Tutorial on data assimilation

Marc Bocquet

CEREA, joint lab École des Ponts ParisTech and EdF R&D, Université Paris-Est, France
Institut Pierre-Simon Laplace

(marc.bocquet@enpc.fr)

Outline

1. Data assimilation: principles
   - Definition
   - Mathematical framework
2. Main techniques
   - 3D-Var and optimal interpolation
   - The Kalman filter
   - The ensemble Kalman filter
   - 4D-Var
3. Advanced techniques
   - Ensemble variational methods
   - Particle filters
4. Uncertainty quantification of the best estimate

Data assimilation: definition

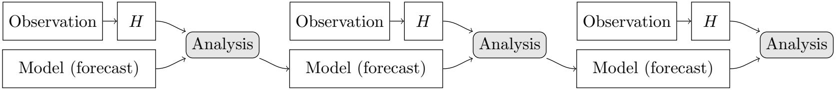
▶ Data assimilation is the set of techniques that allow to optimally combine observations of a physical system with numerical models and prior information of that system, so as to get an estimate of the state of the system.

▶ In the geosciences: Numerical models are often computationally costly. They are often dynamical.

▶ In the geosciences: The state space and observations space are huge (up to $10^9/10^7$ for operational systems, up to $10^7/10^5$ for research systems). A big data problem with costly dynamical models.

▶ What for?: estimate initial state of chaotic systems for forecasting, re-analysis, estimate parameters ($\sim$ inverse modelling).

▶ Example: Data assimilation for prediction.

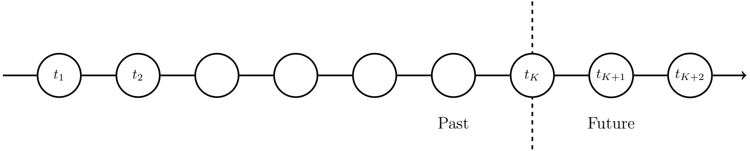


Data assimilation system

- Data assimilation system = observation and evolution models + statistics of the errors. Typically:

$$\mathbf{x}_k = M_{k:k-1}(\mathbf{x}_{k-1}) + \eta_k$$
$$\mathbf{y}_k = H_k(\mathbf{x}_k) + \varepsilon_k$$

with $\eta_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_k)$ and $\varepsilon_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k)$.



- Denoting $\mathbf{x}_{K:1} = \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_K$, $\mathbf{y}_{K:1} = \mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_K$:
  - Prediction: Estimate $\mathbf{x}_k$ for $k > K$, knowing $\mathbf{y}_{K:1}$;
  - Filtering: Estimate $\mathbf{x}_K$, knowing $\mathbf{y}_{K:1}$;
  - Smoothing: Estimate $\mathbf{x}_{K:1}$, knowing $\mathbf{y}_{K:1}$.

The ideal mathematical framework

- Bayes/Laplace approach:

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

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 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 et al., 2003]. Most DA methods can be derived or comply with Bayes' rule.

- But it does not lend to a closed form analytically tractable solution.

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.

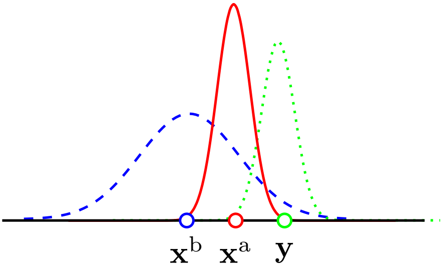
▶ The elementary building block of DA schemes is the statistical BLUE (for Best Linear Unbiased Estimator) analysis. Time is considered fixed. **H** is assumed linear.

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \varepsilon^{\mathrm{o}} \qquad \mathbf{x} = \mathbf{x}^{\mathrm{b}} + \varepsilon^{\mathrm{b}}$$

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

▶ Solution:

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



Outline

3D-Var and optimal interpolation

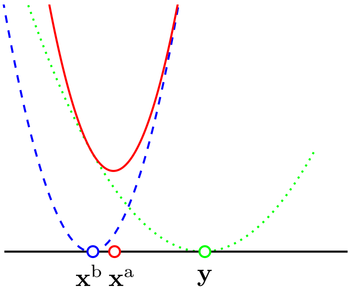
- Variational formulation of the same problem

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

where $\|\mathbf{x}\|^2_{\mathbf{A}} = \mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}$, which is equivalent to BLUE.

- Probabilistic/Bayesian interpretation:

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



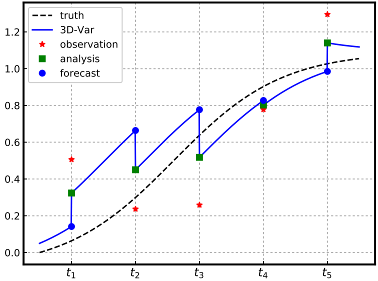
- Capable of handling nonlinear observation operator using standard nonlinear optimisation methods:

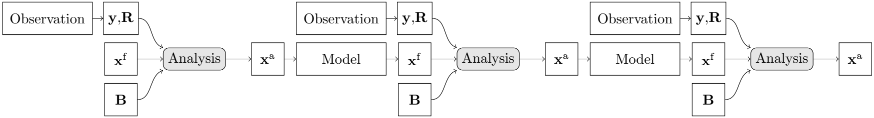
$$J(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}-\mathbf{x}^{\mathbf{b}}\|^2_{\mathbf{B}^{-1}} + \frac{1}{2}\|\mathbf{y}-H(\mathbf{x})\|^2_{\mathbf{R}^{-1}}.$$

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}: $\mathbf{x}^{\mathrm{f}}_{k+1} = M_{k+1:k}(\mathbf{x}^{\mathrm{a}}_k)$.

- 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:

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

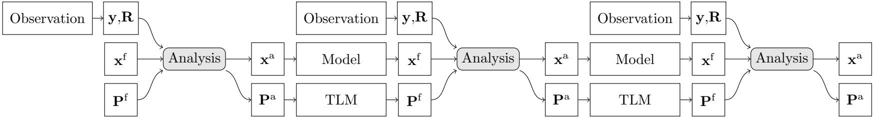
$$\mathbf{K}_k = \mathbf{P}^{\mathrm{f}}_k \mathbf{H}^{\mathrm{T}}_k \left(\mathbf{R}_k + \mathbf{H}_k \mathbf{P}^{\mathrm{f}} \mathbf{H}^{\mathrm{T}}_k \right)^{-1},$$

$$\mathbf{P}^{\mathrm{a}}_k = \left(\mathbf{I} - \mathbf{K}_k \mathbf{H}_k \right) \mathbf{P}^{\mathrm{f}}_k.$$

- Forecast step:

$$\mathbf{x}^{\mathrm{f}}_{k+1} = \mathbf{M}_{k+1:k} \mathbf{x}^{\mathrm{a}}_k,$$

$$\mathbf{P}^{\mathrm{f}}_{k+1} = \mathbf{M}_{k+1:k} \mathbf{P}^{\mathrm{a}}_k \mathbf{M}^{\mathrm{T}}_{k+1:k} + \mathbf{Q}_{k+1}.$$



The extended Kalman filter

- Optimal if the models are linear and if all the initial and observations errors are Gaussian: it gives the perfect Gaussian solution of Bayes' rule.

- Can be extended to nonlinear models: then

$$\mathbf{x}^{\mathrm{f}}_{k+1} = M_{k+1:k}(\mathbf{x}^{\mathrm{a}}_{k}),$$
$$\mathbf{P}^{\mathrm{f}}_{k+1} = \mathbf{M}'_{k+1:k}\mathbf{P}^{\mathrm{f}}_{k}\mathbf{M}'^{\mathrm{T}}_{k+1:k} + \mathbf{Q}_{k+1},$$

where $\mathbf{M}'_{k+1:k}$ is the tangent linear model.

- Extremely costly for large geophysical models: storage space (storage of $\mathbf{P}^{\mathrm{f}}$) and computations ($\mathbf{M}'_{k+1:k}\mathbf{P}^{\mathrm{f}}_{k}\mathbf{M}'^{\mathrm{T}}_{k+1:k}$ requires $2n$ integrations of the model).

The ensemble Kalman filter

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

$$\mathbf{x}^{\mathrm{f}} = \frac{1}{m}\sum_{i=1}^{m} \overline{\mathbf{x}}, \qquad \mathbf{P}^{\mathrm{f}} = \frac{1}{m-1}\sum_{i=1}^{m}\left(\mathbf{x}^{(i)} - \overline{\mathbf{x}}\right)\left(\mathbf{x}^{(i)} - \overline{\mathbf{x}}\right)^{\mathrm{T}}.$$

- Analysis step: Similar to the KF but $\mathbf{P}^{\mathrm{f}}$ explicitly or implicitly taken as the sample covariance estimator.

- Forecast step: The ensemble is propagated using the full nonlinear model (not the tangent linear model!)

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

The ensemble Kalman filter

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

▶ The stochastic EnKF is the closest to traditional KF, but adds stochastic perturbations to the observations of each members to properly account for the observation errors [Burgers et al., 1998]:

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

▶ The deterministic EnKF avoids the introduction of the stochastic perturbations by updating the square root of $\mathbf{P}^{\mathrm{f}} = \mathbf{X}_{\mathrm{f}}\mathbf{X}_{\mathrm{f}}^{\mathrm{T}}$, i.e. $\mathbf{X}_{\mathrm{f}}$. One of the variant (ETKF, [Hunt et al., 2007]) operates the linear algebra in the space of the perturbations:

$$\mathbf{x}^{\mathrm{a}} = \mathbf{x}^{\mathrm{f}} + \mathbf{X}_{\mathrm{f}}\mathbf{w}^{\mathrm{a}}.$$

The analysis in the perturbation space is given by

$$\mathbf{w}^{\mathrm{a}} = \left(\mathbf{I}_m + \mathbf{Y}_{\mathrm{f}}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{Y}_{\mathrm{f}}\right)^{-1}\mathbf{Y}_{\mathrm{f}}^{\mathrm{T}}\mathbf{R}^{-1}\left(\mathbf{y} - \mathbf{H}\mathbf{x}^{\mathrm{f}}\right)$$

where $\mathbf{Y}_{\mathrm{f}} = \mathbf{H}\mathbf{x}^{\mathrm{f}}$. This updates the mean state via $\mathbf{x}^{\mathrm{a}} = \mathbf{x}^{\mathrm{f}} + \mathbf{X}_{\mathrm{f}}\mathbf{w}^{\mathrm{a}}$. The perturbations around it are updated via

$$\mathbf{X}_{\mathrm{a}} = \mathbf{X}_{\mathrm{f}}\left(\mathbf{I}_m - \mathbf{Y}_{\mathrm{f}}^{\mathrm{T}}(\mathbf{Y}_{\mathrm{f}}\mathbf{Y}_{\mathrm{f}}^{\mathrm{T}} + \mathbf{R})^{-1}\mathbf{Y}_{\mathrm{f}}\right)^{\frac{1}{2}}\mathbf{U}, \quad \text{where} \quad \mathbf{U} \in \mathrm{O}(N) \quad \text{and} \quad \mathbf{U}\mathbf{1} = \mathbf{1}.$$

The downside of the EnKF: rank-deficiency

▶ There is a heavy price to pay for replacing the $\mathbf{P}^{\mathrm{f}}$ $n \times n$ covariance matrix of the KF with the $\mathbf{X}_{\mathrm{f}}$ $m \times n$ anomaly matrix: spurious correlations for distant state components. If $\mathbf{P} = \mathbf{X}_{\mathrm{f}}\mathbf{X}_{\mathrm{f}}^{\mathrm{T}}$ and $\mathbf{B}$ is the true error covariance matrix of a Gaussian process:

$$\mathrm{Cov}\left([\mathbf{P}]_{ii},[\mathbf{P}]_{jj}\right) = \frac{2}{N-1}[\mathbf{B}]_{ij}^2, \qquad \mathrm{Cov}\left([\mathbf{P}]_{ij},[\mathbf{P}]_{ij}\right) = \frac{1}{N-1}\left([\mathbf{B}]_{ij}^2 + [\mathbf{B}]_{ii}[\mathbf{B}]_{jj}\right).$$

▶ But, for geophysical systems, we know that most long-range correlations are dampened exponentially. Consequently, the covariances are misestimated (too low variances, too high long-range covariances) and leads to divergence of the EnKF.
$\longrightarrow$ Practically, this is solved using two fixes: inflation and localisation.

▶ Inflation consists in inflating the covariances by a scalar in the hope to compensate for the underestimation of the error statistics [Pham et al., 1998, Anderson et al., 1999]:

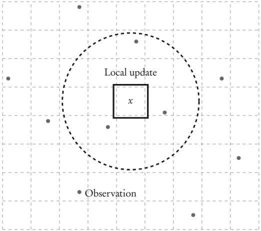
$$\mathbf{x}_{(i)} \longleftarrow \mathbf{x}_{(i)} + \lambda\left(\mathbf{x}_{(i)} - \overline{\mathbf{x}}\right).$$

Can be avoided in a perfect-model context: finite-size EnKF (EnKF-N) [Bocquet et al., 2011-2018].

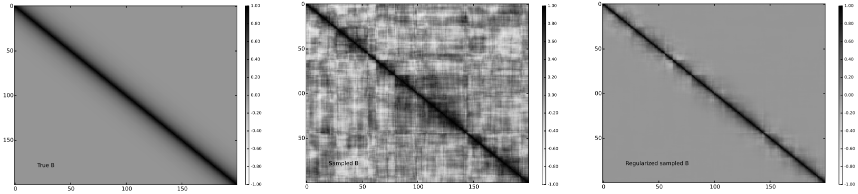
Localisation

- Two flavors of localisation: domain localisation and covariance localisation.

- Domain localisation: divide and conquer.
  The DA analysis is performed in parallel in local domains. The outcomes of these analyses are later sewed together. This is applicable only if the long-range error correlations are negligible.

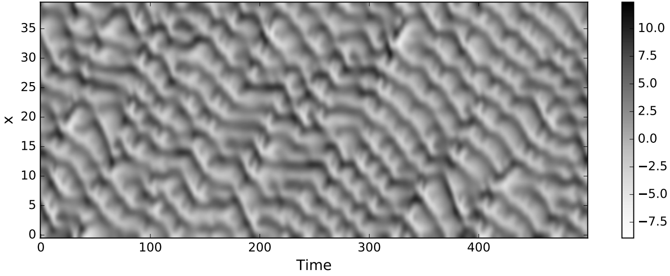


- Covariance localisation: killing off spurious correlation explicitly: $\mathbf{P}^{\mathrm{f}} = \rho \circ \left(\mathbf{X}_{\mathrm{f}}\mathbf{X}_{\mathrm{f}}^{\mathrm{T}}\right)$.



- These strategies have successfully been applied to the EnKF [Hamill et al, 2001; Houtekamer and Mitchell, 2001; Evensen, 2003; Hunt et al., 2007].

Nonlinear chaotic models: the Lorenz-95 low-order model



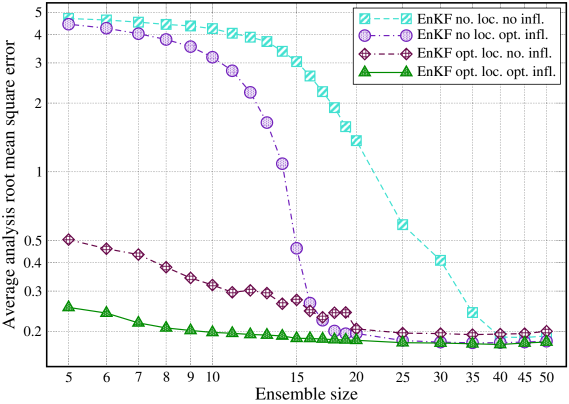
- It represents a mid-latitude zonal circle of the global atmosphere.
- Set of $n = 40$ ordinary differential equations [Lorenz and Emmanuel 1998]:

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = (x_{i+1} - x_{i-2})x_{i-1} - x_i + F\,, \tag{1}$$

where $F = 8$, and the boundary is cyclic.

- Conservative system except for a forcing term F and a dissipation term $-x_i$.
- Chaotic dynamics, 13 positive and 1 neutral Lyapunov exponents, a doubling time of about 0.42 time units.

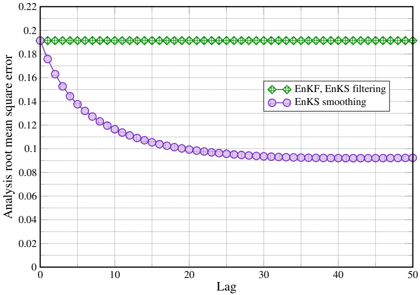
Illustration with the Lorenz-95 model



- Performance of the EnKF in the absence/presence of inflation/localisation.

What about smoothing?

- There are smoothing variants of the Kalman filter [Anderson & Moore, 1979], the Kalman smoother used in the geosciences [Cohn et al., 1994]
- And they have been adapted to the EnKF and variants [Evensen & van Leeuwen, 2000], [Evensen, 2009], [Cosme et al., 2012], [Bocquet & Sakov, 2014], etc.



- Sometimes called asynchronous data assimilation [Sakov et al., 2010; Sakov & Bocquet, 2018].
- With the notable exception of the IEnKS, these smoothers relies on Gaussian assumptions within the DAW.
- 4D-Var is a more natural method to handle nonlinearity within the DAW.

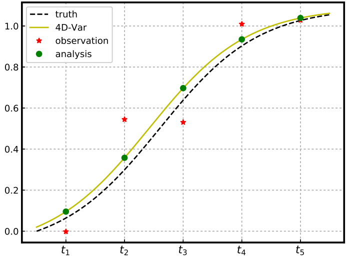
4D-Var

- Strongly constrained 4D-Var, i.e. assuming the model is perfect

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

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

- Fits a model trajectory through the 4D data points.
- In high-dimensional spaces, requires $\nabla_{\mathbf{x}_0} J$ for an efficient minimisation. But $\nabla_{\mathbf{x}_0} J$ depends on the adjoint of $M_{k+1:k}$ and H_k. This can be a very difficult technical task if the model is a huge piece of code for a nonlinear high-dimensional model.



- Weakly constrained 4D-Var, i.e. assuming the model is imperfect

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

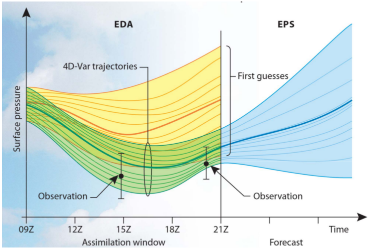
Outline

Hybridising ensemble and variational methods

- ▶ Hybrid: Use flow-dependent statistics from an EnKF into 3D-Var [Hamil & Snyder 2000; Wang et al. 2007].

- ▶ 4D-LETKF [Hunt et al., 2004; Fertig et al., 2007]

- ▶ EDA: ensemble of 4D-Var (ECMWF, Météo-France) [Raynaud et al., 2009; Bonavita et al., 2012; Berre et al., 2015; Jardak & Talagrand 2018]



- ▶ 4DEnVar: Adjoint-less 4D-Var [Liu et al., 2008; Buehner et al. 2010; Zhang and Zhang, 2012; Fairbairn et al. 2014, Desroziers et al. 2014], but ensemble update and nonlinearity still not completely addressed.

- ▶ IEnKS: has it all [Sakov et al. 2012, Bocquet & Sakov 2012-2016].

▶ As ensemble methods, they all require localisation, which is more difficult to implement in a 4D context [Bocquet, 2016] except if the adjoint is available.

⟶ For a review on EnVar methods, see Chapter 7 of the new book [Asch et al., 2016].

The iterative ensemble Kalman smoother (IEnKS)

- Reduced scheme in ensemble space, $\mathbf{x}_0 = \overline{\mathbf{x}}_0 + \mathbf{X}_0\mathbf{w}$, where $\mathbf{X}_0$ is the ensemble anomaly matrix:

$$\widetilde{J}(\mathbf{w}) = J(\overline{\mathbf{x}}_0 + \mathbf{X}_0\mathbf{w}) .$$

- Analysis IEnKS cost function in ensemble space:

$$\widetilde{J}(\mathbf{w}) = \frac{1}{2}\sum_{k=1}^{L} \left\| \mathbf{y}_k - H_k \circ M_{k:0}\left(\overline{\mathbf{x}}_0 + \mathbf{X}_0\mathbf{w}\right) \right\|^2_{\beta_k \mathbf{R}_k^{-1}} + \frac{1}{2}(N-1)\|\mathbf{w}\|^2 .$$

$\{\beta_0, \beta_1, \ldots, \beta_L\}$ weight the observations impact within the window.

- As a variational reduced method, one can use Gauss-Newton [Sakov et al., 2012], Levenberg-Marquardt [Bocquet & Sakov, 2012], quasi-Newton, trust region, etc., minimisation schemes.

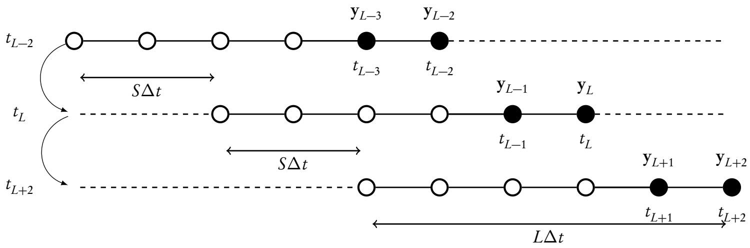
- Perturbation update: same as the ETKF

$$\mathbf{E}_0^\star = \mathbf{x}_0^\star \mathbf{1}^{\mathrm{T}} + \sqrt{N-1}\mathbf{X}_0 \left[\nabla^2_{\mathbf{w}}\widetilde{J}\right]_\star^{-1/2}\mathbf{U} \quad \text{where} \quad \mathbf{U} \in \mathrm{O}(N) \quad \text{and} \quad \mathbf{U}\mathbf{1} = \mathbf{1} .$$

⟶ Cécile Defforge's talk.

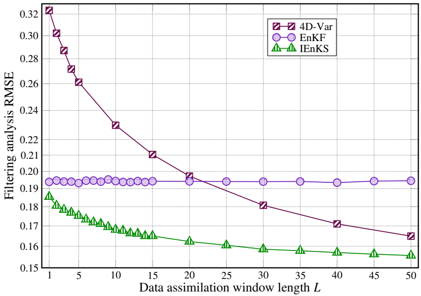
Chaining 4D analyses in time

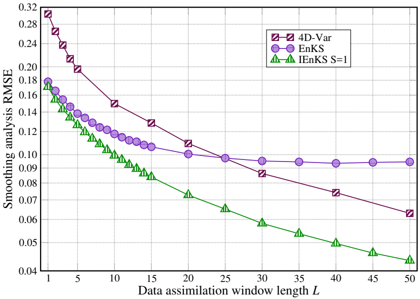
- The IEnKS opens up new perspectives on the chaining of DA cycles which was little relevant for either the EnKF or 4D-Var.
- L: length of the data assimilation window,
- S: shift of the data assimilation window in between two updates.



Variational analysis in ens. space $\rightarrow$ Posterior ens. generation $\rightarrow$ Ens. forecast

Performance comparison with Lorenz-95





- Comparing 4D-Var, the EnKF, the EnKS and the IEnKS.

Taking the bull by the horns: the particle filter

▶ The particle filter is the Monte-Carlo solution of the Bayes' equation. This is a sequential Monte Carlo method.

▶ The most simple algorithm of Monte Carlo type that solves the Bayesian filtering equations is called the bootstrap particle filter [Gordon et al. 1993] .

Sampling: Particles $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_M\}$.
Pdf at time t_k: $p_k(\mathbf{x}) \simeq \sum_{i=1}^{M} \omega_i^k \delta(\mathbf{x} - \mathbf{x}_k^i)$.

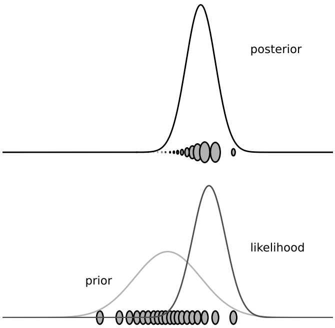
Forecast: Particles propagated by

$$p_{k+1}(\mathbf{x}) \simeq \sum_{i=1}^{M} \omega_k^i \delta(\mathbf{x} - \mathbf{x}_{k+1}^i)$$

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

Analysis: Weights updated according to

$$\omega_{k+1}^{\mathrm{a},i} \propto \omega_{k+1}^{\mathrm{f},i} p(\mathbf{y}_{k+1} | \mathbf{x}_{k+1}^i) .$$



▶ Analysis is carried out with only a few multiplications. No matrix inversion!

Taking the bull by the horns: the particle filter

▶ These normalised statistical weights have a potentially large amplitude of fluctuation. One particle (one trajectory of the model) will stand out among the others. Its weight will largely dominate the others ($\omega_i \lesssim 1$). Then the particle filter becomes very inefficient as an estimating tool since it has lost its variability. This phenomenon is called degeneracy of the particle filter [Kong et al. 1994].

Resampling One way to mitigate this phenomenon is to resample the particles by redrawing a sample with uniform weights from the degenerate distribution. After resampling, all particles have the same weight: $\omega_k^i = 1/M$.

▶ Handles very well, very nonlinear low-dimensional systems. But, without modification, very inefficient for high-dimensional models. Avoiding degeneracy requires a great number of particles that scales exponentially with the size of the system. This is a manifestation of the curse of dimensionality.

Application of the particle filter in the geosciences

- ▶ The applicability of particle filters to high-dimensional models has been investigated in the geosciences [van Leeuwen, 2009; Bocquet, 2010]. The impact of the curse of dimensionality has been quantitatively studied in [Snyder et al., 2008]. It was known [Mackay et al., 2003] that using an importance proposal to guide the particles towards regions of high probability will not change this trend, albeit with a reduced exponential scaling, which was confirmed by [Snyder et al., 2015]: optimal importance sampling particle filter [Doucet et al., 2000; Bocquet, 2010; Snyder; 2011].

- ▶ Particle smoother over a data assimilation window, alternative and more efficient particle filters can be designed, such as the implicit particle filter [Morzfeld et al., 2012].

- ▶ Particle filters can nevertheless be useful for high-dimensional models if the significant degrees of nonlinearity are confined to a small subspace of the state space, e.g. Lagrangian data assimilation [Slivinski et al., 2015] .

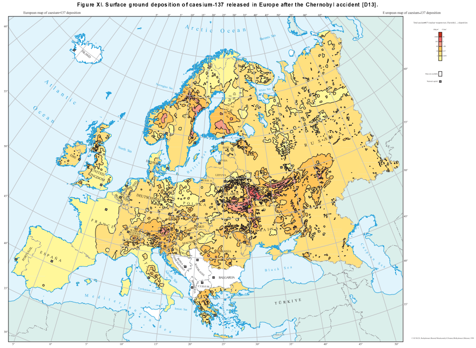
- ▶ It is possible possible to design nonlinear filters for high-dimensional models such as the equal-weight particle filter [van Leeuwen & Ades, 2010-2017].

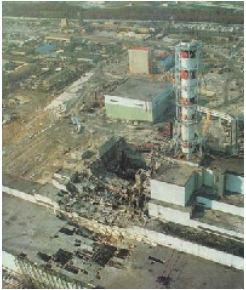
- ▶ Localisation can be (should be?) used in conjunction with the particle filter [Reich et al. 2013; Potterjoy, 2016; Penny & Miyoshi, 2016; Farchi & Bocquet, 2018].
  $\longrightarrow$ Alban Farchi's talk.

- ▶ It has been applied in hydrology, nivology, climate, etc [Goosse, Dubinkina et al.].

Outline

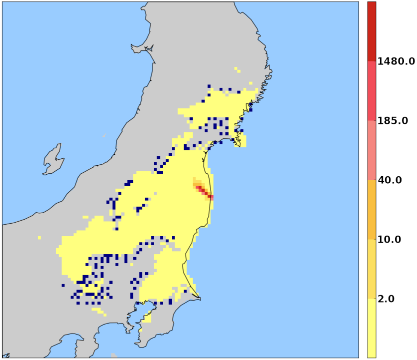
Case study: Chernobyl and Fukushima accidents



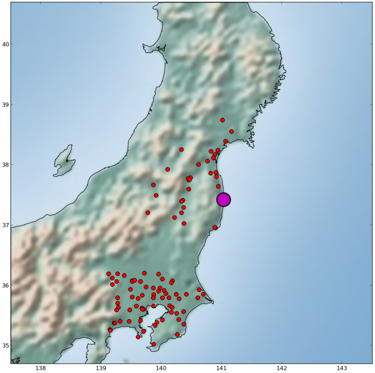


- 30 deaths in the first days of the accident
- 200 000 evacuees
- 30 km exclusion zone
- Mid and long term sanitary impact: thyroid cancer (thousands of cases).

Case study: Chernobyl and Fukushima accidents

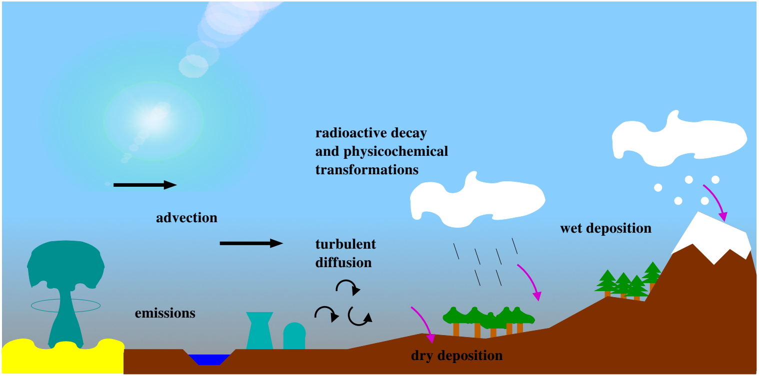


Caesium-137 deposition [IRSN database]



Air quality monitoring network

Case study: Chernobyl and Fukushima accidents



- ▶ Modelled by PDEs of the transport processes and physical and chemical parametrisations.
- ▶ Source term usually unknown.
- ▶ Parameters of the physical parametrisations often poorly know (effective turbulent diffusion, scavenging and dry deposition parameters).

Case study: Cost function

- Source-receptor relationship: **H**. Linear model.
- Problem usually solved using 4D-Var [Bocquet, 2012] or methods equivalent to the representer technique. Here, study focused on UQ of the *best* estimate [Liu et al., 2017].
- Log-normal errors for the prior and for the observations. Non-Gaussian statistics.
- Cost function from Bayes' rule:

$$\begin{aligned}\mathscr{L}(\mathbf{z};\theta) &= -\ln p(\mathbf{z}|\mathbf{y},\theta) = -\ln p(\mathbf{y}|\mathbf{z},\theta) - \ln p(\mathbf{z}|\theta) + \ln p(\mathbf{y}|\theta) \\ &= \frac{1}{2}\left\|\ln \mathbf{y} - \ln \mathbf{H}\overline{\mathbf{x}}e^{\mathbf{z}}\right\|^2_{\mathbf{R}^{-1}} + \frac{1}{2}\left\|\mathbf{z}\right\|^2_{\mathbf{B}^{-1}} + \frac{1}{2}\ln|\mathbf{R}| + \frac{1}{2}\ln|\mathbf{B}| + \xi.\end{aligned}$$

- Two strategies to quantify the uncertainty of the best estimate:
  - Bayesian hierarchy (HB):

$$p(\mathbf{x},\theta|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x},\theta)p(\mathbf{x}|\theta)p(\theta)}{p(\mathbf{y})}, \qquad p(\mathbf{x}|\mathbf{y}) = \int \mathrm{d}\theta\, p(\mathbf{x},\theta|\mathbf{y}). \tag{2}$$

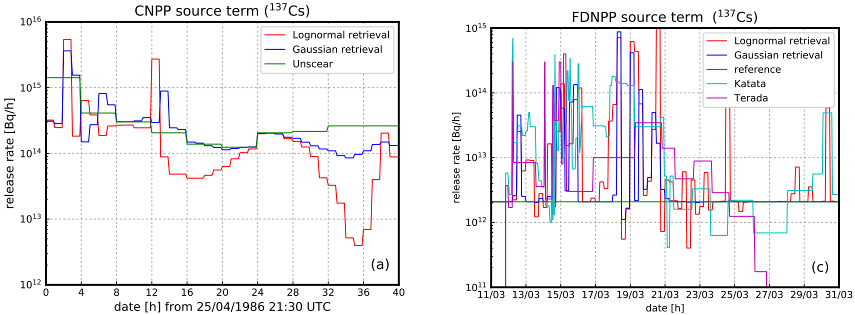
  - Empirical Bayes (EB):

$$p(\mathbf{x}|\mathbf{y}) \approx p(\mathbf{x}|\mathbf{y},\theta^\star). \tag{3}$$

    $\theta^\star$ here estimated by the Expectation-Maximisation (EM) algorithm.

Case study: Inversions (EB)

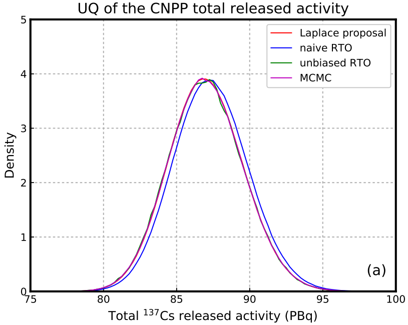
- Uniform hyperparameters: $\mathbf{R} = r^\star \mathbf{I}$, $\mathbf{B} = b^\star \mathbf{I}$, where $r^\star$ and $b^\star$ are obtained from EM.



- Chernobyl and Fukushima-Daiichi source terms with Gaussian and lognormal assumptions on the observation errors. Comparison with the Unscear reference source term.

Case study: UQ of the retrieved total radioactivity (EB)

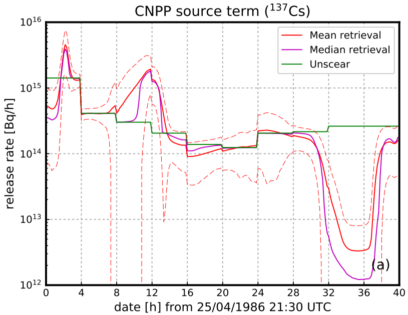
▶ Probability density function of the total released activity for Chernobyl and Fukushima-Daiichi.

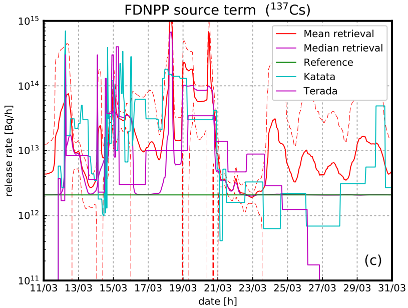


▶ EB: optimal hyperparameters are first determined. Followed by nonlinear sampling of the total activity using three methods: with a Laplace proposal, a random-then-optimise sampling, an unbiased random-then-optimise sampling and a basic MCMC.

Case study: Inversion (HB)

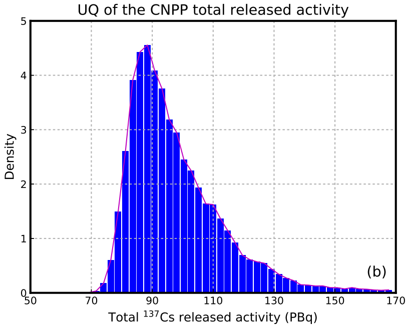
- ▶ Full solution of the Bayesian hierarchy (HB)
- ▶ Obtained from a Monte Carlo Markov Chain (MCMC)
- ▶ Transdimensional analysis (adaptive grid). Here using only 20 grid cells for Chernobyl and 40 grid cells for Fukushima.

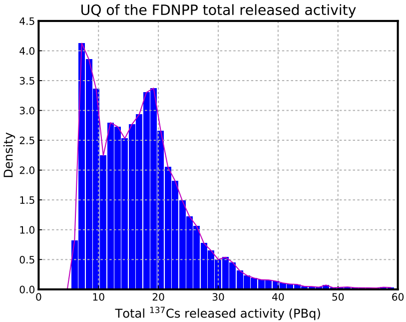




Case study: UQ of the retrieved total radioactivity (HB)

▶ Probability density function of the total released activity for Chernobyl and Fukushima-Daiichi.





▶ Full solution of the Bayesian hierarchy; obtained from an MCMC.

▶ Transdimensional analysis (adaptive grid). Here using only 20 grid cells for Chernobyl and 40 grid cells for Fukushima.

Final word

Thank you for your attention!

- Part I: A gentle introduction to DA.
- Part II: More advanced topics including EnKF and EnVar.
- Part III: Applications of DA including emerging ones such as: glaciology, biology, geomagnetism, medicine, imaging and acoustics, economics and finance, traffic control, etc.