (1968), pp. 1165–1166.

[4] E. Arbogast, G. Desroziers, and L. Berre. “A parallel implementation of a 4DEnVar ensemble”. In: *Q. J. R. Meteorol. Soc.* 143 (2017), pp. 2073–2083.

[5] M. Asch, M. Bocquet, and M. Nodet. *Data Assimilation: Methods, Algorithms, and Applications*. Fundamentals of Algorithms. SIAM, Philadelphia, 2016, p. 324.

[6] C. H. Bishop, B. J. Etherton, and S. J. Majumdar. “Adaptive Sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical Aspects”. In: *Mon. Wea. Rev.* 129 (2001), pp. 420–436.

[7] C. H. Bishop, J. S. Whitaker, and L. Lei. “Gain form of the Ensemble Transform Kalman Filter and its relevance to satellite data assimilation with model space ensemble covariance localization”. In: *Mon. Wea. Rev.* 145 (2017), pp. 4575–4592.

[8] M. Bocquet. “Localization and the iterative ensemble Kalman smoother”. In: *Q. J. R. Meteorol. Soc.* 142 (2016), pp. 1075–1089.

[9] M. Bocquet and A. Farchi. “On the consistency of the perturbation update of local ensemble square root Kalman filters”. In: *Tellus A* 71 (2019), pp. 1–21.

[10] M. Bocquet and P. Sakov. “An iterative ensemble Kalman smoother”. In: *Q. J. R. Meteorol. Soc.* 140 (2014), pp. 1521–1535.

[11] M. Buehner. “Ensemble-derived stationary and flow-dependent background-error covariances: Evaluation in a quasi-operational NWP setting”. In: *Q. J. R. Meteorol. Soc.* 131 (2005), pp. 1013–1043.

[12] G. Burgers, P. J. van Leeuwen, and G. Evensen. “Analysis scheme in the ensemble Kalman filter”. In: *Mon. Wea. Rev.* 126 (1998), pp. 1719–1724.

[13] A. Carrassi et al. “Data Assimilation in the Geosciences: An overview on methods, issues, and perspectives”. In: *WIREs Climate Change* 9 (2018), e535.

[14] S. E. Cohn, N. S. Sivakumaran, and R. Todling. “A Fixed-Lag Kalman Smoother for Retrospective Data Assimilation”. In: *Mon. Wea. Rev.* 122 (1994), pp. 2838–2867.

[15] E. Cosme et al. “Smoothing problems in a Bayesian framework and their linear Gaussian solutions”. In: *Mon. Wea. Rev.* 140 (2012), pp. 683–695.

[16] R. Daley. *Atmospheric Data Analysis*. Cambridge University Press, New-York, 1991, p. 472.

# References II


[17] M. El Gharamti. "Enhanced Adaptive Inflation Algorithm for Ensemble Filters". In: *Mon. Wea. Rev.* 146 (2018), pp. 623–640.

[18] G. Evensen. *Data Assimilation: The Ensemble Kalman Filter*. Second. Springer-Verlag Berlin Heildelberg, 2009, p. 307.

[19] G. Evensen. "Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics". In: *J. Geophys. Res.* 99 (1994), pp. 10143–10162.

[20] G. Evensen. "The Ensemble Kalman Filter: Theoretical Formulation and Practical Implementation". In: *Ocean Dynamics* 53 (2003), pp. 343–367.

[21] G. Evensen and P. J. van Leeuwen. "An Ensemble Kalman Smoother for Nonlinear Dynamics". In: *Mon. Wea. Rev.* 128 (2000), pp. 1852–1867.

[22] A. Farchi and M. Bocquet. "On the efficiency of covariance localisation of the ensemble Kalman filter using augmented ensembles". In: *Front. Appl. Math. Stat.* 5 (2019), p. 3.

[23] A. Farchi and M. Bocquet. "Review article: Comparison of local particle filters and new implementations". In: *Nonlin. Processes Geophys.* 25 (2018), pp. 765–807.

[24] S. J. Fletcher. *Data assimilation for the geosciences: From theory to application*. Elsevier, 2017.

[25] M. Ghil and P. Malanotte-Rizzoli. "Data assimilation in meteorological and oceanography". In: *Advanc. in Geophys.* 33 (1991), pp. 141–266.

[26] N. J. Gordon, D. J. Salmond, and A. F. M. Smith. "Novel approach to nonlinear/non-Gaussian Bayesian state estimation". In: *IEE Proc.-F* 140 (1993), pp. 107–113.

[27] N. Halko, P.-G. Martinsson, and J. A. Tropp. "Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions". In: *SIAM review* 53 (2011), pp. 217–288.

[28] T. M. Hamill, J. S. Whitaker, and C. Snyder. "Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter". In: *Mon. Wea. Rev.* 129 (2001), pp. 2776–2790.

[29] N. J. Higham. *Functions of matrices: theory and computation*. Vol. 104. Siam, 2008, p. 450.

[30] P. L. Houtekamer and H. L. Mitchell. "A sequential ensemble Kalman filter for atmospheric data assimilation". In: *Mon. Wea. Rev.* 129 (2001), pp. 123–137.

[31] P. L. Houtekamer and H. L. Mitchell. "Data assimilation using an ensemble Kalman filter technique". In: *Mon. Wea. Rev.* 126 (1998), pp. 796–811.

[32] B. R. Hunt, E. J. Kostelich, and I. Szunyogh. "Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter". In: *Physica D* 230 (2007), pp. 112–126.

# References III


[33] T. Janjić et al. “On the representation error in data assimilation”. In: *Q. J. R. Meteorol. Soc.* 144 (2018), pp. 1257–1278.

[34] E. T. Jaynes. *Probability theory: The logic of science*. Cambridge university press, 2003, p. 753.

[35] E. Kalnay. *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, Cambridge, 2002, p. 357.

[36] A. Kong, J. S. Liu, and W. H. Wong. “Sequential imputations and Bayesian missing data problems”. In: *Journal of the American statistical association* 89 (1994), pp. 278–288.

[37] P. Laloyaux et al. “Towards an unbiased stratospheric analysis”. In: *Q. J. R. Meteorol. Soc.* 146 (2020), pp. 2392–2409.

[38] W. G. Lawson and J. A. Hansen. “Implications of Stochastic and Determinisitic Filters as Ensemble-Based Data Assimilation Methods in Varying Regimes of Error Growth”. In: *Mon. Wea. Rev.* 132 (2004), pp. 1966–1981.

[39] D. M. Livings, S. L. Dance, and N. K. Nichols. “Unbiased ensemble square root filters”. In: *Physica D* 237 (2008), pp. 1021–1028.

[40] A. C. Lorenc. “The potential of the ensemble Kalman filter for NWP - a comparison with 4D-Var”. In: *Q. J. R. Meteorol. Soc.* 129 (2003), pp. 3183–3203.

[41] E. N. Lorenz and K. A. Emanuel. “Optimal sites for supplementary weather observations: simulation with a small model”. In: *J. Atmos. Sci.* 55 (1998), pp. 399–414.

[42] E. Ott et al. “A local ensemble Kalman filter for atmospheric data assimilation”. In: *Tellus A* 56 (2004), pp. 415–428.

[43] S. G. Penny and T. Miyoshi. “A local particle filter for high dimensional geophysical systems”. In: *Nonlin. Processes Geophys.* 23 (2016), pp. 391–405.

[44] J. Poterjoy. “A localized particle filter for high-dimensional nonlinear systems”. In: *Mon. Wea. Rev.* 144 (2016), pp. 59–76.

[45] P. N. Raanes, M. Bocquet, and A. Carrassi. “Adaptive covariance inflation in the ensemble Kalman filter by Gaussian scale mixtures”. In: *Q. J. R. Meteorol. Soc.* 145 (2019), pp. 53–75. eprint: `arXiv:1801.08474`.

[46] S. Reich. “A nonparametric ensemble transform method for Bayesian inference.”. In: *SIAM J. Sci. Comput.* 35 (2013), A2013–A2014.

[47] S. Reich and C. Cotter. *Probabilistic Forecasting and Bayesian Data Assimilation*. Cambridge University Press, 2015, p. 306.

[48] P. Sakov and L. Bertino. “Relation between two common localisation methods for the EnKF”. In: *Comput. Geosci.* 15 (2011), pp. 225–237.

[49] P. Sakov and M. Bocquet. “Asynchronous data assimilation with the EnKF in presence of additive model error”. In: *Tellus A* 70 (2018), p. 1414545.

# References IV

[50] P. Sakov, G. Evensen, and L. Bertino. "Asynchronous data assimilation with the EnKF". In: *Tellus A* 62 (2010), pp. 24–29.

[51] P. Sakov and P. R. Oke. "A deterministic formulation of the ensemble Kalman filter: an alternative to ensemble square root filters". In: *Tellus A* 60 (2008), pp. 361–371.

[52] P. Sakov and P. R. Oke. "Implications of the Form of the Ensemble Transformation in the Ensemble Square Root Filters". In: *Mon. Wea. Rev.* 136 (2008), pp. 1042–1053.

[53] C. Snyder et al. "Obstacles to High-Dimensional Particle Filtering". In: *Mon. Wea. Rev.* 136 (2008), pp. 4629–4640.

[54] Y. Trémolet. "Accounting for an imperfect model in 4D-Var". In: *Q. J. R. Meteorol. Soc.* 132 (2006), pp. 2483–2504.

[55] J. S. Whitaker and T. M. Hamill. "Ensemble Data Assimilation without Perturbed Observations". In: *Mon. Wea. Rev.* 130 (2002), pp. 1913–1924.