Ensemble filtering and data assimilation for high-dimensional systems

Marc Bocquet
(bocquet@cerea.enpc.fr)

CEREA, École des Ponts ParisTech and EDF R&D
Université Paris-Est and INRIA

Outline

What this is about (disclaimer)

- An overview on ensemble Kalman filtering, and a little about particle filtering,...
- ...in the context of (very) high-dimensional geophysics (atmosphere & ocean): $n \sim 10^2 - 10^9$.
- This talk is about filtering, not smoothing.
- Variational methods (4D-Var $\equiv$ optimal control) are extremely successful in (operational) meteorology. The use of ensemble filters is a long-term effort to bypass the variational methods and to avoid its main disadvantages: the need of an adjoint, and the difficulty to explicitly extract posterior errors.
- The second part of the talk is more focused on my own contribution.

Outline

Particle filtering: a natural approach

- The ultimate goal of (Bayesian) data assimilation:
  - Statistically describe the system state by its complete pdf $p(\mathbf{x})$,
  - and assimilate observations through the Bayes formula

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})} . \tag{1}$$

- Given the system size, discretisation of the pdf is not affordable.

- The only feasible approach is Monte Carlo with N particles.

In the asymptotic limit ($N \longrightarrow \infty$), on should recover the exact Bayesian inference.

The bootstrap filter

It's simple!

- Ensemble of particles: $\left\{\mathbf{x}_h^1, \mathbf{x}_h^2, \ldots, \mathbf{x}_h^N\right\}$ at time t_h.
- Sampling of the system's pdf:

$$p_h(\mathbf{x}_h) \simeq \sum_{n=1}^{N} \omega_{h-1}^n \delta(\mathbf{x}_h - \mathbf{x}_h^n) \,. \tag{2}$$

- Analysis via a direct application of Eq.(1) :

$$\omega_h^n \propto \omega_{h-1}^n \, p(\mathbf{y}_h | \mathbf{x}_h^n) \,. \tag{3}$$

- Propagation:

$$\mathbf{x}_{h+1}^n = M_{h+1}(\mathbf{x}_h^n) + \mathbf{w}_{h+1} \,. \tag{4}$$

It's beautiful!

- No matrix inversion is necessary ($\neq$ EnKF),
- Trivially parallelism ($\simeq$ EnKF),
- The particles are actual solutions of the model ($\geq$ EnKF).

The bootstrap filter

▶ Quite rapidly, the ensemble degenerates. It is necessary to re-sample the ensemble from the weights of each member of the ensemble.

Probabilistic resampling [Metropolis et Ulam, 1944; Gordon, 1993]

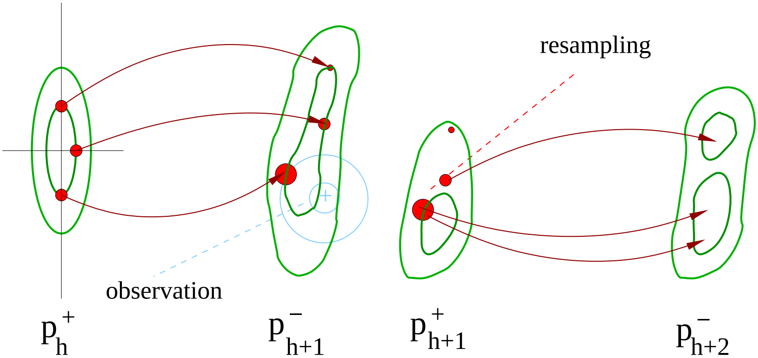
One directly uses the weights ω_h^n, $n = 1, \ldots, N$, as occurring probabilities.
$\longrightarrow$ standard sampling.
$\longrightarrow$ introduce a statistical sampling noise.

Residual resampling[Lui et Chen, 1998]

- If the ensemble size is N, one makes $\mathrm{E}[N\omega_h^n]$ copies of particle n.
- Remains a residue of $N\omega_h^n - \mathrm{E}[N\omega_h^n]$ for each of the particle.
- One draws the rest of the particles up to N particles according to this residual distribution.

$\longrightarrow$ Improvement in the performance of the bootstrap filter, but not essential.

The bootstrap filter



Examples in geophysics

Authors	model	var.	obs./cycle	ens. size
Zhou et al., 2006	land	684	1	800
Kivman, 2003	Lorenz 63	3	3	$250-1000$
Losa et al., 2003	ecosystem	24	-	1000
van Leeuwen, 2003	KdV	100	3	250
van Leeuwen, 2003	ocean QG model	2×10^5	$O(100)$	512
Nakano et al., 2007	Lorenz 95	40	20	$\geq 10^6$
Bocquet et al., 2008	Lorenz 95	10	5	10^4 ($\simeq$ EnKF)

- It does work! occasionally...
- The performance is highly dependent on the dynamics of the model.

The Lorenz 95 model

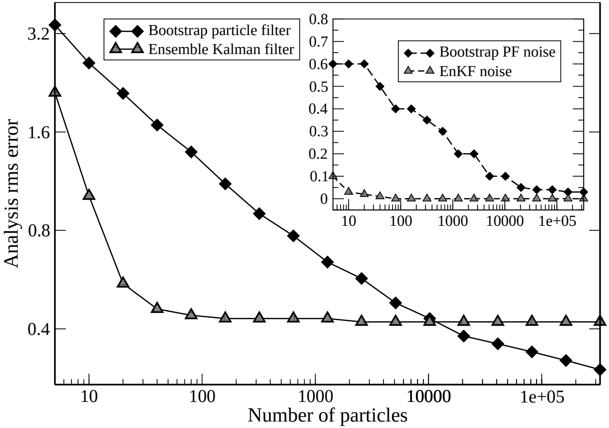
- The toy-model:
  - Represents a mid-latitude zonal circle of the global atmosphere [Lorenz and Emmanuel 1998].
  - $M = 40$ variables $\{x_m\}_{m=1,\ldots,M}$. For $m = 1, \ldots, M$:

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

    where $F = 8$, and the boundary is cyclic.
  - Conservative system except for a forcing term F and a dissipation term $-x_m$.
  - Chaotic dynamics, topological dimension of 13, a doubling time of about 0.42 time units, and a Kaplan-Yorke dimension of about 27.1.

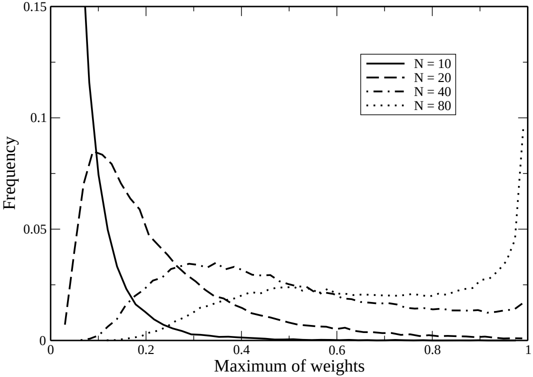
Lorenz 95, 10 variables

- $\Delta t = 0.05$ (6 hours in real time).
- Standard deviation of the obs. $\sigma = 1.5$.
- 1 site over 2 is observed
- EnKF : diagonal error covariance matrix of standard deviation $\chi = 1.5$.
- EnKF : localisation (correlation length $c = 10$).
- Skill of a filter given by the *rmse* of the analysis with the truth.



Degeneracy of the particle filter

- Very rapidly, but on average, the weights go to 0 except for a few particles with large weights.



- Maximal weight for a bootstrap filter with $N = 128$ applied to Lorenz 95 for four system's sizes: $M = 10$, 20, 40, and 80.

Degeneracy of the particle filter

Divergence of the required particle number

[Snyder et al., 2008] have studied the statistics of the highest weight. They have shown on a toy-model that the required size of the ensemble behaves like

$$M \sim \exp(\tau^2/2)\,, \tag{5}$$

where τ is the variance of the log-likelihood of the observations.

- exponentially scales with the dimensions of the state space and observation space.

Damned !

- Related to the *curse of dimensionality* [Bellman, 1961].
- A typical symptom is the shrinking of the hypersphere of radius 1 in the hypercube $[-1,1]^M$. Indeed, the ratio of volume scales like

$$\frac{(\pi/2)^{M/2}}{\Gamma\left(\frac{M}{2}+1\right)} \longrightarrow 0\,. \tag{6}$$

- In a high-dimensional analysis, the background prior and the observation prior overlap less and less!

Outline

The extended Kalman filter (1/2)

- Kalman (i.e. Gaussian) filtering as an alternative to variational data assimilation: less black box (access to errors), but less robust (a priori).
- High-dimensional dynamical system (say $\mathbf{x}_h \in \mathbb{R}^M$, with $M \sim 10^2 - 10^9$):

$$\begin{cases} \mathbf{x}_{h+1} = M_{h+1}(\mathbf{x}_h) + \mathbf{w}_{h+1} \\ \mathbf{y}_h = H_h(\mathbf{x}_h) + \mathbf{v}_h \end{cases}$$

White noise conditions:

$$\begin{aligned} &\mathrm{E}[\mathbf{w}_h] = \mathbf{0} \quad \mathrm{E}[\mathbf{w}_h \mathbf{w}_l^{\mathrm{T}}] = \mathbf{Q}_h \delta_{hl} \\ &\mathrm{E}[\mathbf{v}_h] = \mathbf{0} \quad \mathrm{E}\left[\mathbf{v}_h \mathbf{v}_l^{\mathrm{T}}\right] = \mathbf{R}_h \delta_{hl}\,, \quad \mathrm{E}\left[\mathbf{v}_h \mathbf{w}_l^{\mathrm{T}}\right] = \mathbf{0} \end{aligned} \tag{7}$$

Core assumptions

- Gaussian error statistics (or truncated to second-order moments)
- Linearisation of operators: $M_h \rightarrow \mathbf{M}_h$ and $H_h \rightarrow \mathbf{H}_h$.

The extended Kalman filter (2/2)

1. Initialisation: System state $\mathbf{x}_0^\mathrm{f}$ and error covariance matrix $\mathbf{P}_0^\mathrm{f}$.
2. Analysis at t_h
   - Gain computation: $\mathbf{K}_h = \mathbf{P}_h^\mathrm{f}\mathbf{H}_h^\mathrm{T}\left(\mathbf{H}_h\mathbf{P}_h^\mathrm{f}\mathbf{H}_h^\mathrm{T}+\mathbf{R}_h\right)^{-1}$
   - Estimator
   $$\mathbf{x}_h^\mathrm{a} = \mathbf{x}_h^\mathrm{f} + \mathbf{K}_h\left(\mathbf{y}_h - H_h[\mathbf{x}_h^\mathrm{f}]\right)$$
   - Error covariance matrix
   $$\mathbf{P}_h^a = \left(\mathbf{I}_M - \mathbf{K}_h\mathbf{H}_h\right)\mathbf{P}_h^\mathrm{f}$$
3. Forecast from t_h to t_{h+1}:
   - Forecast estimator $\mathbf{x}_{h+1}^\mathrm{f} = M_{h+1}[\mathbf{x}_h^\mathrm{a}]$
   - Forecast error covariance matrix
   $$\mathbf{P}_{h+1}^\mathrm{f} = \mathbf{M}_{h+1}\mathbf{P}_h^a\mathbf{M}_{h+1}^\mathrm{T} + \mathbf{Q}_{h+1}$$

From the extended Kalman filter to the ensemble Kalman filter

- Inappropriate for high-dimensional geophysical systems (few exceptions though).

What is wrong with the extended Kalman filter?

- Error covariance matrices too big to be stored
- Propagation of errors much too costly
- Linearisation induces errors in the error covariance matrix and in the estimator

Idea: represent uncertainty with an ensemble of N state vectors

[Evensen, 1994; Burgers et al., 1998]

- First and second-order moments obtained from

$$\overline{\mathbf{x}} = \frac{1}{N}\sum_{k=1}^{N}\mathbf{x}_k\,, \qquad \mathbf{P} = \frac{1}{N-1}\sum_{k=1}^{N}\left(\mathbf{x}_k-\overline{\mathbf{x}}\right)\left(\mathbf{x}_k-\overline{\mathbf{x}}\right)^{\mathrm{T}}\,.$$

- Why is this Monte-Carlo approach a good one?
  - Low storage requirements: N state vectors.
  - Exact propagation of the ensemble through the nonlinear model.
  - Still has to compute N model trajectories (much better than $2M$ though!).

The (stochastic) ensemble Kalman filter

1. Initialisation: System state $\mathbf{x}_0^f$ and error covariance matrix $\mathbf{P}_0^{\mathrm{f}}$.
2. Analysis at t_h
   - Create stochastic observation set ($k = 1, .., N$):

$$\mathbf{z}_k = \mathbf{z} + \mathbf{u}_k \qquad \sum_{k=1}^{N} \mathbf{u}_k = 0\,, \qquad \mathbf{R} = \frac{1}{N-1} \sum_{k=1}^{N} \mathbf{u}_k \mathbf{u}_k^{\mathrm{T}}$$

   - Kalman gain $\mathbf{K} = \mathbf{P}^{\mathrm{f}} \mathbf{H}^{\mathrm{T}} \left(\mathbf{H} \mathbf{P}^{\mathrm{f}} \mathbf{H}^{\mathrm{T}} + \mathbf{R} \right)^{-1}$
   - Computation of the analysis estimators $k = 1, .., N$ and their mean

$$\mathbf{x}_k^{\mathrm{a}} = \mathbf{x}_k^{\mathrm{f}} + \mathbf{K} \left(\mathbf{z}_k - H(\mathbf{x}_k^{\mathrm{f}}) \right) \qquad \overline{\mathbf{x}}^{\mathrm{a}} = \frac{1}{N} \sum_{j=1}^{N} \mathbf{x}_k^{\mathrm{a}}$$

   - Error covariance matrix: $\mathbf{P}^a = \dfrac{1}{N-1} \sum_{k=1}^{N} (\mathbf{x}_k^{\mathrm{a}} - \overline{\mathbf{x}}^{\mathrm{a}}) (\mathbf{x}_k^{\mathrm{a}} - \mathbf{x}^{\mathrm{a}})^{\mathrm{T}}$.
3. Forecast of $\left\{ \mathbf{x}_k^{\mathrm{f}} \right\}_{k=1,\dots,N}$, and $\mathbf{P}^{\mathrm{f}}$ from t_h to t_{h+1}:
   - Forecast of $\mathbf{x}_k^{\mathrm{f}} = M_{h+1}(\mathbf{x}_k^{\mathrm{a}})$, for $k = 1, .., N$, and of their mean $\overline{\mathbf{x}}^{\mathrm{f}} = \dfrac{1}{N} \sum_{k=1}^{N} \mathbf{x}_k^{\mathrm{f}}$.
   - Error covariance matrix: $\mathbf{P}^{\mathrm{f}} = \dfrac{1}{N-1} \sum_{k=1}^{N} \left(\mathbf{x}_k^{\mathrm{f}} - \overline{\mathbf{x}}^{\mathrm{f}} \right) \left(\mathbf{x}_k^{\mathrm{f}} - \overline{\mathbf{x}}^{\mathrm{f}} \right)^{\mathrm{T}}$.

The ensemble square root filter(s) (1/2)

- The deterministic variants of EnKF. [Anderson, 2001; Bishop et al., 2001; Whitaker and Hamill, 2002, Tippett et al., 2003]

- If $\mathbf{X}_k = (\mathbf{x}_k - \overline{\mathbf{x}})/\sqrt{N-1}$ are the scaled anomalies, define the scaled anomaly matrix $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_N)$. In an ensemble scheme, the background error covariance matrix $\mathbf{P}_b$ is of the form

$$\mathbf{P}_b = \mathbf{X}_b \mathbf{X}_b^{\mathrm{T}} . \tag{8}$$

- Main idea: factorise the analysis error covariance matrix $\mathbf{P}^a$.

$$\begin{aligned} \mathbf{P}_a &= \mathbf{P}_b - \mathbf{P}_b \mathbf{H}^{\mathrm{T}} \left(\mathbf{R} + \mathbf{H}\mathbf{P}_b\mathbf{H}^{\mathrm{T}} \right)^{-1} \mathbf{H}\mathbf{P}_b \\ &= \mathbf{X}_b \left(\mathbf{I} - (\mathbf{H}\mathbf{X}_b)^{\mathrm{T}} \left(\mathbf{R} + (\mathbf{H}\mathbf{X}_b)(\mathbf{H}\mathbf{X}_b)^{\mathrm{T}} \right)^{-1} (\mathbf{H}\mathbf{X}_b) \right) \mathbf{X}_b^{\mathrm{T}} \\ &\equiv \mathbf{X}_b \mathbf{D} \mathbf{X}_b^{\mathrm{T}} . \end{aligned} \tag{9}$$

- One can choose a decomposition of $\mathbf{D} = (\mathbf{D}^{1/2}\mathbf{U})(\mathbf{D}^{1/2}\mathbf{U})^{\mathrm{T}}$, where $\mathbf{U}$ is an arbitrary orthogonal matrix in ensemble space, so that

$$\mathbf{P}_a = \mathbf{X}_a \mathbf{X}_a^{\mathrm{T}} , \quad \text{with} \quad \mathbf{X}_a = \mathbf{X}_b \mathbf{D}^{1/2} \mathbf{U} . \tag{10}$$

The ensemble square root filter(s) (2/2)

- A particularly elegant class of square root EnKF is the ensemble transform Kalman filter. Apply Sherman-Morrison-Woodbury formula to **D**:

$$\begin{aligned}\mathbf{D} &= \mathbf{I} - (\mathbf{H}\mathbf{X}_b)^{\mathrm{T}}\left(\mathbf{R} + (\mathbf{H}\mathbf{X}_b)(\mathbf{H}\mathbf{X}_b)^{\mathrm{T}}\right)^{-1}(\mathbf{H}\mathbf{X}_b) \\ &= \left(\mathbf{I} + (\mathbf{H}\mathbf{X}_b)^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{H}\mathbf{X}_b)\right)^{-1} \qquad (11)\end{aligned}$$

This SREnKF is called ensemble transform Kalman filter (ETKF).

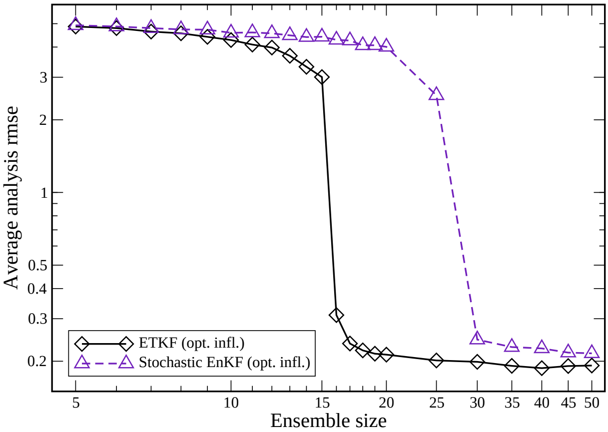
- Use a symmetric square root, such that $\mathbf{U}\mathbf{u} = \mathbf{u}$ where $\mathbf{u} = (1, \dots, 1)^{\mathrm{T}}$.

$$\mathbf{X}_a\mathbf{u} = \mathbf{X}_b\mathbf{D}^{\frac{1}{2}}\mathbf{U}\mathbf{u} = \mathbf{X}_b\mathbf{D}^{\frac{1}{2}}\mathbf{u} = \mathbf{X}_b\mathbf{u} = \mathbf{0}\,, \qquad (12)$$

because $\mathbf{X}_b\mathbf{u} = \mathbf{0}$ by construction. The performance of the symmetric SREnK filters is better.

Traditional EnKF versus ETKF for the Lorenz 95 case

- Time-lag between update: $\Delta t = 0.05$ (6 hours real time).
- All variables observed.
- Observations perturbed with a univariate normal distribution of std.dev. 1.
- Skill of a filter given by the *rmse* of the analysis with the truth.



But stochastic EnKFs are known to be more robust . . .

The European contributions (and others)

- The Reduced Rank Square Root filter [RRSQRT]

[Heemink, Verlaan, Segers, van Loon, Hanea, since 1995]

  - More robust square root form of the Kalman filter
  - Reduced rank: affordable!
  - Propagation of the uncertainty main modes according to the tangent linear.

- The Singular Evolutive Interpolated Kalman filter [SEIK]

[Pham, 2001]

  - It is an ensemble square root Kalman filter.
  - It is symmetric too.

- Others filters: Ensemble Adjustment Kalman filter [Anderson, 2001], hybrid filters [Hanea et al., 2007]

Outline

Sources of error in the ensemble Kalman filter schemes

External sources of error

- Model error.
- Deviation from Gaussianity of the error pdf.

Internal source of errors

- Sampling errors. First and second-order moments obtained from

$$\overline{\mathbf{x}} = \frac{1}{N}\sum_{k=1}^{N}\mathbf{x}_k\,, \qquad \mathbf{P} = \frac{1}{N-1}\sum_{k=1}^{N}\left(\mathbf{x}_k - \overline{\mathbf{x}}\right)\left(\mathbf{x}_k - \overline{\mathbf{x}}\right)^{\mathrm{T}}.$$

Consequence: divergence of the filter [Houtekamer & Mitchell 1998; Whitaker & Hamill 2002]

The ensemble Kalman filter (EnKF)

1. is unstable because of the external errors,
2. and unstable at small and moderate ensemble size because of sampling errors (internal errors).

Consequence and remedies: inflation, localisation

Inflation [Anderson & Anderson 1999; Houtekamer & Mitchell 1999; Hamil et al. 2001]

Rescale the ensemble to balance the underestimation of errors:

$$\mathbf{x}_k \longrightarrow \overline{\mathbf{x}} + r(\mathbf{x}_k - \overline{\mathbf{x}}) \qquad \text{(implies} \quad \mathbf{P}^{\mathrm{f}} \longrightarrow r^2\mathbf{P}^{\mathrm{f}}) .$$

Multi-ensemble configurations [Houtekamer & Mitchell 1998; Mitchell & Houtekamer 2009]

Compute the Kalman gain for one subensemble with the rest of the ensemble. Seems to cure the need for inflation (perfect model context).

Localisation [Houtekamer & Mitchell 1998; Hamil et al. 2001; Ott et al. 2004]

- Schur product of $\mathbf{P}^{\mathrm{f}}$ (or related matrices) with a limited-range covariance matrix ρ:

$$\mathbf{P}^{\mathrm{f}} \longrightarrow \rho \circ \mathbf{P}^{\mathrm{f}} .$$

- Assimilation of local observations within a given distance.

▶ But these are *ad hoc remedies*.

Strategies

Strategies that make current EnKFs work

- Context/model-dependent tuning of inflation, localisation scheme
- Adaptive tuning of inflation, localisation scheme [Mitchell and Houtekamer 1999; Anderson 2001-2009; Brankart et al. 2010; Li et al. 2009; etc.]: state of the art EnKF, mostly inspired by [Dee 1995], or cross-validation ideas [Silverman 1986].
- Objective identification of errors [Furrer and Bengtsson 2007] or of their consequences in the analysis [van Leeuwen 1999; Sacher and Bartello 2008]

Our strategy

- Identify sampling errors,
- and let the data assimilation system know about them.
- Bayesian approach (information flow under control).
- As a first step, rule out external sources of error.

Outline

Getting more from the ensemble

- Compute the prior pdf $p(\mathbf{x}|\mathbf{x}_1,\ldots,\mathbf{x}_N)$, assuming that
  - Members of the ensemble are drawn from an unknown Gaussian distribution of pdf $n(\mathbf{x}_b,\mathbf{B})$ that may differ from $n(\overline{\mathbf{x}},\mathbf{P})$.
  - If one knew $\mathbf{x}_b$ and $\mathbf{B}$ precisely, then the prior would be $p(\mathbf{x}|\mathbf{x}_1,\ldots,\mathbf{x}_N) = n(\mathbf{x}_b,\mathbf{B})$.
- Decomposing over all possible $\mathbf{x}_b$ and $\mathbf{B}$:

$$\begin{aligned}
p(\mathbf{x}|\mathbf{x}_1,\ldots,\mathbf{x}_N) &= \int \mathrm{d}\mathbf{x}_b \mathrm{d}\mathbf{B}\, p(\mathbf{x}|\mathbf{x}_1,\ldots,\mathbf{x}_N,\mathbf{x}_b,\mathbf{B})p(\mathbf{x}_b,\mathbf{B}|\mathbf{x}_1,\ldots,\mathbf{x}_N) \\
&= \int \mathrm{d}\mathbf{x}_b \mathrm{d}\mathbf{B}\, p(\mathbf{x}|\mathbf{x}_b,\mathbf{B})p(\mathbf{x}_b,\mathbf{B}|\mathbf{x}_1,\ldots,\mathbf{x}_N) \\
&\propto \int \mathrm{d}\mathbf{x}_b \mathrm{d}\mathbf{B}\, p(\mathbf{x}|\mathbf{x}_b,\mathbf{B})p(\mathbf{x}_1,\ldots,\mathbf{x}_N|\mathbf{x}_b,\mathbf{B})p(\mathbf{x}_b,\mathbf{B}) .
\end{aligned}$$

- Using the Gaussianity assumption, we get

$$p(\mathbf{x}|\mathbf{x}_1,\ldots,\mathbf{x}_N) \propto \int \mathrm{d}\mathbf{x}_b \mathrm{d}\mathbf{B}\, p(\mathbf{x}_b,\mathbf{B}) \exp\left(-\mathscr{L}(\mathbf{x},\mathbf{x}_b,\mathbf{B})\right) , \quad \text{with}$$

$$\mathscr{L}(\mathbf{x},\mathbf{x}_b,\mathbf{B}) = \frac{1}{2}(\mathbf{x}-\mathbf{x}_b)^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b) + \frac{1}{2}(N+1)\ln|\mathbf{B}| + \frac{1}{2}\sum_{k=1}^{N}(\mathbf{x}_k-\mathbf{x}_b)^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x}_k-\mathbf{x}_b) ,$$

where $|\mathbf{B}|$ is the determinant of $\mathbf{B}$.

Choosing priors for the background statistics

- To progress, we need to make assumptions on the background statistics $p(\mathbf{x}_b, \mathbf{B})$: *the statistics of the error statistics* or *hyperpriors*.
A very simple choice is a weakly informative prior: the Jeffreys' prior [Jeffreys 1961] with an additional assumption of independence for $\mathbf{x}_b$ and $\mathbf{B}$:

$$p(\mathbf{x}_b, \mathbf{B}) \equiv p_\mathrm{J}(\mathbf{x}_b, \mathbf{B}) = p_\mathrm{J}(\mathbf{x}_b) p_\mathrm{J}(\mathbf{B})$$

and

$$p_\mathrm{J}(\mathbf{x}_b) = 1\,, \quad p_\mathrm{J}(\mathbf{B}) = |\mathbf{B}|^{-\frac{M+1}{2}}\,.$$

- It has two desirable properties:
  1. It is invariant by re-parametrisation of state vectors.
  2. It leads to asymptotic Gaussianity: in the limit of a large ensemble, this choice should lead to the usual Gaussian prior used in classical EnKF analysis.

Effective priors

- After integration over $\mathbf{x}_b$ and $\mathbf{B}$, this leads to the $\mathscr{J}_b$ term

$$\mathscr{J}_b(\mathbf{x}) \equiv -\ln p(\mathbf{x}|\mathbf{x}_1, \cdots, \mathbf{x}_N) = \frac{N}{2} \ln \left| \frac{N}{N+1} (\mathbf{x} - \overline{\mathbf{x}})(\mathbf{x} - \overline{\mathbf{x}})^{\mathrm{T}} + (N-1)\mathbf{P} \right| .$$

Outline

Ensemble transform and gauge invariance

- Assume that the analysis is in the form $\mathbf{x} = \overline{\mathbf{x}} + \sum_{k=1}^{N} w_k (\mathbf{x}_k - \overline{\mathbf{x}})$.

If $\mathbf{X}_k = \mathbf{x}_k - \overline{\mathbf{x}}$ are the anomalies, and $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_N)$, then $\mathbf{x} = \overline{\mathbf{x}} + \mathbf{X}\mathbf{w}$. Hence

$$|\mathbf{A}| \equiv \left| \frac{N}{N+1} \mathbf{X}\mathbf{w}\mathbf{w}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}} + \mathbf{X}\mathbf{X}^{\mathrm{T}} \right| = \left|\mathbf{X}\mathbf{X}^{\mathrm{T}}\right| \left| \mathbf{I} + \frac{N}{N+1} \left(\mathbf{X}\mathbf{X}^{\mathrm{T}}\right)^{-1} \mathbf{X}\mathbf{w}\mathbf{w}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}} \right|$$
$$\propto 1 + \frac{N}{N+1} \mathbf{w}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}} \left(\mathbf{X}\mathbf{X}^{\mathrm{T}}\right)^{-1} \mathbf{X}\mathbf{w}\,.$$

- Gauge-fixing term
  - Define the gauge-fixing term $\mathscr{G}(\mathbf{w}) = \dfrac{N}{N+1} \mathbf{w}^{\mathrm{T}} \left(\mathbf{I}_N - \mathbf{X}^{\mathrm{T}} \left(\mathbf{X}\mathbf{X}^{\mathrm{T}}\right)^{-1} \mathbf{X} \right) \mathbf{w}$.
  - Insert it into the cost function

$$\widetilde{\mathscr{J}}_a(\mathbf{w}) = \mathscr{J}_o(\overline{\mathbf{x}} + \mathbf{X}\mathbf{w}) + \frac{N}{2} \ln\left(|\mathbf{A}| + \mathscr{G}(\mathbf{w})\right).$$

- Pivotal properties

The minima of $\widetilde{\mathscr{J}}_a(\mathbf{w})$ and $\mathscr{J}_a(\mathbf{x})$ are identical. Besides, one has $\mathscr{G}(\mathbf{w}^a) = \mathbf{0}$.

Variational analysis and posterior ensemble

- Complete cost (non-convex) function:

$$\widetilde{\mathcal{J}}_a(\mathbf{w}) = \frac{1}{2}\left(\mathbf{y} - H(\bar{\mathbf{x}} + \mathbf{X}\mathbf{w})\right)^{\mathrm{T}} \mathbf{R}^{-1}\left(\mathbf{y} - H(\bar{\mathbf{x}} + \mathbf{X}\mathbf{w})\right) + \frac{N}{2}\ln\left(1 + \frac{1}{N} + \sum_{k=1}^{N} w_k^2\right).$$

Once $\mathbf{w}^a$ is obtained, the posterior state estimator is given by $\mathbf{x}^a = \bar{\mathbf{x}} + \mathbf{X}\mathbf{w}^a$.

- Hessian of $\widetilde{\mathcal{J}}_b$ in ensemble space:

$$\widetilde{\mathcal{H}}_b = \nabla^2_{\mathbf{w}} \widetilde{\mathcal{J}}_b(\mathbf{w}) = N \frac{\left(1 + \frac{1}{N} + \mathbf{w}^{\mathrm{T}}\mathbf{w}\right)\mathbf{I}_N - 2\mathbf{w}\mathbf{w}^{\mathrm{T}}}{\left(1 + \frac{1}{N} + \mathbf{w}^{\mathrm{T}}\mathbf{w}\right)^2}.$$

Approximation: the analysis error cov. mat. is given by the inverse of the local Hessian

$$\widetilde{\mathbf{P}}_a \simeq \widetilde{\mathcal{H}}_a^{-1} = \left(\widetilde{\mathcal{H}}_b(\mathbf{w}^a) + \widetilde{\mathcal{H}}_o(\mathbf{w}^a)\right)^{-1}.$$

The posterior ensemble anomalies, in ensemble space, are given by the columns $\mathbf{W}^a_k$ of

$$\mathbf{W}^a = \left((N-1)\widetilde{\mathbf{P}}_a\right)^{1/2}, \qquad \mathbf{x}^a_k = \mathbf{x}^a + \mathbf{X}\mathbf{W}^a_k.$$

- Property: the posterior ensemble is centred on $\mathbf{x}^a$. Important for the consistency and the skills of the filter [Wang et al. 2004; Hunt et al. 2007; Livings et al. 2008; Sakov and Oke 2008].

Interpretation

- Assume the analysis is distant from the ensemble mean:

$$\sum_{k=1}^{N} w_k^2 \geq O(1) \,.$$

The ln function is barely constraining: priority given to observation.

- On the contrary, the analysis is close to the ensemble mean

$$\sum_{k=1}^{N} w_k^2 \ll 1 \,.$$

However, because of the $1/N$ offset in the ln function, the prior term cannot vanish even when the ensemble mean is taken as the optimal state.

$\longrightarrow$ Comes from the uncertainty of the ensemble mean at finite N. Same term $1 + 1/N$ as [Sacher and Bartello 2008].

- Reminiscent of Huber norm (for the ln part).

Robust and alternate ETKF-N

- Assume one trusts the ensemble forecasted mean to be the ensemble mean $\mathbf{x}_b = \bar{\mathbf{x}}$.
- Alternate finite-size ensemble transform Kalman filter:

$$\widetilde{\mathscr{J}}_b^{\mathrm{alt}} = \frac{N}{2} \ln \left(1 + \sum_{k=1}^{N} w_k^2 \right) .$$

- The only difference is in the $1/N$ (uncertainty of the empirical mean).

Local ETKF-N

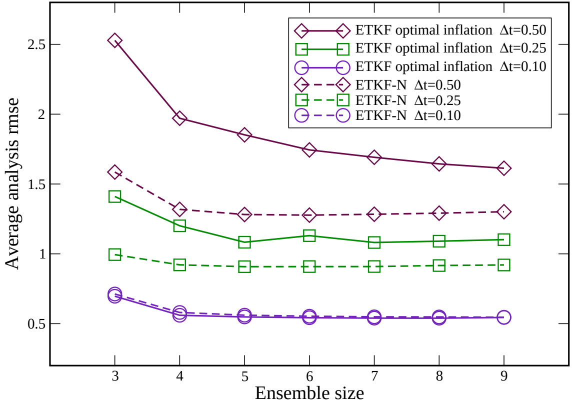
- We call this new EnKF scheme, the ETKF-N.
- Unfortunately, localisation is still mandatory !
- Following [Hunt et al. 2007; Harlim and Hunt, 2007], it is easy to generalise ETKF-N to a finite-size local ensemble transform Kalman filter, or LETKF-N (assimilation of observation within a given range).

Outline

The Lorenz 63 model

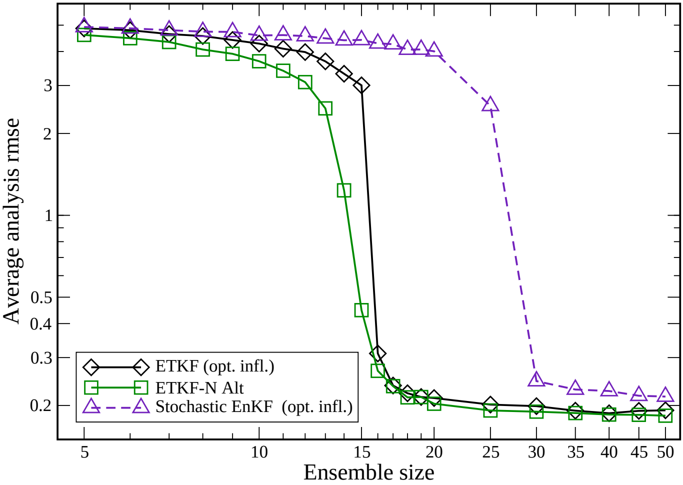
- The toy-model: $\frac{\mathrm{d}x}{\mathrm{d}t} = \sigma(y-x) \qquad \frac{\mathrm{d}y}{\mathrm{d}t} = \rho x - y - xz \qquad \frac{\mathrm{d}z}{\mathrm{d}t} = xy - \beta z\,.$

The parameters are set to the original values [Lorenz, 1963] $(\sigma,\rho,\beta) = (10, 28, 8/3)$.

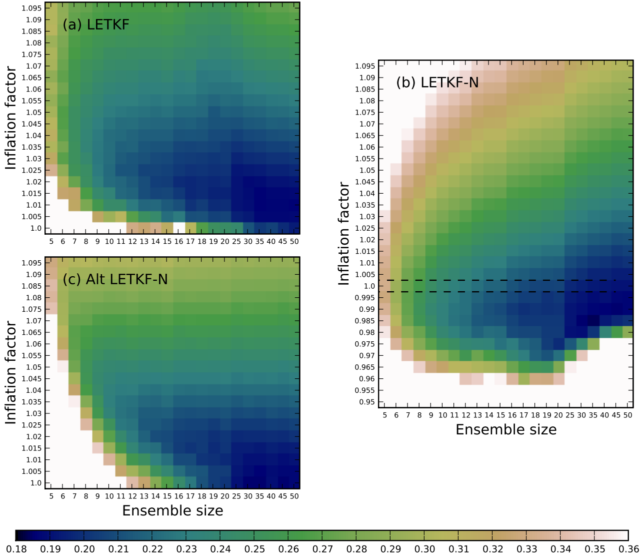


The Lorenz 95 model

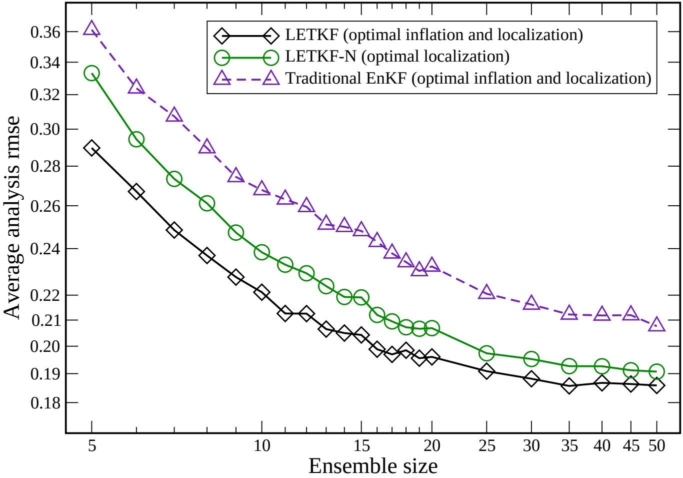
- Same setup as before, $\Delta t = 0.05$.



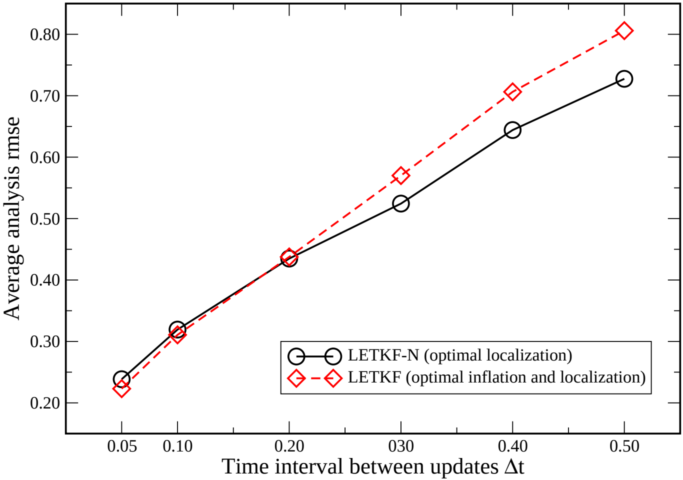
LETKF-N: ensemble size - inflation diagrams ($\Delta t = 0.05$)



LETKF-N: Skills ($\Delta t = 0.05$)



LETKF-N: Skills (several Δt and $N = 10$)



Outline

Conclusions

1. A large collection of ensemble Kalman filtering algorithms (stochastic and deterministic). Some of them are now operationally implemented and can compete with 4D-Var (Environment Canada).
2. Even advanced EnKFs need solutions for their structural deficiencies (mostly sampling errors).
3. ETKF-N seems to cure the need for inflation to a large extent, on toy-models.
4. Localisation still a very difficult issue.
5. Model error treatment is especially interesting within the EnKF schemes (inflation of external origin, calibration of the ensemble, adaptive schemes, etc.). This was not discussed today!
6. Particle filters still in their infancy for high-dimensional geophysical systems. But solutions are near?

Beyond Gaussian statistical modeling in geophysical data assimilation, Bocquet M., Pires C. A. and Wu L., Mon. Wea. Rev., **138**, 2997-3023, 2010.

Ensemble Kalman Filtering without the intrinsic need for inflation. Bocquet M., Mon. Wea. Rev., in revision, 2011.