Lecture 2 on data assimilation: The ensemble Kalman filter (the algebra of)

Marc Bocquet

With help from Alban Farchi, inspiration from Pavel Sakov

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

(marc.bocquet@enpc.fr)

Synopsis of the course

- Monday, October 28 10:30-12:30
  Lecture 1: Elementary principles of geophysical data assimilation. The Bayesian standpoint. Classical methods of data assimilation: 3D-Var, the Kalman filter, 4D-Var.
- Tuesday, October 29, 10:30-12:30
  Lecture 2: The ensemble Kalman filter and its variants (focus on the algorithmic/mathematical aspects.)
- Thursday, October 31, 10:30-12:30
  Lecture 3: Recent advances: hybrid and ensemble variational techniques. Discussion on what to expect from machine learning/deep learning.

Followed next week by:

- A course on data assimilation and stochastic filtering, particle filters by Dan Crisan (Imperial College, London)
- A course on big data and uncertainty quantification by Omar Ghattas (Uni. of Texas, Austin)

Outline

1. The ensemble Kalman filter
   - Reminder
   - Principles
   - Mathematical prerequisites
   - The ETKF
   - The EnSRF
   - The DEnKF
   - The serial EnKF
2. Making EnKF work: localisation and inflation
   - Diagnostic
   - Localisation
   - Inflation
   - Why they are necessary
   - Hybrid localisation
3. References

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, $\boldsymbol{\eta}_k, \boldsymbol{\epsilon}_k : k = 1, \ldots, K$ are assumed to be uncorrelated in time, mutually independent, and they follow the pdfs $p_{\boldsymbol{\eta}}$ and $p_{\boldsymbol{\epsilon}}$.

Formal sequential Bayesian solution

- 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 which propagates this pdf, using the Chapman-Kolmogorov equation, forward in time until the new observation batch:

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

Sequential Bayesian estimation: the Kalman filter

- Even though these equations are well suited for sequential DA with chaotic models, they are still impractical to solve. However, the Kalman filter solves them exactly under the assumptions of linearity of the models and Gaussianity of the statistics.

- Analysis step:

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

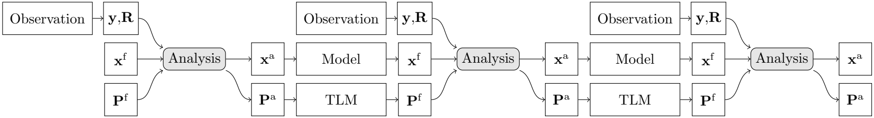
$$\mathbf{K}_k = \mathbf{P}_k^{\mathrm{f}} \mathbf{H}_k^\top \left(\mathbf{R}_k + \mathbf{H}_k \mathbf{P}^{\mathrm{f}} \mathbf{H}_k^\top \right)^{-1},$$

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

- Forecast step:

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

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



The extended Kalman filter

- As seen in lecture 1, the Kalman filter can be extended to handle nonlinear models:

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

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

- Drawbacks 1 & 2: 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).

- Drawback 3: The model linearisation in the error covariances is an approximation.

- Solutions: The reduced-rank / ensemble Kalman filters.

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 $i = 1 \ldots N_\mathrm{e}$:

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

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 $\mathbf{Y}_\mathrm{f} = \mathbf{H}\mathbf{X}_\mathrm{f} + \boldsymbol{\Theta}$, 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_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: What's nice about it?

The ensemble forecast has a complexity of N_e model runs

Yes, it is far better than the extended Kalman filter and game-changing.
But there will be a heavy tribute for this.

The ensemble forecast uses the nonlinear model in place of the tangent linear model

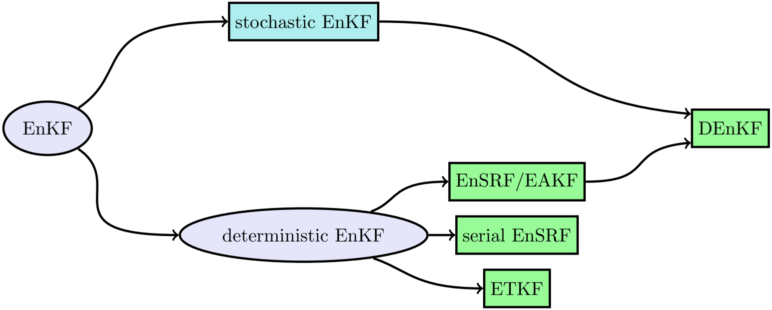
Yes, it's nice and better from a Bayesian standpoint.
But not as critical as it was originally sold. In that respect, the EnKF is outperformed by the iterative ensemble Kalman filter and smoother ($\rightarrow$ lecture 3).

It emulates the tangent linear of the observation model

Definitely a good point and at the origin of nonlinear EnVar techniques ($\rightarrow$ lecture 3).

The ensemble Kalman filter: a bunch of methods

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



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

Key algebraic identities

- Sherman-Morrison-Woodbury (SMW) identity (**A** and **C** invertible):

$$(\mathbf{A}+\mathbf{UCV})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}\left(\mathbf{C}^{-1}+\mathbf{VA}^{-1}\mathbf{U}\right)^{-1}\mathbf{VA}^{-1}.$$

- Typical applications:
  - Analysis error covariances:

$$\mathbf{P}^{\mathrm{a}} = \left(\mathbf{B}^{-1}+\mathbf{H}^{\top}\mathbf{R}^{-1}\mathbf{H}\right)^{-1} = \mathbf{B} - \mathbf{BH}^{\top}\left(\mathbf{R}+\mathbf{HBH}^{\top}\right)^{-1}\mathbf{HB}.$$

  - Kalman gain:

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

Key algebraic identities

▶ Matrix shift lemma (SML): Let **A** and **B** two matrices of compatible dimensions and $x \mapsto f(x)$ be a function defined on the spectra of **AB** and **BA**, then :

$$\mathbf{A} f(\mathbf{BA}) = f(\mathbf{AB})\mathbf{A}.$$

→ Proof in [Higham 2008].

▶ Typical application, $\mathbf{A} \in \mathbb{R}^{N_x \times N_y}$ and $\mathbf{B} \in \mathbb{R}^{N_y \times N_x}$ are positive semi-definite:

$$\mathbf{A}\left(\mathbf{I}_{\mathrm{y}} + \mathbf{BA}\right)^{-1} = \left(\mathbf{I}_{\mathrm{x}} + \mathbf{AB}\right)^{-1}\mathbf{A}.$$

Key algebraic identities

▶ Let f be a function such that $f(0) = 1$, and which is analytic in a connected domain $\mathcal{D}$ of contour $\mathcal{C}$ in the complex plane $\mathbb{C}$ which encloses the eigenvalues of both **AB** and **AB**. Define $g(x) = (f(x) - 1)/x$. Then

$$f(\mathbf{AB}) = \mathbf{I} + \mathbf{A}g(\mathbf{BA})\mathbf{B}.$$

→ Proof in [Higham 2008].

▶ Application: let us assume that the eigenvalues of **AB** and **BA** have a non-negative real part, then

$$(\mathbf{I}_{\mathrm{x}} + \mathbf{AB})^{-\frac{1}{2}} = \mathbf{I}_{\mathrm{x}} - \mathbf{A}\left(\mathbf{I}_{\mathrm{y}} + \mathbf{BA} + \left[\mathbf{I}_{\mathrm{y}} + \mathbf{BA}\right]^{\frac{1}{2}}\right)^{-1}\mathbf{B},$$

where we chose $f(x) = (1+x)^{-\frac{1}{2}}$ and $g(x) = -(1 + x + \sqrt{1+x})^{-1}$.
→ Proof in [Bocquet and Farchi 2019].

Deterministic Kalman filters and matrix square root definition

▶ The deterministic EnKFs avoid the introduction of the stochastic perturbations by updating the anomaly matrix $\mathbf{X}_\mathrm{f}$ in

$$\mathbf{P}^\mathrm{f} = \mathbf{X}_\mathrm{f}\mathbf{X}_\mathrm{f}^\top ,$$

rather than updating $\mathbf{P}^\mathrm{f}$.

▶ In the following, $\mathbf{X}_\mathrm{f}$ is called a factor of $\mathbf{P}^\mathrm{f}$, not a "square root" of $\mathbf{P}^\mathrm{f}$ as sometimes seen in geophysical data assimilation literature. This would clash with the mathematical definition of a square root matrix.

▶ Let $\mathbf{M}$ be a diagonalisable matrix with non-negative eigenvalues, i.e. $\mathbf{M} = \mathbf{G}\mathbf{D}\mathbf{G}^{-1}$, where $\mathbf{G}$ is an invertible matrix and $\mathbf{D}$ is the diagonal matrix containing the non-negative eigenvalues of $\mathbf{M}$. Then the square root of $\mathbf{M}$ is

$$\mathbf{M}^{\frac{1}{2}} = \mathbf{G}\mathbf{D}^{\frac{1}{2}}\mathbf{G}^{-1},$$

where $\mathbf{D}^{\frac{1}{2}}$ is the diagonal matrix with the square root of the eigenvalues of $\mathbf{M}$.

▶ Note that $\mathbf{M}$ does not have to be symmetric.

The ensemble transform Kalman filter: mean update

- One of the variant (ETKF, [Hunt et al. 2007] on an idea by [Bishop, Etherton, et al. 2001]) operates the linear algebra in the space of the perturbations, or ensemble subspace:

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

- Inserting this decomposition into the Kalman state update equation:

$$\mathbf{x}^{\mathrm{f}} + \mathbf{X}_{\mathrm{f}}\mathbf{w}^{\mathrm{a}} = \mathbf{x}^{\mathrm{f}} + \mathbf{X}_{\mathrm{f}}\mathbf{X}_{\mathrm{f}}^{\top}\mathbf{H}^{\top}\left(\mathbf{H}\mathbf{X}_{\mathrm{f}}\mathbf{X}_{\mathrm{f}}^{\top}\mathbf{H}^{\top} + \mathbf{R}\right)^{-1}\delta, \qquad \text{where} \qquad \delta = \mathbf{y} - \mathcal{H}(\mathbf{x}^{\mathrm{f}}),$$

which suggests

$$\mathbf{w}^{\mathrm{a}} \equiv \mathbf{X}_{\mathrm{f}}^{\top}\mathbf{H}^{\top}\left(\mathbf{H}\mathbf{X}_{\mathrm{f}}\mathbf{X}_{\mathrm{f}}^{\top}\mathbf{H}^{\top} + \mathbf{R}\right)^{-1}\delta = \mathbf{Y}_{\mathrm{f}}^{\top}\left(\mathbf{Y}_{\mathrm{f}}\mathbf{Y}_{\mathrm{f}}^{\top} + \mathbf{R}\right)^{-1}\delta.$$

- Using the SMW identity, we finally obtain:

$$\mathbf{w}^{\mathrm{a}} = \left(\mathbf{I}_{\mathrm{e}} + \mathbf{Y}_{\mathrm{f}}^{\top}\mathbf{R}^{-1}\mathbf{Y}_{\mathrm{f}}\right)^{-1}\mathbf{Y}_{\mathrm{f}}^{\top}\mathbf{R}^{-1}\delta.$$

The ensemble transform Kalman filter: perturbations update

▶ From the analysis error covariance matrix of the Kalman filter, let us infer what the analysis anomaly matrix could be:

$$\begin{aligned}\mathbf{P}^{\mathrm{a}} &= (\mathbf{I}_{\mathrm{x}} - \mathbf{K}\mathbf{H})\mathbf{P}^{\mathrm{f}} \\ &\approx \left(\mathbf{I}_{\mathrm{x}} - \mathbf{X}_{\mathrm{f}}\mathbf{Y}_{\mathrm{f}}^{\top}\left(\mathbf{Y}_{\mathrm{f}}\mathbf{Y}_{\mathrm{f}}^{\top} + \mathbf{R}\right)^{-1}\mathbf{H}\right)\mathbf{X}_{\mathrm{f}}\mathbf{X}_{\mathrm{f}}^{\top} \\ &\approx \mathbf{X}_{\mathrm{f}}\left(\mathbf{I}_{\mathrm{e}} - \mathbf{Y}_{\mathrm{f}}^{\top}\left(\mathbf{Y}_{\mathrm{f}}\mathbf{Y}_{\mathrm{f}}^{\top} + \mathbf{R}\right)^{-1}\mathbf{Y}_{\mathrm{f}}\right)\mathbf{X}_{\mathrm{f}}^{\top},\end{aligned}$$

which suggests to choose the following factor:

$$\mathbf{X}_{\mathrm{a}} = \mathbf{X}_{\mathrm{f}}\left(\mathbf{I}_{\mathrm{e}} - \mathbf{Y}_{\mathrm{f}}^{\top}(\mathbf{Y}_{\mathrm{f}}\mathbf{Y}_{\mathrm{f}}^{\top} + \mathbf{R})^{-1}\mathbf{Y}_{\mathrm{f}}\right)^{1/2}.$$

The ensemble transform Kalman filter: perturbations update

- This expression can be simplified into

$$\begin{aligned}
\mathbf{X}_\mathrm{a} &= \mathbf{X}_\mathrm{f}\left(\mathbf{I}_\mathrm{e} - \mathbf{Y}_\mathrm{f}^\top\left(\mathbf{Y}_\mathrm{f}\mathbf{Y}_\mathrm{f}^\top + \mathbf{R}\right)^{-1}\mathbf{Y}_\mathrm{f}\right)^{1/2} \\
&\overset{\mathrm{SMW}}{=} \mathbf{X}_\mathrm{f}\left(\mathbf{I}_\mathrm{e} - \left(\mathbf{I}_\mathrm{e} + \mathbf{Y}_\mathrm{f}^\top\mathbf{R}^{-1}\mathbf{Y}_\mathrm{f}\right)^{-1}\mathbf{Y}_\mathrm{f}^\top\mathbf{R}^{-1}\mathbf{Y}_\mathrm{f}\right)^{1/2} \\
&= \mathbf{X}_\mathrm{f}\left[\left(\mathbf{I}_\mathrm{e} + \mathbf{Y}_\mathrm{f}^\top\mathbf{R}^{-1}\mathbf{Y}_\mathrm{f}\right)^{-1}\left(\mathbf{I}_\mathrm{e} + \mathbf{Y}_\mathrm{f}^\top\mathbf{R}^{-1}\mathbf{Y}_\mathrm{f} - \mathbf{Y}_\mathrm{f}^\top\mathbf{R}^{-1}\mathbf{Y}_\mathrm{f}\right)\right]^{1/2} \\
&= \mathbf{X}_\mathrm{f}\left(\mathbf{I}_\mathrm{e} + \mathbf{Y}_\mathrm{f}^\top\mathbf{R}^{-1}\mathbf{Y}_\mathrm{f}\right)^{-1/2}.
\end{aligned}$$

- We conclude

$$\mathbf{X}_\mathrm{a} = \mathbf{X}_\mathrm{f}\mathbf{T}, \qquad \text{with} \qquad \mathbf{T} = \left(\mathbf{I}_\mathrm{e} + \mathbf{Y}_\mathrm{f}^\top\mathbf{R}^{-1}\mathbf{Y}_\mathrm{f}\right)^{-1/2}.$$

- Now, we can build the posterior ensemble as

$$i = 1,\ldots,N_\mathrm{e}: \qquad \mathbf{x}_i^\mathrm{a} = \overline{\mathbf{x}}^\mathrm{a} + \sqrt{N_\mathrm{e}-1}\,\mathbf{X}_\mathrm{f}\left[\mathbf{T}\right]_i = \overline{\mathbf{x}}^\mathrm{f} + \mathbf{X}_\mathrm{f}\left(\mathbf{w}^\mathrm{a} + \sqrt{N_\mathrm{e}-1}\left[\mathbf{T}\right]_i\right).$$

The ensemble transform Kalman filter: rotation matrix

- A more general anomaly update is

$$\mathbf{X}_\mathrm{a} = \mathbf{X}_\mathrm{f}\mathbf{T}\mathbf{U}, \qquad \text{where} \qquad \mathbf{U} \in \mathrm{O}(N_\mathrm{e}).$$

- It is important to require:

$$\mathbf{U}\mathbf{1} = \mathbf{1}, \qquad \text{where} \qquad \mathbf{1} = [1, \ldots, 1]^\top \in \mathbb{R}^{N_\mathrm{e}}.$$

This ensures that the updated ensemble is centred on $\mathbf{x}^\mathrm{a}$ [Livings et al. 2008; Sakov and Oke 2008b]. Indeed, we have

$$\mathbf{X}_\mathrm{a}\mathbf{1} = \mathbf{X}_\mathrm{f}\mathbf{T}\mathbf{U}\mathbf{1} = \mathbf{X}_\mathrm{f}\mathbf{T}\mathbf{1} = \mathbf{X}_\mathrm{f}\mathbf{1} = \mathbf{0},$$

and

$$\frac{1}{N_\mathrm{e}}\sum_{i=1}^{N_\mathrm{e}} \mathbf{x}_i^\mathrm{a} = \overline{\mathbf{x}}^\mathrm{a} + \frac{\sqrt{N_\mathrm{e}-1}}{N_\mathrm{e}}\mathbf{X}_\mathrm{a}\mathbf{1} = \overline{\mathbf{x}}^\mathrm{a}.$$

- $\mathbf{U} = \mathbf{I}_\mathrm{e}$ minimises the distance between $\mathbf{X}_\mathrm{a}$ and $\mathbf{X}_\mathrm{f}$ [Ott et al. 2004].
However, choosing random $\mathbf{U}$ may make the update more Gaussian and hence be more consistent with the EnKF assumptions [Lawson and Hansen 2004; Sakov and Oke 2008b].

- $\mathbf{U} = \mathbf{I}_\mathrm{e}$ in the following for the sake of simplicity.

The ensemble square-root Kalman filter (EnSRF)

▶ This is a variant of the deterministic EnKF where the update is carried out in state space, rather than in ensemble subspace as for the ETKF.

▶ Mean update: same as all the other EnKFs.

▶ Perturbation update [Sakov and Bertino 2011]:

$$\begin{aligned}\mathbf{X}_\mathrm{a} &= \mathbf{X}_\mathrm{f}\left(\mathbf{I}_\mathrm{e}+\mathbf{Y}_\mathrm{f}^\top\mathbf{R}^{-1}\mathbf{H}\mathbf{X}_\mathrm{f}\right)^{-\frac{1}{2}}\\ &\overset{\mathrm{SML}}{=} \left(\mathbf{I}_\mathrm{x}+\mathbf{X}_\mathrm{f}\mathbf{X}_\mathrm{f}^\top\mathbf{H}^\top\mathbf{R}^{-1}\mathbf{H}\right)^{-\frac{1}{2}}\mathbf{X}_\mathrm{f}\\ &= \left(\mathbf{I}_\mathrm{x}+\mathbf{P}^\mathrm{f}\mathbf{H}^\top\mathbf{R}^{-1}\mathbf{H}\right)^{-\frac{1}{2}}\mathbf{X}_\mathrm{f}.\end{aligned}$$

Very elegant formula though not practical!
Note that $\mathbf{I}_\mathrm{e}+\mathbf{P}^\mathrm{f}\mathbf{H}^\top\mathbf{R}^{-1}\mathbf{H}$ is in general not symmetric but it is diagonalisable with positive spectrum hence, it has a square root, which is unique.

▶ The EnSRF is algebraically equivalent and shares the left transform update with the adjustment ensemble Kalman filter (EAKF) [J. L. Anderson 2001].

DEnKF: the *deterministic* ensemble Kalman filter

- Reformulation of the perturbation update on the left:

We use $(\mathbf{I}_\mathrm{x}+\mathbf{A}\mathbf{B})^{-\frac{1}{2}} = \mathbf{I}_\mathrm{x} - \mathbf{A}\left(\mathbf{I}_\mathrm{y}+\mathbf{B}\mathbf{A}+\left[\mathbf{I}_\mathrm{y}+\mathbf{B}\mathbf{A}\right]^{\frac{1}{2}}\right)^{-1}\mathbf{B}$ with $\mathbf{A}=\mathbf{P}^\mathrm{f}\mathbf{H}^\top$ and $\mathbf{B}=\mathbf{R}^{-1}\mathbf{H}$ and we obtain:

$$\begin{aligned}\mathbf{X}_\mathrm{a} &= \left(\mathbf{I}_\mathrm{e}+\mathbf{P}^\mathrm{f}\mathbf{H}^\top\mathbf{R}^{-1}\mathbf{H}\right)^{\frac{1}{2}}\mathbf{X}_\mathrm{f} \\ &= \left\{\mathbf{I}_\mathrm{x} - \mathbf{P}^\mathrm{f}\mathbf{H}^\top\left(\mathbf{R}+\mathbf{H}\mathbf{P}^\mathrm{f}\mathbf{H}^\top+\mathbf{R}\left[\mathbf{I}_\mathrm{y}+\mathbf{R}^{-1}\mathbf{H}\mathbf{P}^\mathrm{f}\mathbf{H}^\top\right]^{\frac{1}{2}}\right)^{-1}\mathbf{H}\right\}\mathbf{X}_\mathrm{f}.\end{aligned}$$

- Effective gain in a deterministic setup:

Mimicking the stochastic EnKF, the effective gain for the updated perturbations (not the mean!) is

$$\widetilde{\mathbf{K}} = \mathbf{P}^\mathrm{f}\mathbf{H}^\top\left(\mathbf{R}+\mathbf{H}\mathbf{P}^\mathrm{f}\mathbf{H}^\top+\mathbf{R}\left[\mathbf{I}_\mathrm{y}+\mathbf{R}^{-1}\mathbf{H}\mathbf{P}^\mathrm{f}\mathbf{H}^\top\right]^{\frac{1}{2}}\right)^{-1},$$

as shown by [Whitaker and Hamill 2002] following [Andrews 1968], [Farchi and Bocquet 2019]. This can be reformulated as

$$\widetilde{\mathbf{K}} = \mathbf{K}\left\{\mathbf{I}_\mathrm{y}+\left(\mathbf{I}_\mathrm{y}+\mathbf{H}\mathbf{P}^\mathrm{f}\mathbf{H}^\top\mathbf{R}^{-1}\right)^{-\frac{1}{2}}\right\}^{-1}.$$

DEnKF: the deterministic ensemble Kalman filter

- An approximation of the EnSRF that mimics the update of the stochastic EnKF.
- Mean update: same as all the other EnKFs.
- In the weak assimilation regime, we have:

$$\left\{\mathbf{I}_{\mathrm{y}}+\left(\mathbf{I}_{\mathrm{y}}+\mathbf{H}\mathbf{P}^{\mathrm{f}}\mathbf{H}^{\top}\mathbf{R}^{-1}\right)^{-\frac{1}{2}}\right\}^{-1}\approx\frac{1}{2}\mathbf{I}_{\mathrm{y}}.$$

which suggests that the effective gain matrix can be approximated as

$$\widehat{\mathbf{K}}=\frac{1}{2}\mathbf{K},$$

i.e.

$$\mathbf{X}_{\mathrm{a}}\approx\left(\mathbf{I}_{\mathrm{x}}-\frac{1}{2}\mathbf{K}\mathbf{H}\right)\mathbf{X}_{\mathrm{f}}.$$

- Avoids the need to compute the square root $\rightarrow$ very similar to the stochastic EnKF (but deterministic).

DEnKF: the deterministic ensemble Kalman filter

- Why this filter is robust:

$$
\begin{aligned}
\widehat{\mathbf{P}}^{\mathrm{a}} = \widehat{\mathbf{X}}_{\mathrm{a}}\widehat{\mathbf{X}}_{\mathrm{a}}^{\top} &= \left(\mathbf{I}_{\mathrm{x}} - \frac{1}{2}\mathbf{K}\mathbf{H}\right)\mathbf{X}_{\mathrm{f}}\mathbf{X}_{\mathrm{f}}^{\top}\left(\mathbf{I}_{\mathrm{x}} - \frac{1}{2}\mathbf{H}^{\top}\mathbf{K}^{\top}\right) \\
&= \mathbf{P}^{\mathrm{f}} - \frac{1}{2}\mathbf{K}\mathbf{H}\mathbf{P}^{\mathrm{f}} - \frac{1}{2}\mathbf{P}^{\mathrm{f}}\mathbf{H}^{\top}\mathbf{K}^{\top} + \frac{1}{4}\mathbf{K}\mathbf{H}\mathbf{P}^{\mathrm{f}}\mathbf{H}^{\top}\mathbf{K}^{\top} \\
&= \left(\mathbf{I}_{\mathrm{x}} - \mathbf{K}\mathbf{H}\right)\mathbf{P}^{\mathrm{f}} + \frac{1}{4}\mathbf{K}\mathbf{H}\mathbf{P}^{\mathrm{f}}\mathbf{H}^{\top}\mathbf{K}^{\top} \\
&\geqslant \left(\mathbf{I}_{\mathrm{x}} - \mathbf{K}\mathbf{H}\right)\mathbf{P}^{\mathrm{f}} = \mathbf{P}^{\mathrm{a}},
\end{aligned}
$$

i.e. the analysis error covariance matrix of the DEnKF ($\widehat{\mathbf{P}^{\mathrm{a}}}$) is bounded by the exact one:

$$
\widehat{\mathbf{P}^{\mathrm{a}}} \geqslant \mathbf{P}^{\mathrm{a}}.
$$

- Ensemble update: In summary,

$$
\mathbf{x}_i^{\mathrm{a}} = \mathbf{x}_i^{\mathrm{f}} + \mathbf{K}\left[\mathbf{y} - \mathcal{H}\left(\frac{\mathbf{x}_i^{\mathrm{f}} + \overline{\mathbf{x}}^{\mathrm{f}}}{2}\right)\right].
$$

This nicely mimics the stochastic EnKF – the update can be carried out in parallel.

- Used in several intermediate and operational systems.

Serial EnKF

- Alternatively, the observations can be assimilated one at a time.
  - Drawback: can lead to suboptimality whenever an approximation is introduced.
  - Advantage: simple (especially the Potter scheme) and localisation is effective and elegant in this framework.

→ Used in the NCAR DART DA suite, and in most of J. L. Anderson's papers.

- Mean update:

$$\mathbf{x}^{\mathrm{a}} = \mathbf{x}^{\mathrm{f}} + \mathbf{K}(y - \mathfrak{h}(\mathbf{x}^{\mathrm{f}})) \qquad \mathbf{K} = \mathbf{P}^{\mathrm{f}}\mathbf{h}^{\top}/(r + \mathbf{h}\mathbf{P}^{\mathrm{f}}\mathbf{h}^{\top}).$$

- Perturbation update:

$$\widetilde{\mathbf{K}} = \frac{\mathbf{K}}{1 + 1/\sqrt{1 + r^{-1}\mathbf{h}\mathbf{P}^{\mathrm{f}}\mathbf{h}^{\top}}}.$$

Outline

1. The ensemble Kalman filter
   - Reminder
   - Principles
   - Mathematical prerequisites
   - The ETKF
   - The EnSRF
   - The DEnKF
   - The serial EnKF
2. Making EnKF work: localisation and inflation
   - Diagnostic
   - Localisation
   - Inflation
   - Why they are necessary
   - Hybrid localisation
3. References

Remedies to make EnKF work in high dimension

- Limited number $N_{\rm e}$ of anomalies: the sample covariance matrix is highly rank-deficient.

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

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

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

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

- Since $[\mathbf{B}]_{ij}$ vanish exponentially with the distance, we expect $\mathbb{E}\left([\mathbf{P}^{\rm e}-\mathbf{B}]_{ij}^2\right)$ to also vanish exponentially with the distance. Hence with $N_{\rm e}$ finite, the sample covariance $[\mathbf{P}^{\rm e}]_{ij}$ 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.

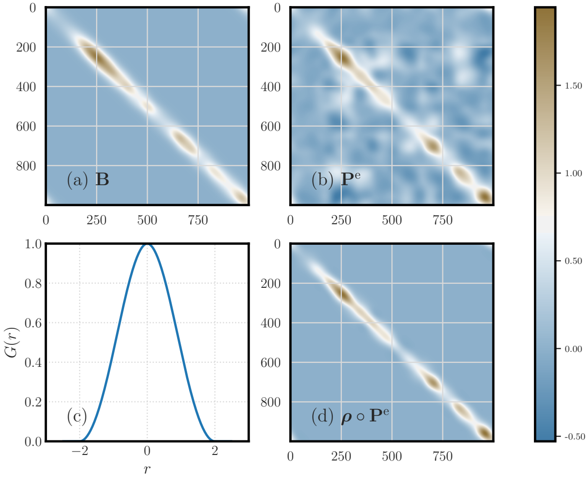
The Schur product of $\boldsymbol{\rho}$ and $\mathbf{B}$ is defined by (tapering of covariances)

$$[\boldsymbol{\rho} \circ \mathbf{P}^{\mathrm{e}}]_{ij} = [\boldsymbol{\rho}]_{ij} [\mathbf{P}^{\mathrm{e}}]_{ij}. \tag{3}$$

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.

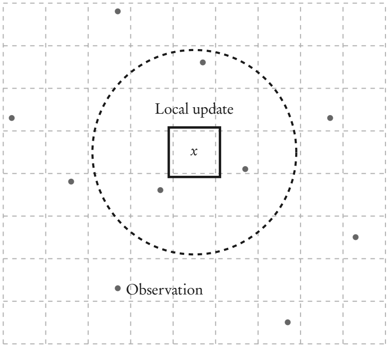
Domain localisation

▶ 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 still diverge!

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

$$\mathbf{P}^{\mathrm{e}} \longrightarrow \lambda^2 \mathbf{P}^{\mathrm{e}}, \tag{4}$$

or, alternatively,

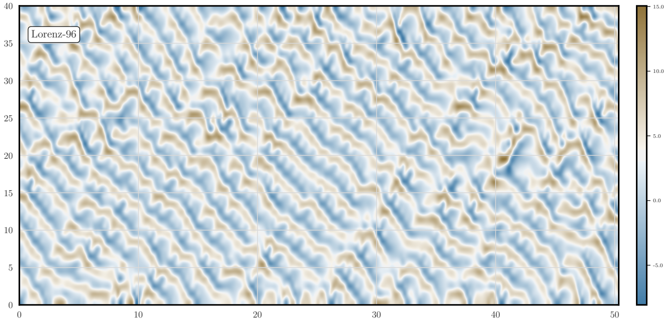
$$\mathbf{x}_{[n]} \longrightarrow \overline{\mathbf{x}} + \lambda \left(\mathbf{x}_{[n]} - \overline{\mathbf{x}} \right). \tag{5}$$

- 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].

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



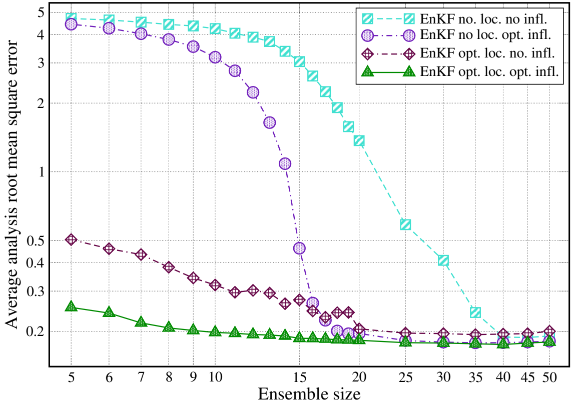
- It represents a mid-latitude zonal circle of the global atmosphere.
- Set of $N_x = 40$ ordinary differential equations [Lorenz and Emanuel 1998]:

$$\frac{\mathrm{d}x_n}{\mathrm{d}t} = (x_{n+1} - x_{n-2})x_{n-1} - x_n + F\,, \tag{6}$$

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

- Conservative system except for a forcing term F and a dissipation term $-x_n$.
- 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.

The local ensemble transform Kalman filter (LETKF)

▶ Since the ETKF update is carried out in ensemble subspace, only domain localisation can be used. Hence an ETKF update is performed for each local domain.

▶ Advantages: The scheme is simple. Local ETKF updates are computed in parallel.

▶ Drawback: it is not possible to assimilate nonlocal observations such as radiances, without drastic approximations.

▶ Updating N_x variables with an ETKF could be seen as a formidable task. However, (i) the updates are parallel (ii) each local update operates on a reduce observation vector which drastically reduces the local numerical cost.

Mean update of the local EnKF (except for the LETKF)

▶ The mean analysis in the local EnKF is carried out using the Kalman gain matrix

$$\mathbf{K} = \mathbf{B}\mathbf{H}^\top \left(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^\top\right)^{-1}, \quad (7)$$

where $\mathbf{H}$ is the observation operator (or tangent-linear thereof), and where the regularised

$$\mathbf{B} = \rho \circ \mathbf{P}^{\mathrm{e}} \quad (8)$$

is used in place of the sample $\mathbf{P}^{\mathrm{e}}$.

$\rightarrow$ numerically very costly!

▶ Usually applied in observation space whenever the observations can be seen as point-wise, i.e. local. Then $\mathbf{B}\mathbf{H}^\top \approx \rho_{\mathrm{xy}} \circ \left(\mathbf{P}^{\mathrm{e}}\mathbf{H}^\top\right)$ and $\mathbf{H}\mathbf{B}\mathbf{H}^\top \approx \rho_{\mathrm{yy}} \circ \left(\mathbf{H}\mathbf{P}^{\mathrm{e}}\mathbf{H}^\top\right)$ where ρ_{xy} represents ρ acting in the cross product of the state and observations spaces and ρ_{yy} represents ρ acting in the observations space. As a result:

$$\mathbf{K} \approx \rho_{\mathrm{xy}} \circ \left(\mathbf{P}^{\mathrm{e}}\mathbf{H}^\top\right) \left[\mathbf{R} + \rho_{\mathrm{yy}} \circ \left(\mathbf{H}\mathbf{P}^{\mathrm{e}}\mathbf{H}^\top\right)\right]^{-1}. \quad (9)$$

The local ensemble square root Kalman filter (LEnSRF)

- Perturbation update of the global EnSRF (in state space by definition):

$$\mathbf{X}_{\mathrm{a}} = \mathbf{T}\mathbf{X}_{\mathrm{f}} \qquad \text{with} \qquad \mathbf{T}_{\mathrm{x}} = \left(\mathbf{I}_{\mathrm{x}} + \mathbf{X}_{\mathrm{f}}\mathbf{X}_{\mathrm{f}}^{\top}\mathbf{H}^{\top}\mathbf{R}^{-1}\mathbf{H}\right)^{-\frac{1}{2}}.$$

- Covariance localisation: $\mathbf{X}_{\mathrm{f}}\mathbf{X}_{\mathrm{f}}^{\top} \longrightarrow \mathbf{B} = \boldsymbol{\rho} \circ \left(\mathbf{X}_{\mathrm{f}}\mathbf{X}_{\mathrm{f}}^{\top}\right)$,

$$\mathbf{X}_{\mathrm{a}} = \mathbf{T}\mathbf{X}_{\mathrm{f}} \qquad \text{with} \qquad \mathbf{T}_{\mathrm{x}} = \left(\mathbf{I}_{\mathrm{x}} + \mathbf{B}\mathbf{H}^{\top}\mathbf{R}^{-1}\mathbf{H}\right)^{-\frac{1}{2}}.$$

The LEnSRF: mode expansion

- The LSEnSRF requires the inverse square root of an $N_x \times N_x$ matrix. Too costly!

- We wish to make a mode expansion $\mathbf{B} = \boldsymbol{\rho} \circ \left(\mathbf{X}_\mathrm{f}\mathbf{X}_\mathrm{f}^\top\right) \approx \mathbf{X}_\mathrm{r}\mathbf{X}_\mathrm{r}^\top$, where $\mathbf{X}_\mathrm{r} \in \mathbb{R}^{N_x \times N_\mathrm{r}}$.
If we can do so, we will be able to make a perturbation à la ETKF in the expansion mode subspace rather than in the ensemble subspace.

- For high-dimensional chaotic models, we would typically have: $N_\mathrm{e} \ll N_\mathrm{r} \ll N_x$.

- The mathematical problem
Given the matrix $\mathbf{B} = \boldsymbol{\rho} \circ \left(\mathbf{X}_\mathrm{f}\mathbf{X}_\mathrm{f}^\top\right)$, we want to construct a matrix $\mathbf{X}_\mathrm{r} \in \mathbb{R}^{N_x \times N_\mathrm{r}}$such that

$$\mathbf{X}_\mathrm{r}\mathbf{X}_\mathrm{r}^\top \approx \mathbf{B} \quad \text{and} \quad \mathbf{X}_\mathrm{r}\mathbf{1} = \mathbf{0}.$$

The LEnSRF: modulation

- Suppose that there is a matrix $\mathbf{W}$ with N_r columns such that $\rho \approx \mathbf{W}\mathbf{W}^\top$.

- We define the modulation product of $\mathbf{W}$ and $\mathbf{X}_\mathrm{f}$ as the matrix with $N_\mathrm{r} N_\mathrm{e}$ columns:

$$[\mathbf{W}\Delta\mathbf{X}_\mathrm{f}]_n^{jN_\mathrm{e}+i} = [\mathbf{W}]_n^j \, [\mathbf{X}_\mathrm{f}]_n^i \, .$$

This is a mix between a Schur product (for the state variable index n) and a tensor product (for the ensemble indices i and j) [Buehner 2005].

The matrix $\mathbf{X}_\mathrm{r} = \mathbf{W}\Delta\mathbf{X}_\mathrm{f}$ is a solution with $N_\mathrm{r} = N_\mathrm{m} N_\mathrm{e}$ columns to the problem

$$\mathbf{X}_\mathrm{m}\mathbf{X}_\mathrm{m}^\top \approx \mathbf{B} \quad \text{and} \quad \mathbf{X}_\mathrm{m}\mathbf{1} = \mathbf{0}.$$

- The modulation product is based on a factorisation property shown by [Lorenc 2003] and is currently used for covariance localisation [Bishop, Whitaker, et al. 2017], including in operational centres [Arbogast et al. 2017].

The LEnSRF: the randomised SVD approach

▶ Direct mode expansion of $\rho \circ \mathbf{P}^{\mathrm{e}}$: a singular value decomposition (SVD) is unfeasable!

▶ The randomised SVD is an alternative to the Lanczos method.
(i) It defines a reduced random subspace in the column-space of $\rho \circ \mathbf{P}^{\mathrm{e}}$.
This subspace is generated by the application of $\rho \cdot \mathbf{P}^{\mathrm{e}}$ on random vectors $\mathbf{v}$: $\rho \cdot \mathbf{P}^{\mathrm{e}} \cdot \mathbf{v}$.
(ii) A regular svd is then performed in the generated subspace.

▶ Rigorous probilistic bounds can be obtained on the SVD, given the number of desired modes [Halko et al. 2011].

▶ Critical advantage: the application of $\rho \circ \mathbf{P}^{\mathrm{e}}$ on the random vectors $\mathbf{v}$ are independent and are hence carried out in parallel.

▶ It was applied to the local EnSRF in [Farchi and Bocquet 2019] .

$\longrightarrow$ Much more on randomised SVD in Omar Ghattas' lectures next week!

The LEnSRF: mode expansion

▶ Let us assume a mode expansion $\mathbf{B} = \rho \circ \left(\mathbf{X}_\mathrm{f}\mathbf{X}_\mathrm{f}^\top\right) \approx \mathbf{X}_\mathrm{r}\mathbf{X}_\mathrm{r}^\top$.

$$\mathbf{X}_\mathrm{a} \approx \mathbf{T}\mathbf{X}_\mathrm{f} \qquad \text{with} \qquad \mathbf{T}_\mathrm{x} = \left(\mathbf{I}_\mathrm{x} + \mathbf{X}_\mathrm{r}\mathbf{X}_\mathrm{r}^\top \mathbf{H}^\top \mathbf{R}^{-1}\mathbf{H}\right)^{-\frac{1}{2}}.$$

Let us use the last algebraic identity and obtain

$$\mathbf{X}_\mathrm{a} = \mathbf{T}_\mathrm{r}\mathbf{X}_\mathrm{f} \quad \text{with} \quad \mathbf{T}_\mathrm{r} = \mathbf{I}_\mathrm{x} - \mathbf{X}_\mathrm{r}\left(\mathbf{I}_\mathrm{r} + \mathbf{Y}_\mathrm{r}^\top\mathbf{R}^{-1}\mathbf{Y}_\mathrm{r} + \left[\mathbf{I}_\mathrm{r} + \mathbf{Y}_\mathrm{r}^\top\mathbf{R}^{-1}\mathbf{Y}_\mathrm{r}\right]^{\frac{1}{2}}\right)^{-1}\mathbf{Y}_\mathrm{r}^\top\mathbf{R}^{-1}\mathbf{H}.$$

Now, the algebra is performed in the reduced/mode subspace. It has been proposed in [Bocquet 2016] and later called the Gain Form of the ensemble transform Kalman filter in [Bishop, Whitaker, et al. 2017].

▶ An approximation which avoids the square root, similar to the DEnKF, is

$$\mathbf{X}_\mathrm{a} \approx \mathbf{X}_\mathrm{f} - \frac{1}{2}\left(\mathbf{I}_\mathrm{r} + \mathbf{Y}_\mathrm{r}^\top\mathbf{R}^{-1}\mathbf{Y}_\mathrm{r}\right)^{-1}\mathbf{Y}_\mathrm{r}^\top\mathbf{R}^{-1}\mathbf{H}\mathbf{X}_\mathrm{f}.$$

A multilayer extension of the L96 model

▶ We introduce the mL96 model, which consists of $P_z = 32$ coupled layers of the L96 model with $P_\mathrm{h} = 40$ variables:

$$\frac{\mathrm{d}x_{(z,h)}}{\mathrm{d}t} = \left(x_{(z,h+1)} - x_{(z,h-2)}\right) x_{(z,h-1)} - x_{(z,h)} + F_z + \underbrace{\delta_{\{z>0\}}\left(x_{(z-1,h)} - x_{(z,h)}\right)}_{\text{Coupling from below}} + \underbrace{\delta_{\{z<P_z\}}\left(x_{(z+1,h)} - x_{(z,h)}\right)}_{\text{Coupling from above}}.$$

▶ The forcing term linearly (and realistically) decreases from $F_1 = 8$ to $F_{32} = 4$.

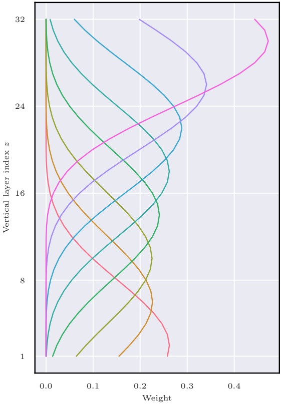
Satellite observations for the mL96 model

- Each column is observed independently via:

$$y_{c,h} = \sum_{z=1}^{P_z} [\mathbf{\Omega}]_{c,z} x_{z,h} + v_{c,h}, \quad v_{c,h} \sim \mathcal{N}(0,1),$$

where $\mathbf{\Omega}$ is a weighting matrix with $N_\mathrm{c} = 8$ channels that is designed to mimic satellite radiances.

- The 8×40 observations are available every $\Delta t = 0.05$.
- The runs are $10^4 \Delta t$ long.
- All algorithms use an ensemble of $N_\mathrm{e} = 8$ members.



Covariance localisation (with augmented ensembles) is used only in the vertical direction. Domain localisation (LETKF-like) is used in the horizontal direction.

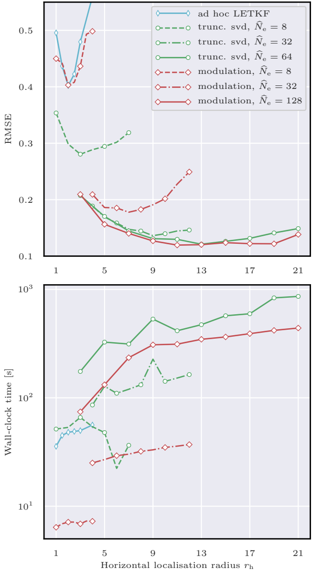
Results with the mL96 model

▶ Using covariance localisation in the vertical direction yields better RMSE scores than the LETKF.

▶ The modulation method requires a larger augmented ensemble size to yield similar RMSE scores as the randomised SVD method.

▶ Both methods benefit from the parallelisation of the local analyses, but the parallelisation potential of the randomised SVD method is not fully exploited because of a limited number of threads.

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

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

[25] 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.

[26] 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.

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

[28] 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.

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

[30] 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.

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

[32] 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.

References III

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

[34] 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.

[35] 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.

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

[37] 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`.

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

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

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

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

[42] 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.

[43] 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.

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