Quasi static ensemble variational data assimilation

A. Fillion[1,2] M. Bocquet[1] S. Gratton[3]

[1]CEREA,
Joint laboratory École des Ponts ParisTech and EdF R&D,
Université Paris-Est, Marne-la-Vallée, France

[2]CERFACS,
Toulouse, France

[3]INPT-IRIT,
Université de Toulouse and ENSEEIHT, Toulouse, France

12th international EnKF workshop, june 2017

Outline

Variational data assimilation

- The analysis relies on a cost function minimization.
- This method can miss the global minimum.
- Quasi static (QS) minimizations use the cost function temporal structure to localize the global minimum.
  - Pires et al. 1996[1] introduced it in one cycle of a variational assimilation.
  - We place it in multiple cycles of an ensemble variational assimilation.

[1] C. Pires, R. Vautard, and O. Talagrand. On extending the limits of variational assimilation in nonlinear chaotic systems.
Tellus A, 48:96–121, 1996

Outline

Objective: Justify the use of QS minimizations in sequential ensemble variational data assimilation with a perfect model.

1. Basic algorithms
   - 4D-Var and IEnKS
2. Performance quantification of an assimilation
   - Empirical and Theoretical
3. Long term impact of the cycling
   - Simplest theoritical case
   - Chaotic case
4. Quasi static algorithms
5. Numerical experiments

Outline

The data assimilation window (DAW)

Cycling is controled by the DAW parameters:

- During the analysis, observations (obs) from time t_K to t_L are assimilated.
- During the propagation, the DAW is shifted S time steps in the future.
- Single data assimilation imposes $K = L - S + 1$ (no overlap). Thus S is also the DAW number of observations.
- Filtering: $K = L = 0$ and $S = 1$.

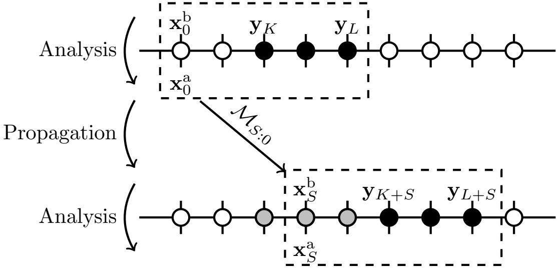
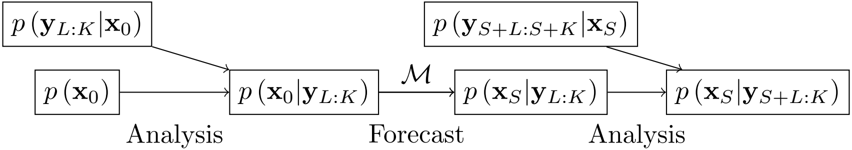


Figure: Two first cycles of an assimilation with $K = 2$, $L = 4$ and $S = 3$.

Bayes' framework



- The exact cost function at the kth cycle is

$$G\left(\mathbf{x}_{kS}|\mathbf{y}_{kS+L:K}\right) \propto -\ln p\left(\mathbf{x}_{kS}|\mathbf{y}_{kS+L:K}\right). \tag{1}$$

- Bayes' theorem yields

$$G\left(\mathbf{x}_{kS}|\mathbf{y}_{kS+L:K}\right) \propto -\ln p\left(\mathbf{x}_{kS}|\mathbf{y}_{(k-1)S+L:K}\right), \quad \text{(bg term)} \tag{2}$$

$$-\ln p\left(\mathbf{y}_{kS+K:kS+L}|\mathbf{x}_{kS}\right). \quad \text{(obs term)} \tag{3}$$

Observation and background term

- With Gaussian errors, the observation term is

$$-\ln p\left(\mathbf{y}_{kS+K:kS+L}|\mathbf{x}_{kS}=\mathbf{x}\right) \propto \frac{1}{2}\sum_{l=kS+K}^{kS+L}\left\|\mathbf{y}_l-\mathcal{H}\circ\mathcal{M}_{l\leftarrow kS}\left(\mathbf{x}\right)\right\|^2_{\mathbf{R}^{-1}}. \tag{4}$$

- If the operators $\mathcal{H}, \mathcal{M}$ are non-linear the background (bg) term

$$-\ln p\left(\mathbf{x}_{kS}|\mathbf{y}_{(k-1)S+L:K}\right), \tag{5}$$

is complex.

- Thus it has to be approximated, this approximation determines our algorithm.

The 4D-Var[2]

- The 4D-Var background approximation is

$$-\ln p\left(\mathbf{x}_{kS} = \mathbf{x} | \mathbf{y}_{(k-1)S+L:K}\right) \propto \frac{1}{2}\left\|\mathbf{x}^{\mathrm{b}}_{kS} - \mathbf{x}\right\|^2_{\mathbf{B}^{-1}}, \tag{6}$$

  where
  - $\mathbf{x}^{\mathrm{b}}_{kS} \equiv \mathcal{M}_{kS\leftarrow(k-1)S}\left(\mathbf{x}^{\mathrm{a}}_{(k-1)S}\right)$ is the bg mean and $\mathbf{x}^{\mathrm{a}}_{(k-1)S}$ is the last cycle analysis,
  - $\mathbf{B}$ is a constant bg error covariance matrix.
- It is a Gaussian background approximation, only the first moment is tracked.

[2] E. Blayo, M. Bocquet, E. Cosme, and L.F. Cugliandolo. *Advanced Data Assimilation for Geosciences.*
Lecture Notes of the Les Houches School of Physics: Special Issue, June 2012. 2014

The IEnKS[3]

- The IEnKS background approximation in the ensemble space is

$$-\ln p\left(\mathbf{x}_{kS} = \bar{\mathbf{x}}^{\mathrm{b}}_{kS} + \mathbf{X}^{\mathrm{b}}_{kS}\mathbf{w} | \mathbf{y}_{(k-1)S+L:K}\right) \propto \frac{1}{2}\left\|\mathbf{w}\right\|^2, \quad (7)$$

where

  - $\bar{\mathbf{x}}^{\mathrm{b}}_{kS} \equiv \mathbf{E}^{\mathrm{b}}_{kS}\frac{\mathbf{1}}{n}$ is the bg ensemble mean,
  - $\mathbf{X}^{\mathrm{b}}_{kS} \equiv \frac{1}{\sqrt{n-1}}\left(\mathbf{E}^{\mathrm{b}}_{kS} - \bar{\mathbf{x}}^{\mathrm{b}}_{kS}\mathbf{1}^{\mathrm{T}}\right)$ is the bg ensemble normalized anomalies,
  - $\mathbf{E}^{\mathrm{b}}_{kS} \equiv \mathcal{M}_{kS\leftarrow(k-1)S}\left(\mathbf{E}^{\mathrm{a}}_{(k-1)S}\right)$ is the bg ensemble and $\mathbf{E}^{\mathrm{a}}_{(k-1)S}$ is the last cycle analyzed ensemble.

- It is a Gaussian background approximation, the two first moments are tracked. So it is (quite) exact when the operators $\mathcal{H}, \mathcal{M}$ are linear.

[3] M. Bocquet and P. Sakov. An iterative ensemble kalman smoother. *Quarterly Journal of the Royal Meteorological Society*, 140:1521–1535, 2014

Outline

Empirical performance

- The smoothing RMSE at cycle k with lag $L - l$ is defined by

$$\mathrm{RMSE}_{kS+l} \equiv \left\| \mathbf{x}_{kS+l} - \mathbf{x}^{\mathrm{a}}_{kS+l} \right\| . \tag{8}$$

- It's a random variable. In our numerical experiments the RMSE is averaged over cycles:

$$\mathrm{aRMSE}_{l} \equiv \frac{1}{N} \sum_{k=0}^{N-1} \left\| \mathbf{x}_{kS+l} - \mathbf{x}^{\mathrm{a}}_{kS+l} \right\| . \tag{9}$$

- It measures the long term impact of cycling on the assimilation performance[4].

[4] The aRMSE convergence depends on ergodic properties which are beyond the scope of this presentation.

Theoretical performance

- The former quantity is difficult to exploit analytically. In theoretical developments we prefer the expected MSE

$$\mathrm{eMSE}_{kS+l} \equiv \mathbb{E}\left[\left\|\mathbf{x}_{kS+l} - \mathbf{x}^{\mathrm{a}}_{kS+l}\right\|^2\right]. \tag{10}$$

- Its asymptotic limit measures the long term impact of cycling on the assimilation performance

$$\mathrm{eMSE}_{\infty+l} \equiv \lim_{k\to\infty} \mathrm{eMSE}_{kS+l}. \tag{11}$$

- In the following, simplifying assumptions will be made to express this limit.

Outline

Outline

Simplifying assumptions

- The state space dimension is $m = 2$.
- $\mathcal{H} = \mathbf{B} = \mathbf{R} = \mathbf{I}_2$.
- $\mathcal{M}_{i \leftarrow j} = \mathbf{M}^{i-j} = \begin{pmatrix} \alpha_1^{i-j} & 0 \\ 0 & \alpha_2^{i-j} \end{pmatrix}$
  - $|\alpha_1| > 1$ to have an unstable direction,
  - $|\alpha_2| < 1$ to have a stable direction.
- The IEnKS becomes a Kalman smoother (no sampling errors).

Performance expression

The 4D-Var and IEnKS asymptotic eMSE is expressible

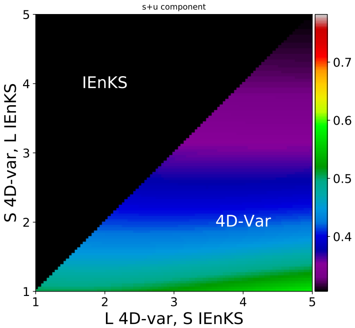


Figure: IEnKS, 4D-Var asymptotic filtering eMSE as a function of S, L

- The 4D-Var asymptotic filtering eMSE is constant with L and decreases with S.
- The IEnKS asymptotic filtering eMSE is constant with L, S.

Interpretations

- The error forecast in filtering eMSE compensates performance gain with remote observations
  - Few dependancy with L.
- The IEnKS Gaussian background approximation is exact.
  - No loss of information during the propagation, thus each S configuration is equivalent.
- The 4D-Var Gaussian background approximation is not exact.
  - There is a loss of information during the propagation.
  - The greater S the lesser the assimilation relies on bg approximations so it is more performant.

Assimilating 100 observations

- S=1 requires 100 cycles.
- S=10 requires 10 cycles.
- S=100 requires 1 cycle.

Outline

aRMSE with L95

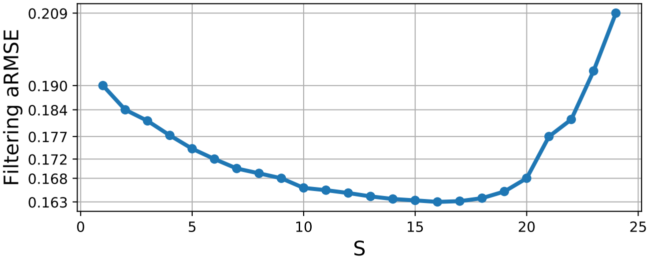


Figure: IEnKS filtering RMSE as a function of S ($L = S$) averaged over 5×10^5 cycles with the 40 variable Lorenz'95 model. The ensemble contains 20 members (logarithmic scale).

- The aRMSE decreases until $S = 15$ then it increases.

Interpretation: $S < 15$

- The model is now non-linear, Gaussianity is lost, the IEnKS background approximation is not exact anymore.
  - Analogy with the 4DVar.
  - There is a loss of information during the propagation.
  - The greater S the lesser the assimilation relies on bg approximations so it is more performant.
- To understand the case $S > 15$, let's have a look at the IEnKS analysis.

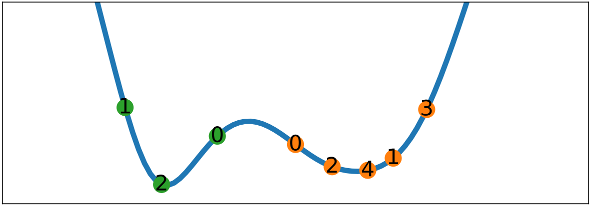
IEnKS analysis

- The IEnKS cost function in the ensemble space is

$$J(\mathbf{w}) = \frac{1}{2}\|\mathbf{w}\|^2 + \frac{1}{2}\sum_{l=K}^{L}\left\|\mathbf{y}_l - \mathcal{H}\circ\mathcal{M}_{l:0}\left(\bar{\mathbf{x}}^{\mathrm{b}} + \mathbf{X}^{\mathrm{b}}\mathbf{w}\right)\right\|^2_{\mathbf{R}^{-1}}, \quad (12)$$

where $\bar{\mathbf{x}}^{\mathrm{b}} = \mathbf{E}^{\mathrm{b}}\frac{\mathbf{1}_n}{n}$ is the bg ensemble mean and $\mathbf{X}^{\mathrm{b}} = \frac{\mathbf{E}^{\mathrm{b}} - \bar{\mathbf{x}}^{\mathrm{b}}\mathbf{1}_n^{\mathrm{T}}}{\sqrt{n-1}}$ is the bg normalized anomaly.

- It's minimization relies on a Gauss-Newton algorithm which is a non-global method.



Cost functions

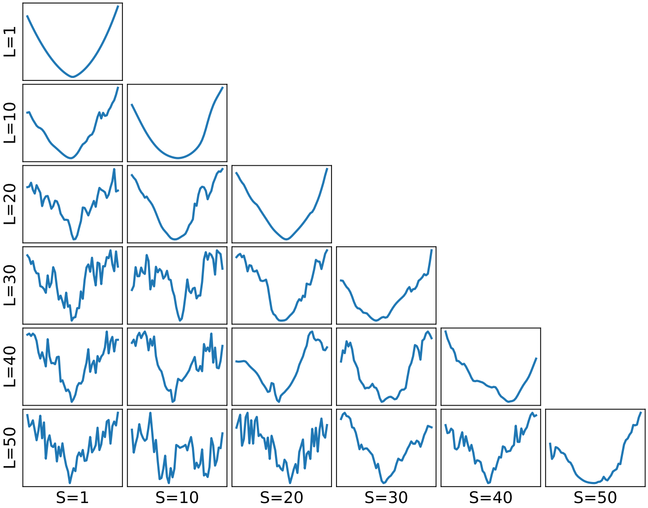


Figure: IEnKS cost functions in one direction of the ensemble space with Lorenz'95 model.

The need for QS minimizations

- The bigger L is, the narrower the cost function global minimum basin of attraction is.
- If the Gauss-Newton starting point $\bar{\mathbf{x}}^{\mathrm{b}}$ falls outside of this basin of attraction, the analysis will be deteriorated.
- Quasi-static minimizations consists in multiple minimizations of cost functions with increasing L.
  - Each minimum becomes the next minimization starting point.
- On the previous figure it corresponds to minimizing the cost functions in diagonal.

Outline

The IEnKS-QS

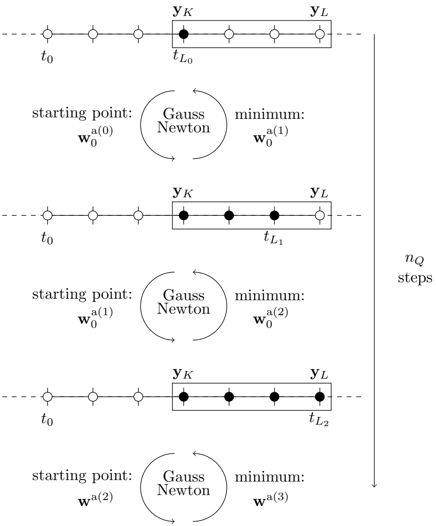


Figure: The analysis of the IEnKS-QS

Defaults

- Repeating the GN minimizations is numerically costly.
- Precision on intermediate minimums is not necessary, just imports to be in the next minimum attraction basin.
- One can limit the number of intermediate GN iterations.

The IEnKS-QC

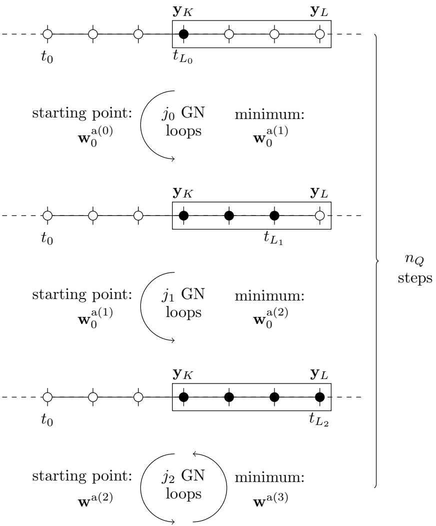


Figure: The analysis of the IEnKS-QC (Quasi Convergent)

Outline

aRMSE as function of S,L

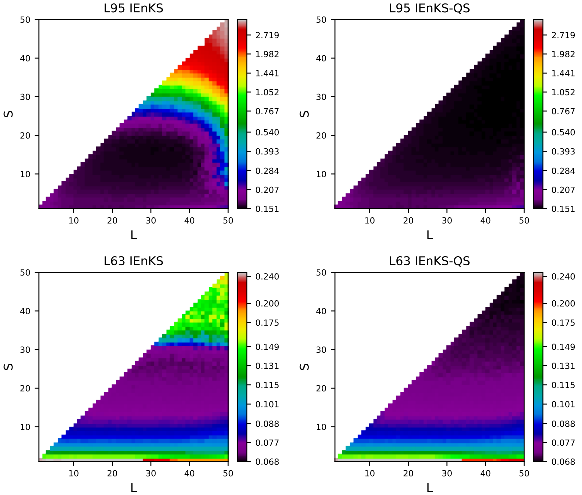


Figure: IEnKS and IEnKS-QS ($n_Q = S$) filtering aRMSE as a function of S, L with Lorenz' 63 and Lorenz' 95 models.

IEnKS-QS vs IEnKS

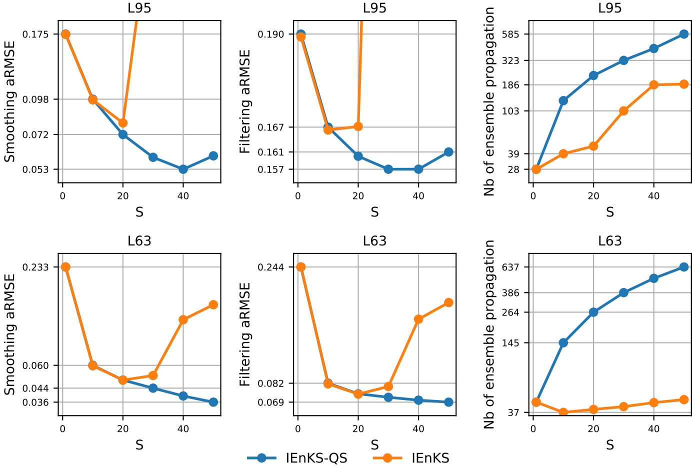


Figure: IEnKS-QS ($n_Q = S$, $L = S$), IEnKS ($L = S$) smoothing, filtering aRMSE and number of ensemble propagations as a function of S with Lorenz'63 / 95 models (logarithmic scale).

IEnKS-QS vs IEnKS-QC

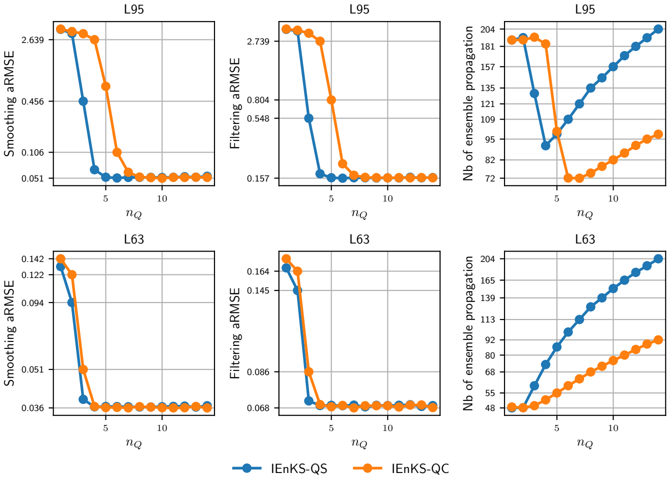


Figure: IEnKS-QS and IEnKS-QC ($S = L = 50$) smoothing, filtering aRMSE and number of ensemble propagations as a function of n_Q with Lorenz'63 / 95 models (logarithmic scale).

Conclusions

- The 4D-Var and IEnKS performance increase with S the DAW number of observations.
  - Because the assimilation relies less on the Gaussian background approximation.
- However, with a chaotic model, the cost function global minimum basin of attraction shrinks as S increases.
  - It causes Gauss-Newton to miss the global minimum, which deteriorates the analysis performance.
- QS minimizations avoid this problem
  - It brings the minimization starting point closer to the global minimimum as its basin of attraction shrinks.
  - But repeating the minimizations is costly.
- QC minimizations avoid this problem
  - The Gauss-Newton multiple increments unavoidable to minimize a non-quadratic cost function are reported in time.

For further reading

Thanks for your attention !

Paper in preparation : A. Fillion, M. Bocquet, and S. Gratton. Quasi static ensemble variational data assimilation.
Nonlinear Processes in Geophysics, 2017