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:

\[ x_k = M_{k:k-1}(x_{k-1}, \lambda) + \eta_k, \quad y_k = H_k(x_k) + \epsilon_k. \]

The model and the observational errors, \( \eta_k, \epsilon_k : k = 1, \ldots, K \) are assumed to be uncorrelated in time, mutually independent, and they follow the pdfs \( p_\eta \) and \( p_\epsilon \).

Formal sequential Bayesian solution

An analysis step, in which the conditional pdf \( p(x_k|y_{k:0}) \) is updated using the latest observation vector, \( y_k \),

\[ p(x_k|y_{k:0}) \propto p_\eta(y_k - H_k(x_k)) p(x_k|y_{k-1:0}), \]

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

\[ p(x_{k+1}|y_{k:0}) = \int \! dx \ p_\eta(x_k - M_{k:k-1}(x_{k-1})) p(x_k|y_{k:0}). \]
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:**

\[
x_a^k = x_f^k + K_k \left( y_k - H_k x_f^k \right),
\]

\[
K_k = P_f^k H_k^\top \left( R_k + H_k P_f^k H_k^\top \right)^{-1},
\]

\[
P_a^k = (I_x - K_k H_k) P_f^k.
\]

**Forecast step:**

\[
x_{f}^{k+1} = M_{k+1:k} x_a^k,
\]

\[
P_{f}^{k+1} = M_{k+1:k} P_a^k M_{k+1:k}^\top + Q_{k+1}.
\]
The extended Kalman filter

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

\[
x_{k+1} = \mathcal{M}_{k+1:k}(x_{k}^{a}),
\]

\[
P_{k+1} = \mathcal{M}_{k+1:k} P_{k}^{a} \mathcal{M}_{k+1:k}^{\top} + Q_{k+1},
\]

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

**Drawbacks 1 & 2:** Extremely costly for large geophysical models: storage space (storage of \( P_{k}^{f} \)) and computations (\( \mathcal{M}_{k+1:k} P_{k}^{f} \mathcal{M}_{k+1:k}^{\top} \) 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 $P$ ($P^a$ and $P^f$) with an ensemble of states $x_1, x_2, \ldots, x_{Ne}$. The moments of the error could theoretically be approximated by the sample/empirical moments:

$$\bar{x}^f = \frac{1}{Ne} \sum_{i=1}^{Ne} x^f_i, \quad P^f \approx \frac{1}{Ne-1} \sum_{i=1}^{Ne} (x^f_i - \bar{x}^f)(x^f_i - \bar{x}^f)\top.$$

Define the normalised anomaly or perturbation matrix $\in \mathbb{R}^{Nx \times Ne}$

$$[X^f]_i = \frac{x^f_i - \bar{x}^f}{\sqrt{Ne-1}} \quad \Rightarrow \quad P^f \approx X^f X^f\top.$$

Likewise

$$\bar{x}^a = \frac{1}{Ne} \sum_{i=1}^{Ne} x^a_i, \quad P^a \approx X^a X^a\top \quad \text{where} \quad [X^a]_i = \frac{x^a_i - \bar{x}^a}{\sqrt{Ne-1}}.$$
The ensemble Kalman filter: Ansatz and mean update

An educated guess would suggest, for $i = 1 \ldots N_e$:

$$x_i^a = x_i^f + K \left( y - Hx_i^f \right).$$

but the correct answer is actually

$$x_i^a = x_i^f + K \left( y + \epsilon_i - Hx_i^f \right).$$

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

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

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

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

$$\bar{e} = \frac{1}{N_e} \sum_{i=1}^{N_e} e_i, \quad \Theta = \frac{1}{\sqrt{N_e-1}} [e_1 - \bar{e} \quad e_2 - \bar{e} \quad \ldots \quad e_{N_e} - \bar{e}] .$$

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

$$X_a = (I_x - KH)X_f + K\Theta,$$

which yields the empirical analysis error covariances:

$$P^a = (I_x - KH)P^f(I_x - KH)^\top + K\Theta\Theta^\top K^\top + (I_x - KH)X_f\Theta^\top K^\top + K\Theta X_f^\top (I_x - KH)^\top,$$

whose average on $\Theta$ is

$$\mathbb{E}[P^a] = (I_x - KH)P^f(I_x - KH)^\top + KRK^\top = (I_x - KH)P^f.$$

The last identity is valid if $K$ 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 $Y_f = HX_f + \Theta$, we have $K = X_fY_f^T(Y_fY_f^T)^{-1}$
  - **Deterministic**: denoting $Y_f = HX_f$, we have $K = X_fY_f^T(R + Y_fY_f^T)^{-1}$

- Forecast step: The ensemble is propagated using the full nonlinear model
  \[
  x_{i,k+1}^f = M_{k+1:k}(x_{i,k}^a),
  \]
  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 $H$

- Instead of estimating $P_fH^\top = X_fY_f^\top$ and $HP_fH^\top = Y_fY_f^\top$ in the Kalman gain, we can use the ensemble:

$$\bar{y}_f = \frac{1}{N_e} \sum_{i=1}^{N_e} H(x_i^f),$$

$$P_fH^\top = \frac{1}{N_e - 1} \sum_{i=1}^{N_e} \left( x_i^f - \bar{x}_f \right) \left[ H(x_i^f) - \bar{y}_f \right]^\top,$$

$$HP_fH^\top = \frac{1}{N_e - 1} \sum_{i=1}^{N_e} \left[ H(x_i^f) - \bar{y}_f \right] \left[ H(x_i^f) - \bar{y}_f \right]^\top.$$

These approximations rely on the key assumption:

$$[Y_f]_i = H\left( x_i^f - \bar{x}_f \right) \approx H(x_i^f) - \bar{y}_f.$$

- 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 (→ lecture 3).

It emulates the tangent linear of the observation model

Definitely a good point and at the origin of nonlinear EnVar techniques (→ lecture 3).
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.
The ensemble Kalman filter

Mathematical prerequisites

Key algebraic identities

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

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

► **Typical applications:**

- **Analysis error covariances:**

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

- **Kalman gain:**

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

$$A f(BA) = f(AB) A.$$  

→ **Proof in** [Higham 2008].

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

$$A \left( I_y + BA \right)^{-1} = \left( I_x + AB \right)^{-1} A.$$
Key algebraic identities

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

$$f(AB) = I + Ag(BA)B.$$  

Proof in [Higham 2008].

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

$$\left( I_x + AB \right)^{-\frac{1}{2}} = I_x - A \left( I_y + BA + \left[ I_y + BA \right]^{\frac{1}{2}} \right)^{-1} B,$$

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 $X_f$ in

$$P^f = X_f X_f^\top,$$

rather than updating $P^f$.

In the following, $X_f$ is called a factor of $P^f$, not a “square root” of $P^f$ as sometimes seen in geophysical data assimilation literature. This would clash with the mathematical definition of a square root matrix.

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

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

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

Note that $M$ 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:

\[ x^a = x^f + X_f w^a. \]

▶ Inserting this decomposition into the Kalman state update equation:

\[ x^f + X_f w^a = x^f + X_f X_f^T H^T \left( HX_f X_f^T H^T + R \right)^{-1} \delta, \]

where \( \delta = y - \mathcal{H}(x^f) \),

which suggests

\[ w^a \equiv X_f^T H^T \left( HX_f X_f^T H^T + R \right)^{-1} \delta = Y_f^T \left( Y_f Y_f^T + R \right)^{-1} \delta. \]

▶ Using the SMW identity, we finally obtain:

\[ w^a = \left( I_e + Y_f^T R^{-1} Y_f \right)^{-1} Y_f^T R^{-1} \delta. \]
From the analysis error covariance matrix of the Kalman filter, let us infer what the analysis anomaly matrix could be:

\[
P^a = (I_x - KH)P^f
\]

\[
\approx \left( I_x - X_f Y_f^\top \left( Y_f Y_f^\top + R \right)^{-1} H \right) X_f X_f^\top
\]

\[
\approx X_f \left( I_e - Y_f^\top \left( Y_f Y_f^\top + R \right)^{-1} Y_f \right) X_f^\top,
\]

which suggests to choose the following factor:

\[
X_a = X_f \left( I_e - Y_f^\top \left( Y_f Y_f^\top + R \right)^{-1} Y_f \right)^{1/2}.
\]
The ensemble transform Kalman filter: perturbations update

This expression can be simplified into

\[ X_a = X_f \left( I_e - Y_f^\top (Y_f Y_f^\top + R)^{-1} Y_f \right)^{1/2} \]

\[ \overset{\text{SMW}}{=} X_f \left( I_e - \left( I_e + Y_f^\top R^{-1} Y_f \right)^{-1} Y_f^\top R^{-1} Y_f \right)^{1/2} \]

\[ = X_f \left[ \left( I_e + Y_f^\top R^{-1} Y_f \right)^{-1} \left( I_e + Y_f^\top R^{-1} Y_f - Y_f^\top R^{-1} Y_f \right) \right]^{1/2} \]

\[ = X_f \left( I_e + Y_f^\top R^{-1} Y_f \right)^{-1/2}. \]

We conclude

\[ X_a = X_f T, \quad \text{with} \quad T = \left( I_e + Y_f^\top R^{-1} Y_f \right)^{-1/2}. \]

Now, we can build the posterior ensemble as

\[ i = 1, \ldots, N_e : \quad x_i^a = \bar{x}^a + \sqrt{N_e - 1} X_f [T]_i = \bar{x}^f + X_f \left( w^a + \sqrt{N_e - 1} [T]_i \right). \]
The ensemble transform Kalman filter: rotation matrix

- A more general anomaly update is
  \[ X_a = X_f TU, \]  
  where \( U \in O(N_e) \).

- It is important to require:
  \[ U1 = 1, \]  
  where \( 1 = [1, \ldots, 1]^\top \in \mathbb{R}^{N_e} \).

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

\[ X_a 1 = X_f TU1 = X_f T1 = X_f 1 = 0, \]

and

\[ \frac{1}{N_e} \sum_{i=1}^{N_e} x_i^a = \bar{x}^a + \frac{\sqrt{N_e - 1}}{N_e} X_a 1 = \bar{x}^a. \]

- \( U = I_e \) minimises the distance between \( X_a \) and \( X_f \) [Ott et al. 2004]. However, choosing random \( U \) may make the update more Gaussian and hence be more consistent with the EnKF assumptions [Lawson and Hansen 2004; Sakov and Oke 2008b].

- \( U = I_e \) 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]:

\[
X_a = X_f \left( I_e + Y_f^\top R^{-1} H X_f \right)^{-\frac{1}{2}} \\
SML = \left( I_x + X_f X_f^\top H^\top R^{-1} H \right)^{-\frac{1}{2}} X_f \\
= \left( I_x + P^f H^\top R^{-1} H \right)^{-\frac{1}{2}} X_f.
\]

Very elegant formula though not practical!

Note that \( I_e + P^f H^\top R^{-1} H \) 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 \((I_x + AB)^{-\frac{1}{2}} = I_x - A \left( I_y + BA + \left[ I_y + BA \right]^\frac{1}{2} \right)^{-1} B\) with \(A = P^fH^\top\) and \(B = R^{-1}H\) and we obtain:

\[
X_a = \left( I_e + P^fH^\top R^{-1}H \right)^{\frac{1}{2}} X_f
= \left\{ I_x - P^fH^\top \left( R + HP^fH^\top + R \left[ I_y + R^{-1}HP^fH^\top \right]^\frac{1}{2} \right)^{-1} H \right\} X_f.
\]

▶ Effective gain in a deterministic setup:

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

\[
\tilde{K} = P^fH^\top \left( R + HP^fH^\top + R \left[ I_y + R^{-1}HP^fH^\top \right]^\frac{1}{2} \right)^{-1},
\]

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

\[
\tilde{K} = K \left\{ I_y + \left( I_y + HP^fH^\top R^{-1} \right)^{-\frac{1}{2}} \right\}^{-1}.
\]
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:

\[
\left\{ I_y + \left( I_y + HP^f H^T R^{-1} \right)^{-\frac{1}{2}} \right\}^{-1} \approx \frac{1}{2} I_y.
\]

which suggests that the effective gain matrix can be approximated as

\[
\hat{K} = \frac{1}{2} K,
\]

i.e.

\[
X_a \approx \left( I_x - \frac{1}{2} KH \right) X_f.
\]

- Avoids the need to compute the square root → very similar to the stochastic EnKF (but deterministic).
DEnKF: the deterministic ensemble Kalman filter

Why this filter is robust:

\[
\hat{P}^a = \hat{X}_a \hat{X}_a^\top = \left( I_x - \frac{1}{2} KH \right) X_f X_f^\top \left( I_x - \frac{1}{2} H^\top K^\top \right)
\]

\[
= P_f - \frac{1}{2} KHP_f - \frac{1}{2} P_f H^\top K^\top + \frac{1}{4} KHP_f H^\top K^\top
\]

\[
= (I_x - KH) P_f + \frac{1}{4} KHP_f H^\top K^\top
\]

\[
\geq (I_x - KH) P_f = P^a,
\]

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

\[
\hat{P}^a \geq P^a.
\]

Ensemble update: In summary,

\[
x_i^a = x_i^f + K \left[ y - \mathcal{H} \left( \frac{x_i^f + \bar{x}^f}{2} \right) \right].
\]

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:**
  
  \[ x^a = x^f + K(y - h(x^f)) \quad K = P^f h^\top / (r + h P^f h^\top). \]

- **Perturbation update:**
  
  \[ \tilde{K} = \frac{K}{1 + 1/\sqrt{1 + r^{-1} h P^f h^\top}}. \]
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 $N_e$ of anomalies: the sample covariance matrix is highly rank-deficient.

- If $B$ is the true covariance matrix and $P_e$ is the ($N_e$-member) sample covariance matrix which approximates $B$, then:

$$
\mathbb{E}\left(\left[P_e - B\right]_{ij}^2\right) = \frac{1}{N_e - 1} \left(\left[B\right]_{ij}^2 + \left[B\right]_{ii} \left[B\right]_{jj}\right).
$$

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

$$
\mathbb{E}\left(\left[P_e - B\right]_{ij}^2\right)_{|i-j| \to \infty} \sim \frac{1}{N_e - 1} [B]_{ii} [B]_{jj}.
$$

- Since $[B]_{ij}$ vanish exponentially with the distance, we expect $\mathbb{E}\left(\left[P_e - B\right]_{ij}^2\right)$ to also vanish exponentially with the distance. Hence with $N_e$ finite, the sample covariance $[P_e]_{ij}$ 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 $P^e$ and the appearance of spurious correlations.

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

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

$$[\rho \circ P^e]_{ij} = [\rho]_{ij} [P^e]_{ij}. \quad (3)$$

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 $\rho$, $\rho \circ P^e$ turns out to be full-rank.
Covariance localisation with the Gaspari-Cohn function

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 $\lambda^2 \geq 1$:

  $$P^e \longrightarrow \lambda^2 P^e,$$

  or, alternatively,

  $$x_{[n]} \longrightarrow \bar{x} + \lambda (x_{[n]} - \bar{x}).$$

- Inflation can also come in an *additive form*: $x_{[n]} \longrightarrow x_{[n]} + \epsilon_{[n]}$.

- 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]:

\[
\frac{dx_n}{dt} = (x_{n+1} - x_{n-2})x_{n-1} - x_n + F ,
\]

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.
Making EnKF work: localisation and inflation

Why they are necessary

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

\[ K = BH^\top \left( R + HBH^\top \right)^{-1}, \]  

(7)

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

\[ B = \rho \circ P^e \]  

(8)

is used in place of the sample \( P^e \).

→ numerically very costly!

- Usually applied in observation space whenever the observations can be seen as point-wise, i.e. local. Then \( BH^\top \approx \rho_{xy} \circ (P^e H^\top) \) and \( HBH^\top \approx \rho_{yy} \circ (HP^e H^\top) \) where \( \rho_{xy} \) represents \( \rho \) acting in the cross product of the state and observations spaces and \( \rho_{yy} \) represents \( \rho \) acting in the observations space. As a result:

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

(9)
The local ensemble square root Kalman filter (LEnSRF)

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

\[ X_a = T X_f \quad \text{with} \quad T_x = \left( I_x + X_f X_f^\top H^\top R^{-1} H \right)^{-\frac{1}{2}}. \]

Covariance localisation:

\[ X_f X_f^\top \rightarrow B = \rho \circ (X_f X_f^\top), \]

\[ X_a = T X_f \quad \text{with} \quad T_x = \left( I_x + B H^\top R^{-1} H \right)^{-\frac{1}{2}}. \]
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 $B = \rho \circ (X_f X_f^\top) \approx X_r X_r^\top$, where $X_r \in \mathbb{R}^{N_x \times N_r}$. 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: $N_e \ll N_r \ll N_x$.

- The mathematical problem
Given the matrix $B = \rho \circ (X_f X_f^\top)$, we want to construct a matrix $X_r \in \mathbb{R}^{N_x \times N_r}$ such that

$$X_r X_r^\top \approx B \quad \text{and} \quad X_r 1 = 0.$$
The LEnSRF: modulation

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

▶ We define the modulation product of $\mathbf{W}$ and $\mathbf{X}_f$ as the matrix with $N_r N_e$ columns:

$$[\mathbf{W}\Delta \mathbf{X}_f]_{jN_e+i} = [\mathbf{W}]_j [\mathbf{X}_f]_i.$$ 

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 $\mathbf{X}_r = \mathbf{W}\Delta \mathbf{X}_f$ is a solution with $N_r = N_m N_e$ columns to the problem

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

▶ 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 $\rho \circ P^e$: a singular value decomposition (SVD) is unfeasible!

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

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

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

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

→ Much more on randomised SVD in Omar Ghattas’ lectures next week!
The LEnSRF: mode expansion

Let us assume a mode expansion $\mathbf{B} = \rho \circ (\mathbf{X}_f \mathbf{X}_f^\top) \approx \mathbf{X}_r \mathbf{X}_r^\top$.

$$\mathbf{X}_a \approx \mathbf{T} \mathbf{X}_f \quad \text{with} \quad \mathbf{T}_x = \left( \mathbf{I}_x + \mathbf{X}_r \mathbf{X}_r^\top \mathbf{H}^\top \mathbf{R}^{-1} \mathbf{H} \right)^{-\frac{1}{2}}.$$

Let us use the last algebraic identity and obtain

$$\mathbf{X}_a = \mathbf{T}_r \mathbf{X}_f \quad \text{with} \quad \mathbf{T}_r = \mathbf{I}_x - \mathbf{X}_r \left( \mathbf{I}_r + \mathbf{Y}_r^\top \mathbf{R}^{-1} \mathbf{Y}_r + \left[ \mathbf{I}_r + \mathbf{Y}_r^\top \mathbf{R}^{-1} \mathbf{Y}_r \right]^{\frac{1}{2}} \right)^{-1} \mathbf{Y}_r^\top \mathbf{R}^{-1} \mathbf{H}.$$

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

$$\mathbf{X}_a \approx \mathbf{X}_f - \frac{1}{2} \left( \mathbf{I}_r + \mathbf{Y}_r^\top \mathbf{R}^{-1} \mathbf{Y}_r \right)^{-1} \mathbf{Y}_r^\top \mathbf{R}^{-1} \mathbf{H} \mathbf{X}_f.$$
A multilayer extension of the L96 model

We introduce the \textit{mL96 model}, which consists of $P_z = 32$ coupled layers of the L96 model with $P_h = 40$ variables:

\[
\frac{dx_{(z,h)}}{dt} = \left( x_{(z,h+1)} - x_{(z,h-2)} \right) x_{(z,h-1)} - x_{(z,h)} + F_z + \delta_{\{z>0\}} \left( x_{(z-1,h)} - x_{(z,h)} \right) \qquad \text{Coupling from below}
\]

\[
+ \delta_{\{z<P_z\}} \left( x_{(z+1,h)} - x_{(z,h)} \right) \quad \text{Coupling from above}
\]

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:

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

where \( \Omega \) is a weighting matrix with \( N_c = 8 \) 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 \( N_e = 8 \) 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


References


References III


