Data assimilation in the geosciences
An overview

Alberto Carrassi[1], Olivier Talagrand[2], Marc Bocquet[3]

(1) NERSC, Bergen, Norway
(2) LMD, École Normale Supérieure, IPSL, France
(3) CEREA, joint lab École des Ponts ParisTech and EdF R&D, IPSL, France

Outline of part III

1. Making EnKF work: localisation and inflation
   - Diagnostic
   - Localisation
   - Inflation
   - Why they are necessary
2. Ensemble variational methods
   - Hybrid approaches
   - Ensemble of data assimilation
   - 4D-En-Var
   - Iterative ensemble Kalman smoother
   - Numerical comparison
3. Particle filters
   - Principle
   - Can they be useful in the geosciences?

Outline

1. Making EnKF work: localisation and inflation
   - Diagnostic
   - Localisation
   - Inflation
   - Why they are necessary
2. Ensemble variational methods
   - Hybrid approaches
   - Ensemble of data assimilation
   - 4D-En-Var
   - Iterative ensemble Kalman smoother
   - Numerical comparison
3. Particle filters
   - Principle
   - Can they be useful in the geosciences?

Remedies to make EnKF working in high dimension

▶ Limited number N of anomalies: the sample covariance matrix is highly rank-deficient.

▶ If $\mathbf{B}$ is the true covariance matrix and $\mathbf{P}^{\mathrm{e}}$ is the (N-member) sample covariance matrix which approximates $\mathbf{B}$, then:

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

In most geophysical systems, $[\mathbf{B}]_{ij}$ vanish exponentially with $|i-j| \to \infty$.
The $[\mathbf{B}]_{ii}$ are the variances and remain finite, so that

$$\mathbb{E}\left([\mathbf{P}^{\mathrm{e}} - \mathbf{B}]_{ij}^2\right) \underset{|i-j|\to\infty}{\sim} \frac{1}{N-1}[\mathbf{B}]_{ii}[\mathbf{B}]_{jj}.$$

▶ Since $[\mathbf{B}]_{ij}$ vanish exponentially with the distance, we want $\mathbb{E}\left([\mathbf{P}^{\mathrm{e}} - \mathbf{B}]_{ij}^2\right)$ to also vanish exponentially with the distance. Hence with N finite, the sample covariance $[\mathbf{P}^{\mathrm{e}}]_{ij}$ is potentially a bad approximation especially for large distances $|i-j|$.

▶ The errors of such an approximation are usually referred to as **sampling errors**.

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.

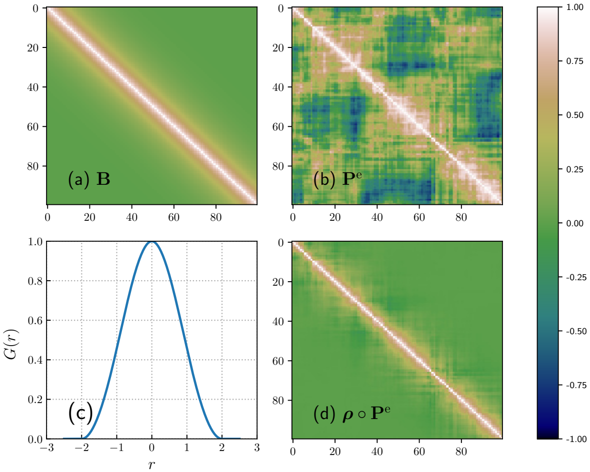
▶ Solution: compute the **Schur product** of $\mathbf{P}^{\mathrm{e}}$ with a well chosen smooth **correlation matrix** $\boldsymbol{\rho}$, that has exponentially vanishing correlations for distant parts.

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

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

▶ 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



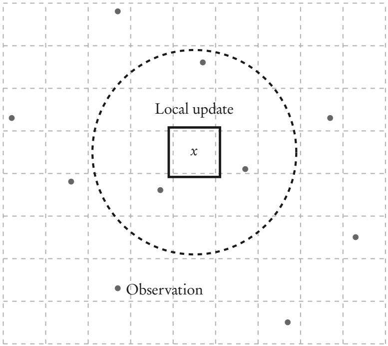
Panel (a): True covariance matrix. Panel (b): Sample covariance matrix.
Panel (c): Gaspari-Cohn correlation matrix used for covariance localisation.
Panel (d): Tapered covariance matrix.

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.



▶ 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 still diverge!

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

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

or, alternatively,

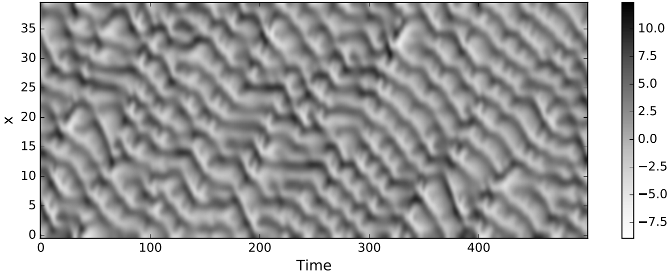
$$\mathbf{x}_{[n]} \longrightarrow \overline{\mathbf{x}} + \lambda \left(\mathbf{x}_{[n]} - \overline{\mathbf{x}} \right) .$$

- Inflation can also come in an **additive form**: $\mathbf{x}_{[n]} \longrightarrow \mathbf{x}_{[n]} + \varepsilon_{[n]}$.

- 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.

Nonlinear chaotic models: the Lorenz-96 low-order model



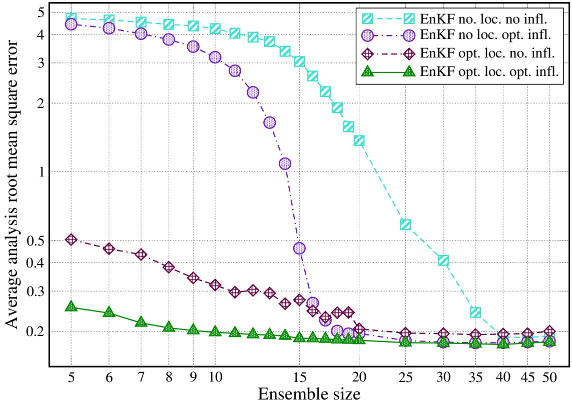
- It represents a mid-latitude zonal circle of the global atmosphere.
- Set of $n = 40$ ordinary differential equations [Lorenz and Emmanuel 1998]:

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

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-96 model



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

Outline

Hybrid of EnKF and Var methods

- Hybrid can refer to a combination of a **variational** method and of an **EnKF** method.

- Yet, it often refers to the **hybridising** of a **static** error covariance matrix with a **dynamical** one sampled from an ensemble [Hamill and Snyder, 2000].

- Example: use the static covariance of a 3DVar and perform a variational analysis to be used for the analysis step of an EnKF with the prior:

$$\mathbf{B} = \alpha \mathbf{C} + (1-\alpha)\mathbf{X}^{\mathrm{f}} \left(\mathbf{X}^{\mathrm{f}}\right)^{\mathrm{T}},$$

where $\mathbf{C}$ is the static error covariance matrix, $\mathbf{X}^{\mathrm{f}}$ is the matrix of the forecast ensemble anomalies, and $\alpha \in [0,1]$ is a scalar that weights the **static** and **dynamical** contributions.

- **Updated ensemble** easily obtained in the framework of the stochastic EnKF. More difficult to obtain in the framework of deterministic EnKFs [Sakov and Bertino, 2011; Bocquet et al, 2015; Auligné et al., 2016]

- With an EnKF based on localisation:

$$\mathbf{B} = \alpha \mathbf{C} + (1-\alpha)\boldsymbol{\rho} \circ \left[\mathbf{X}^{\mathrm{f}} \left(\mathbf{X}^{\mathrm{f}}\right)^{\mathrm{T}}\right].$$

Ensemble of data assimilation: EDA

▶ Methods known as **ensemble of data assimilations** (EDA) perform an ensemble of analysis based on a given method and perturbed errors (observation and background): e.g., En-3D-Var, En-4D-var, En-4D-En-Var. **Mimics the stochastic EnKF.**

▶ Stem from NWP centres (Météo-France and ECMWF) that operate 4D-Var and, by offering an ensemble of analysis and forecast, **can generate time-dependent error statistics.** [Raynaud et al., 2009-2012; Bonavita et al., 2011-2012]

▶ Each analysis, indexed by i, uses a different first guess $\mathbf{x}_0^i$, and observations perturbed with $\varepsilon_k^i \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k)$ to maintain **statistical consistency**. Hence, each analysis i carries out the minimisation of

$$\mathcal{J}_i^{\mathrm{EDA}}(\mathbf{x}_0) = \frac{1}{2} \sum_{k=0}^{K} \left\| \mathbf{y}_k + \varepsilon_k^i - \mathcal{H}_k \circ \mathcal{M}_{k:0}(\mathbf{x}_0) \right\|_{\mathbf{R}_k^{-1}}^2 + \frac{1}{2} \left\| \mathbf{x}_0 - \mathbf{x}_0^i \right\|_{\mathbf{B}^{-1}}^2 .$$

▶ The background covariance **B** is typically hybrid because it still uses the static covariances of the traditional 4DVar and incorporates the sample covariances from the dynamical perturbations.

Four-dimensional ensemble variational: 4DEnVar

▶ NWP centres operating 4D-Var have difficulties maintaining the adjoint models.

▶ Generalises how the EnKF estimates the sensitivities of the observation to the state variables to the dynamical model over the 4D-Var time window [Liu et al., 2008-2009]:
$\longrightarrow$ an ensemble of N nonlinear trajectories within the 4D-Var window.

▶ The 4D-Var analysis is carried out in the subspace generated by the perturbations [Robert et al., 2005]. **No model adjoint**. This yields **4DEnVar**.

▶ In 4DEnVar, the perturbations are usually generated stochastically, for instance resorting to a stochastic EnKF. Hence, flow-dependent error estimation is introduced.

▶ The prior is often **hybrid** since the method is based on a preceding 4D-Var system.

▶ **Localisation** is needed as in the EnKF. But more difficult problem than in 4D-Var [Bocquet et al., 2016; Desroziers et al., 2016].

▶ Many variants of the 4DEnVar are possible depending on the way the perturbations are generated, or if the adjoint model is available or not [Buehner et al. 2010; Zhang et al., 2012; Poterjoy et al., 2015]. Full 4DEnVar operational systems are now implemented or are in the course of being so [Buehner et al. 2013-2015, Gustafsson et al. 2014; Desroziers et al. 2014, Lorenc et al. 2015, kleist et al. 2015, bowler et al. 2017].

The iterative ensemble Kalman smoother: IEnKS

▶ The IEnKS is derived from Bayes' rule and provides an exemplar of **deterministic nonlinear four-dimensional EnVar method**.

▶ It mimics a deterministic EnKF, in particular the ETKF. The analysis cost function over a time window is:

$$\mathcal{J}(\mathbf{w}) = \frac{1}{2}\left\|\mathbf{w}\right\|^2 + \sum_{k=L-S+1}^{L} \frac{1}{2}\left\|\mathbf{y}_k - \mathcal{H}_k \circ \mathcal{M}_{k:0}\left(\overline{\mathbf{x}}_0 + \mathbf{X}_0\mathbf{w}\right)\right\|^2_{\mathbf{R}_k^{-1}},$$

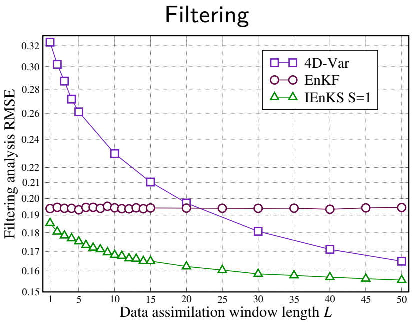
The minimisation can be carried out (i) with or without the adjoint model (ii) using any efficient minimisation technique, usually Gauss-Newton, or Quasi-Newton, Levenberg-Marquardt or trust region.

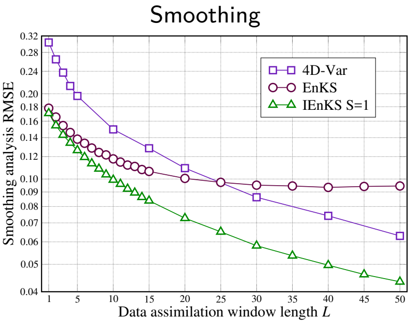
▶ It generates an **updated ensemble consistent with the analysis**, as opposed to current implementations of 4DEnVar.

▶ It allows for **overlapping time-windows** (a technique contemplated by the ECMWF).

▶ But **it does require localisation** (for the exact same reason as 4DEnVar).

Numerical comparison





▶ Measures the gain in accuracy obtained from a fully nonlinear 4D EnVar technique (the IEnKS): combine a nonlinear variational analysis with time-dependent error statistics.

Outline

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].

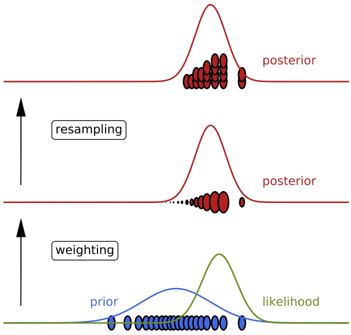
Sampling: Particles $\left\{\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^N\right\}$.
Pdf at time t_k: $p_k(\mathbf{x}_k|\mathbf{y}_k) \simeq \sum_{i=1}^N \omega_i^k \delta(\mathbf{x}_k - \mathbf{x}_k^i)$.
Analysis: Weights updated according to

$$\omega_k^{\mathrm{a},i} \propto \omega_k^{\mathrm{f},i} p(\mathbf{y}_k|\mathbf{x}_k^i)\,.$$

Forecast: Particles propagated by

$$p_{k+1}(\mathbf{x}_{k+1}|\mathbf{y}_k) \simeq \sum_{i=1}^N \omega_k^{\mathrm{a},i} \delta(\mathbf{x}_{k+1} - \mathbf{x}_{k+1}^i)$$

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

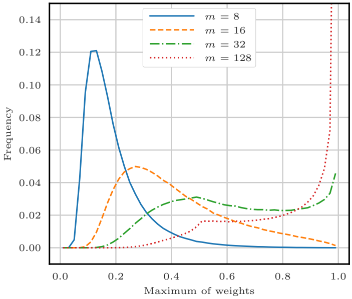


▶ Analysis is carried out with only a few multiplications. **No matrix inversion!**

Taking the bull by the horns: the particle filter

▶ These normalised statistical weights have a potentially large amplitude of fluctuation. One particle will stand out among the others ($\omega_i \lesssim 1$). Then the particle filter becomes very inefficient as an estimating tool.

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: $\omega_k^i = 1/N$.

▶ Handles 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.
$\longrightarrow$ This is a manifestation of the **curse of dimensionality**.

Application of the particle filter in the geosciences

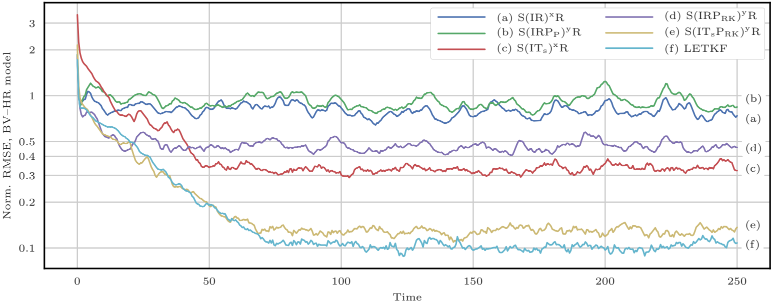
- The applicability of particle filters to high-dimensional models has been investigated in the geosciences [van Leeuwen, 2009; Bocquet, 2010]. The impact of the curse of dimensionality has been quantitatively studied in [Snyder et al., 2008]. It was known [Mackay et al., 2003] that using an importance 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, as in **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].

Application of the particle filter in the geosciences



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

- The particle filter has been applied in hydrology, convection, nivology, climate, etc [Goosse, Dubinkina, Haslehner, van Leeuwen, etc.].

A selection of reviews and textbooks

- **Assimilation of Observations, an Introduction.** Talagrand, O., J. Meteor. Soc. Japan, 75, 191–209, 1997.

- **Atmospheric Modeling, Data Assimilation and Predictability.** E. Kalnay, Cambridge University Press, 2002.

- **Data Assimilation: The Ensemble Kalman Filter.** G. Evensen, G., Springer-Verlag, 2007.

- **Data assimilation: Methods, Algorithms and Applications.** M. Asch, M. Bocquet and M. Nodet, Society for Industrial and Applied Mathematics (SIAM), 2016.

- **Data Assimilation in the Geosciences: An overview on methods, issues and perspectives.** A. Carrassi, M. Bocquet, L. Bertino, and G. Evensen, submitted to WIREs Climate Change, 2018.