A data-driven method for improving the correlation estimation in serial ensemble Kalman filter

Michèle De La Chevrotière,[1] John Harlim[2]

[1]Department of Mathematics, Penn State University, [2]Department of Mathematics and Department of Meteorology, Penn State University

The 7th EnKF Data Assimilation Workshop
May 2016

PennState
Eberly College of Science

Outline

1. Correlation statistics in EnKF
2. Learning algorithm for improving the correlation estimation
3. Serial EnKF with transformed correlations
4. Numerical experiments on the Lorenz-96 model
5. Conclusion

Correlation statistics in EnKF

EnKF in linear and Gaussian settings

- In the linear and Gaussian, and under certain assumptions (controllability, observability, uncorrelated noises), the Kalman posterior error covariance matrix P^a will converge to the fixed point $\hat{C}$ of

$$\hat{C} = (I - KH)(F\hat{C}F^\top + Q),$$

where $K = \hat{C}H^\top(H\hat{C}H^\top + R)^{-1}$ is the Kalman gain, F is the deterministic linear forecast model with noise covariance matrix Q, H is the linear observation operator, and R is the observation error covariance matrix.

EnKF in linear and Gaussian settings

- In the linear and Gaussian, and under certain assumptions (controllability, observability, uncorrelated noises), the Kalman posterior error covariance matrix P^a will converge to the fixed point $\hat{C}$ of

$$\hat{C} = (I - KH)(F\hat{C}F^\top + Q),$$

  where $K = \hat{C}H^\top(H\hat{C}H^\top + R)^{-1}$ is the Kalman gain, F is the deterministic linear forecast model with noise covariance matrix Q, H is the linear observation operator, and R is the observation error covariance matrix.
- EnKF construct an ensemble of analysis solutions $\{\boldsymbol{x}^{a,k}\}_{k=1}^{K}$ so that the ensemble mean and covariance *match the Kalman solution*.

EnKF in linear and Gaussian settings

- In the linear and Gaussian, and under certain assumptions (controllability, observability, uncorrelated noises), the Kalman posterior error covariance matrix P^a will converge to the fixed point $\hat{C}$ of

$$\hat{C} = (I - KH)(F\hat{C}F^\top + Q),$$

where $K = \hat{C}H^\top(H\hat{C}H^\top + R)^{-1}$ is the Kalman gain, F is the deterministic linear forecast model with noise covariance matrix Q, H is the linear observation operator, and R is the observation error covariance matrix.

- EnKF construct an ensemble of analysis solutions $\{\boldsymbol{x}^{a,k}\}_{k=1}^{K}$ so that the ensemble mean and covariance *match the Kalman solution*.

  Here $\boldsymbol{x}^a \in \mathbb{R}^N$, and K is the ensemble size.

EnKF in linear and Gaussian settings

- In the linear and Gaussian, and under certain assumptions (controllability, observability, uncorrelated noises), the Kalman posterior error covariance matrix P^a will converge to the fixed point $\hat{C}$ of

$$\hat{C} = (I - KH)(F\hat{C}F^\top + Q),$$

  where $K = \hat{C}H^\top(H\hat{C}H^\top + R)^{-1}$ is the Kalman gain, F is the deterministic linear forecast model with noise covariance matrix Q, H is the linear observation operator, and R is the observation error covariance matrix.
- EnKF construct an ensemble of analysis solutions $\{\boldsymbol{x}^{a,k}\}_{k=1}^K$ so that the ensemble mean and covariance *match the Kalman solution*.

  Here $\boldsymbol{x}^a \in \mathbb{R}^N$, and K is the ensemble size.
- The scheme approximates both prior and posterior covariances by the ensemble covariance matrix, e.g.,

$$P^a \approx \frac{1}{K-1}XX^T,$$

  where X is the matrix $\mathbb{R}^{N\times K}$ of analysis ensemble perturbations from the ensemble mean (the k^{th} column of X is $\boldsymbol{X}^{a,k} = \boldsymbol{x}^{a,k} - \overline{\boldsymbol{x}}^{a,k}$).

EnKF in linear and Gaussian settings

- In the linear and Gaussian, and under certain assumptions (controllability, observability, uncorrelated noises), the Kalman posterior error covariance matrix P^a will converge to the fixed point $\hat{C}$ of

$$\hat{C} = (I - KH)(F\hat{C}F^\top + Q),$$

  where $K = \hat{C}H^\top(H\hat{C}H^\top + R)^{-1}$ is the Kalman gain, F is the deterministic linear forecast model with noise covariance matrix Q, H is the linear observation operator, and R is the observation error covariance matrix.
- EnKF construct an ensemble of analysis solutions $\{\boldsymbol{x}^{a,k}\}_{k=1}^{K}$ so that the ensemble mean and covariance *match the Kalman solution*.

  Here $\boldsymbol{x}^a \in \mathbb{R}^N$, and K is the ensemble size.
- The scheme approximates both prior and posterior covariances by the ensemble covariance matrix, e.g.,

$$P^a \approx \frac{1}{K-1} X X^T,$$

  where X is the matrix $\mathbb{R}^{N\times K}$ of analysis ensemble perturbations from the ensemble mean (the k^{th} column of X is $\boldsymbol{X}^{a,k} = \boldsymbol{x}^{a,k} - \overline{\boldsymbol{x}}^{a,k}$).
- In the *linear and Gaussian* setting, the EnKF solution is *optimal* with finite ensemble if

$$\boldsymbol{x}_t^{a,k} - \overline{\boldsymbol{x}}_t^a \sim \mathcal{N}(0, \hat{C})$$

  in the limit as $t \to \infty$.

Motivation

- *Nonlinear and non-Gaussian case*: estimating the correlation $\rho = D^{-1/2} C D^{-1/2}$, $D = diag(C)$, is much more difficult because ρ is time-dependent and does not equilibrate to an invariant ρ_{eq}. Furthermore, the EnKF sample correlation, r^K, is such that
$$\hat{r}(t) \equiv \lim_{K \to \infty} r^K(t) \neq \rho(t), \quad \forall t.$$

Motivation

- *Nonlinear and non-Gaussian case*: estimating the correlation $\rho = D^{-1/2} C D^{-1/2}$, $D = diag(C)$, is much more difficult because ρ is time-dependent and does not equilibrate to an invariant ρ_{eq}. Furthermore, the EnKF sample correlation, r^K, is such that
$$\hat{r}(t) \equiv \lim_{K \to \infty} r^K(t) \neq \rho(t), \quad \forall t.$$
- Since estimating the true correlation ρ is very difficult, then instead, we try to estimate $\hat{r}$.

Motivation

- *Nonlinear and non-Gaussian case*: estimating the correlation $\rho = D^{-1/2} C D^{-1/2}$, $D = diag(C)$, is much more difficult because ρ is time-dependent and does not equilibrate to an invariant ρ_{eq}. Furthermore, the EnKF sample correlation, r^K, is such that

$$\hat{r}(t) \equiv \lim_{K \to \infty} r^K(t) \neq \rho(t), \quad \forall t.$$

- Since estimating the true correlation ρ is very difficult, then instead, we try to estimate $\hat{r}$.

How can we transform r^K into $\hat{r}$?

Motivation

- *Nonlinear and non-Gaussian case*: estimating the correlation $\rho = D^{-1/2} C D^{-1/2}$, $D = diag(C)$, is much more difficult because ρ is time-dependent and does not equilibrate to an invariant ρ_{eq}. Furthermore, the EnKF sample correlation, r^K, is such that

$$\hat{r}(t) \equiv \lim_{K \to \infty} r^K(t) \neq \rho(t), \quad \forall t.$$

- Since estimating the true correlation ρ is very difficult, then instead, we try to estimate $\hat{r}$.

How can we transform r^K into $\hat{r}$?

- The idea:
  I. Find a *linear map* $\mathcal{L}$ that transforms the undersampled correlation matrix r^K into a sample of improved correlation matrix:

$$\mathcal{L}^T r^K \approx \hat{r}, \qquad \mathcal{L} \in \mathbb{R}^{N \times N \times M}.$$

Motivation

- *Nonlinear and non-Gaussian case*: estimating the correlation $\rho = D^{-1/2} C D^{-1/2}$, $D = diag(C)$, is much more difficult because ρ is time-dependent and does not equilibrate to an invariant ρ_{eq}. Furthermore, the EnKF sample correlation, r^K, is such that

$$\hat{r}(t) \equiv \lim_{K \to \infty} r^K(t) \neq \rho(t), \quad \forall t.$$

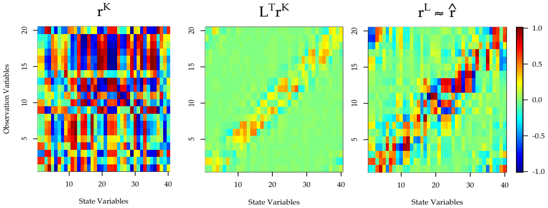
- Since estimating the true correlation ρ is very difficult, then instead, we try to estimate $\hat{r}$.

How can we transform r^K into $\hat{r}$?

- The idea:
  I. Find a *linear map* $\mathcal{L}$ that transforms the undersampled correlation matrix r^K into a sample of improved correlation matrix:

$$\mathcal{L}^T r^K \approx \hat{r}, \qquad \mathcal{L} \in \mathbb{R}^{N \times N \times M}.$$

$r^L \approx \hat{r}$
EnKF L=500



Motivation

- *Nonlinear and non-Gaussian case*: estimating the correlation $\rho = D^{-1/2} C D^{-1/2}$, $D = diag(C)$, is much more difficult because ρ is time-dependent and does not equilibrate to an invariant ρ_{eq}. Furthermore, the EnKF sample correlation, r^K, is such that

$$\hat{r}(t) \equiv \lim_{K \to \infty} r^K(t) \neq \rho(t), \quad \forall t.$$

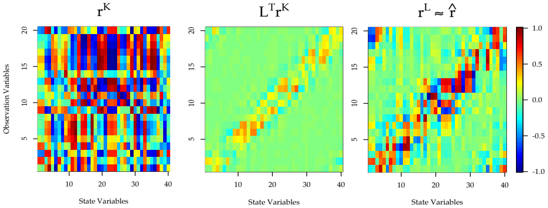
- Since estimating the true correlation ρ is very difficult, then instead, we try to estimate $\hat{r}$.

How can we transform r^K into $\hat{r}$?

- The idea:
  I. Find a *linear map* $\mathcal{L}$ that transforms the undersampled correlation matrix r^K into a sample of improved correlation matrix:

$$\mathcal{L}^T r^K \approx \hat{r}, \qquad \mathcal{L} \in \mathbb{R}^{N \times N \times M}.$$

$r^L \approx \hat{r}$
EnKF L=500



II. Implement the map $\mathcal{L}$ into the *serial LLS EnKF* of Anderson (2003).

Learning algorithm for improving the correlation estimation

A minimization problem

- For nonlinear filtering problems, where $h(x) : \mathbb{R}^N \to \mathbb{R}^M$, we consider the sample *cross correlation:*

$$r_{\boldsymbol{xy}}^K = D_X^{-1/2}\Big(\frac{1}{K-1}XY^\top\Big)D_Y^{-1/2} \in \mathbb{R}^{N\times M}.$$

[The K^{th} column of Y is $\boldsymbol{Y}_k = h(\boldsymbol{x}^{a,k}) - \overline{\boldsymbol{y}}^a$, where $\overline{\boldsymbol{y}}^a = \sum_{k=1}^K h(\boldsymbol{x}^{a,k})/K$]

A minimization problem

- For nonlinear filtering problems, where $h(x) : \mathbb{R}^N \to \mathbb{R}^M$, we consider the sample *cross correlation:*

$$r_{\boldsymbol{xy}}^K = D_X^{-1/2}\left(\frac{1}{K-1} XY^\top\right) D_Y^{-1/2} \in \mathbb{R}^{N\times M}.$$

[The K^{th} column of Y is $\boldsymbol{Y}_k = h(\boldsymbol{x}^{a,k}) - \overline{\boldsymbol{y}}^a$, where $\overline{\boldsymbol{y}}^a = \sum_{k=1}^K h(\boldsymbol{x}^{a,k})/K$]

Onward: suppress the subscript $\boldsymbol{xy}$ (write $r_{\boldsymbol{xy}}^K$ as r^K; $\hat{r} = \lim_{K\to\infty} r^K$).

A minimization problem

- For nonlinear filtering problems, where $h(x) : \mathbb{R}^N \to \mathbb{R}^M$, we consider the sample *cross correlation:*

$$r^K_{\boldsymbol{xy}} = D_X^{-1/2}\Big(\frac{1}{K-1} XY^\top\Big) D_Y^{-1/2} \in \mathbb{R}^{N\times M}.$$

  [The K^{th} column of Y is $\boldsymbol{Y}_k = h(\boldsymbol{x}^{a,k}) - \overline{\boldsymbol{y}}^a$, where $\overline{\boldsymbol{y}}^a = \sum_{k=1}^K h(\boldsymbol{x}^{a,k})/K$]

  Onward: suppress the subscript $\boldsymbol{xy}$ (write $r^K_{\boldsymbol{xy}}$ as r^K; $\hat{r} = \lim_{K\to\infty} r^K$).

- We seek to find a linear map $\mathcal{L}$ such that *for every pair* (i,j), $\mathcal{L}(\cdot, i, j)$ takes the sample correlations $r^K(\cdot, y_j)$ to the limiting correlation $\hat{r}(x_i, y_j)$.

A minimization problem

- For nonlinear filtering problems, where $h(x) : \mathbb{R}^N \to \mathbb{R}^M$, we consider the sample *cross correlation:*

$$r_{\boldsymbol{xy}}^K = D_X^{-1/2}\left(\frac{1}{K-1} XY^\top\right) D_Y^{-1/2} \in \mathbb{R}^{N\times M}.$$

[The K^{th} column of Y is $\boldsymbol{Y}_k = h(\boldsymbol{x}^{a,k}) - \overline{\boldsymbol{y}}^a$, where $\overline{\boldsymbol{y}}^a = \sum_{k=1}^K h(\boldsymbol{x}^{a,k})/K$]

Onward: suppress the subscript $\boldsymbol{xy}$ (write $r_{\boldsymbol{xy}}^K$ as r^K; $\hat{r} = \lim_{K\to\infty} r^K$).

- We seek to find a linear map $\mathcal{L}$ such that *for every pair* (i,j), $\mathcal{L}(\cdot, i, j)$ takes the sample correlations $r^K(\cdot, y_j)$ to the limiting correlation $\hat{r}(x_i, y_j)$.

  Inspired by Anderson (2012), we formulate:

$$\min_{\mathcal{L}(\cdot,i,j)} \int_{[-1,1]\times[-1,1]} \left(r^K(\cdot, y_j)^\top \mathcal{L}(\cdot, i, j) - \hat{r}(x_i, y_j)\right)^2 p(r^K|\hat{r})p(\hat{r})\, dr^K\, d\hat{r}.$$

A minimization problem

- For nonlinear filtering problems, where $h(x) : \mathbb{R}^N \to \mathbb{R}^M$, we consider the sample *cross correlation:*

$$r^K_{\boldsymbol{xy}} = D_X^{-1/2}\Big(\frac{1}{K-1} XY^\top\Big) D_Y^{-1/2} \in \mathbb{R}^{N\times M}.$$

[The K^{th} column of Y is $\boldsymbol{Y}_k = h(\boldsymbol{x}^{a,k}) - \overline{\boldsymbol{y}}^a$, where $\overline{\boldsymbol{y}}^a = \sum_{k=1}^K h(\boldsymbol{x}^{a,k})/K$]

Onward: suppress the subscript $\boldsymbol{xy}$ (write $r^K_{\boldsymbol{xy}}$ as r^K; $\hat{r} = \lim_{K\to\infty} r^K$).

- We seek to find a linear map $\mathcal{L}$ such that *for every pair (i,j)*, $\mathcal{L}(\cdot,i,j)$ takes the sample correlations $r^K(\cdot, y_j)$ to the limiting correlation $\hat{r}(x_i, y_j)$.

  Inspired by Anderson (2012), we formulate:

$$\min_{\mathcal{L}(\cdot,i,j)} \int_{[-1,1]\times[-1,1]} \big(r^K(\cdot,y_j)^\top \mathcal{L}(\cdot,i,j) - \hat{r}(x_i,y_j)\big)^2 p(r^K|\hat{r})p(\hat{r})\, dr^K\, d\hat{r}.$$

- In practice, can use any reliable correlation statistics r^L to approximate the limiting correlation $\hat{r}$, so that $r^L \sim p(\hat{r})$. (e.g. EnKF with large ensemble L).

Learning the map $\mathcal{L}$

1. *Train offline:* Generate an OSSE assimilation analysis ensemble $\{\boldsymbol{x}_t^{a,k}\}_{t,k=1}^{T,L}$ with $L \gg 1$ and T assimilation cycles.

Learning the map $\mathcal{L}$

1. *Train offline:* Generate an OSSE assimilation analysis ensemble $\{\boldsymbol{x}_t^{a,k}\}_{t,k=1}^{T,L}$ with $L \gg 1$ and T assimilation cycles.
2. Compute $\{r_t^L\}_{t=1}^T$, so that $r_t^L \sim p(\hat{r})$.

Learning the map $\mathcal{L}$

1. *Train offline:* Generate an OSSE assimilation analysis ensemble $\{\boldsymbol{x}_t^{a,k}\}_{t,k=1}^{T,L}$ with $L \gg 1$ and T assimilation cycles.
2. Compute $\{r_t^L\}_{t=1}^T$, so that $r_t^L \sim p(\hat{r})$.
3. Compute $\{r_t^K\}_{t=1}^T$, subsampling K out of L members, so that $r_t^K \sim p(r^K | \hat{r} = r_t^L)$.

Learning the map $\mathcal{L}$

1. *Train offline:* Generate an OSSE assimilation analysis ensemble $\{\boldsymbol{x}_t^{a,k}\}_{t,k=1}^{T,L}$ with $L \gg 1$ and T assimilation cycles.
2. Compute $\{r_t^L\}_{t=1}^T$, so that $r_t^L \sim p(\hat{r})$.
3. Compute $\{r_t^K\}_{t=1}^T$, subsampling K out of L members, so that $r_t^K \sim p(r^K | \hat{r} = r_t^L)$.
4. Monte Carlo approximation to the minimization problem:

$$\min_{\mathcal{L}(\cdot,i,j)} \frac{1}{T} \sum_{t=1}^{T} \Big(\underbrace{r_t^K(\cdot, y_j)^\top}_{\text{row of } A} \underbrace{\mathcal{L}(\cdot, i, j)}_{\boldsymbol{u}} - \underbrace{r_t^L(x_i, y_j)}_{\text{entry of } \boldsymbol{b}} \Big)^2 .$$

or

$$\text{LSP}: \qquad \min_{\boldsymbol{u}} \|A\boldsymbol{u} - \boldsymbol{b}\|^2 \text{ when } \boldsymbol{u} = (A^\top A)^{-1} A^\top \boldsymbol{b}$$

Learning the map $\mathcal{L}$

1. *Train offline:* Generate an OSSE assimilation analysis ensemble $\{\boldsymbol{x}_t^{a,k}\}_{t,k=1}^{T,L}$ with $L \gg 1$ and T assimilation cycles.
2. Compute $\{r_t^L\}_{t=1}^T$, so that $r_t^L \sim p(\hat{r})$.
3. Compute $\{r_t^K\}_{t=1}^T$, subsampling K out of L members, so that $r_t^K \sim p(r^K | \hat{r} = r_t^L)$.
4. Monte Carlo approximation to the minimization problem:

$$\min_{\mathcal{L}(\cdot,i,j)} \frac{1}{T} \sum_{t=1}^{T} \Big(\underbrace{r_t^K(\cdot, y_j)^\top}_{\text{row of } A} \underbrace{\mathcal{L}(\cdot, i, j)}_{\boldsymbol{u}} - \underbrace{r_t^L(x_i, y_j)}_{\text{entry of } \boldsymbol{b}} \Big)^2.$$

or

$$\text{LSP}: \qquad \min_{\boldsymbol{u}} \|A\boldsymbol{u} - \boldsymbol{b}\|^2 \text{ when } \boldsymbol{u} = (A^\top A)^{-1} A^\top \boldsymbol{b}$$

- Offline training, little computational overhead, almost no tuning (K, T, L).

Learning the map $\mathcal{L}$

1. *Train offline:* Generate an OSSE assimilation analysis ensemble $\{\boldsymbol{x}_t^{a,k}\}_{t,k=1}^{T,L}$ with $L \gg 1$ and T assimilation cycles.
2. Compute $\{r_t^L\}_{t=1}^T$, so that $r_t^L \sim p(\hat{r})$.
3. Compute $\{r_t^K\}_{t=1}^T$, subsampling K out of L members, so that $r_t^K \sim p(r^K | \hat{r} = r_t^L)$.
4. Monte Carlo approximation to the minimization problem:

$$\min_{\mathcal{L}(\cdot,i,j)} \frac{1}{T} \sum_{t=1}^{T} \Big(\underbrace{r_t^K(\cdot, y_j)^\top}_{\text{row of } A} \underbrace{\mathcal{L}(\cdot, i, j)}_{\boldsymbol{u}} - \underbrace{r_t^L(x_i, y_j)}_{\text{entry of } \boldsymbol{b}} \Big)^2.$$

or

$$\text{LSP}: \qquad \min_{\boldsymbol{u}} \|A\boldsymbol{u} - \boldsymbol{b}\|^2 \text{ when } \boldsymbol{u} = (A^\top A)^{-1} A^\top \boldsymbol{b}$$

- Offline training, little computational overhead, almost no tuning (K, T, L).
- $\mathcal{L}$ **is a scalar**: the map found can be viewed as a type of *data-driven localization function* ($\mathcal{L}_d$).

Serial EnKF with transformed correlations

Adapted serial LLS EnKF (Anderson, 2013)

LLS EnKF assimilates observations $j = 1, \ldots, M$ one at a time:

1 **for** $j = 1$ **to** M **do** `/* Loop over observations */`

2 $\quad \overline{y}_j^a = \overline{y}_j^b + (P^b_{y_j y_j} + R_{jj})^{-1} P^b_{y_j y_j} (y_j^o - \overline{y}_j^b)$;

3 $\quad y_j^{a,k} = \overline{y}_j^a + \sqrt{\frac{R_{jj}}{R_{jj} + P^b_{y_j y_j}}} (y_j^{b,k} - \overline{y}_j^b)$;

4 $\quad \Delta y_j^k = y_j^{a,k} - y_j^{b,k}$; `/* compute the obs update */`

5 $\quad \boldsymbol{x}^{a,k} = \boldsymbol{x}^{b,k} + \frac{P^b_{\boldsymbol{x} y_j}}{P^b_{y_j y_j}} \Delta y_j^k$; `/* regresses the update onto` $\boldsymbol{x}$ `*/`

6 **end**

The straightforward modification is to replace in line 5:

$$P^b_{\boldsymbol{x} y_j} = D_X^{1/2} r(\cdot, y_j) D_Y^{1/2} \longleftarrow D_X^{1/2} \mathcal{L}(\cdot, \cdot, j)^\top r(\cdot, y_j) D_Y^{1/2},$$

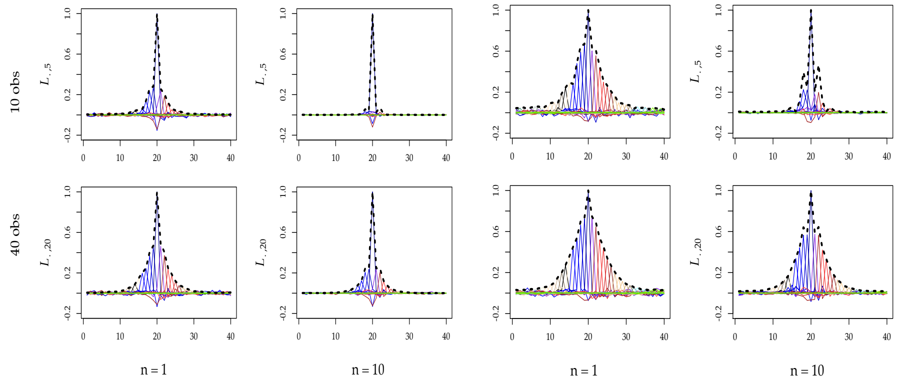
where D_X, D_Y and r are estimated form the EnKF with ensemble size K.

Numerical experiments on the Lorenz-96 model

Linear direct observations

Maps $\mathcal{L}$ and $\mathcal{L}_d$ found by regressing onto ETKF L500

N = 40 state variables, 10 (top) and 40 (bottom) observations
observation time steps 1 (left) and 10 (right)

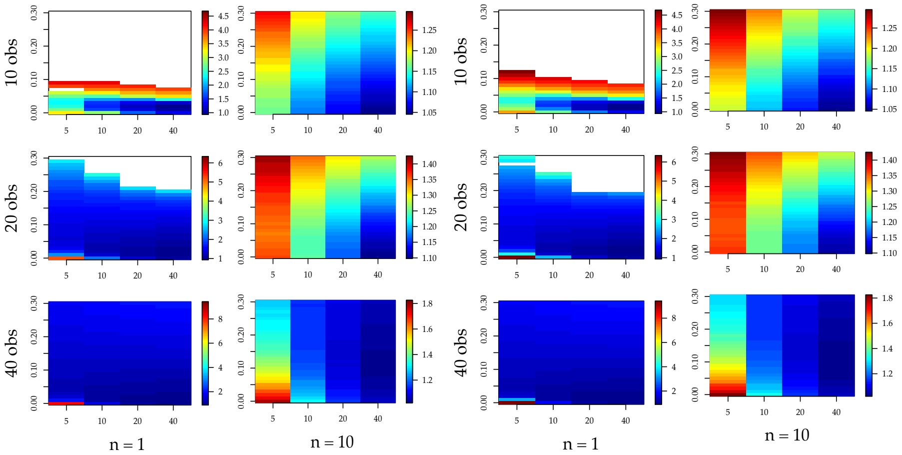


(a) 5 ensemble members. (b) 40 ensemble members.

Localized structure of the mappings; $\mathcal{L}_d$ (dashed black) is an envelope for the mappings $\mathcal{L}$.

Linear direct observations

Filtering results with adapted LLS EnKF + covariance inflation



(c) $\mathcal{L}$.

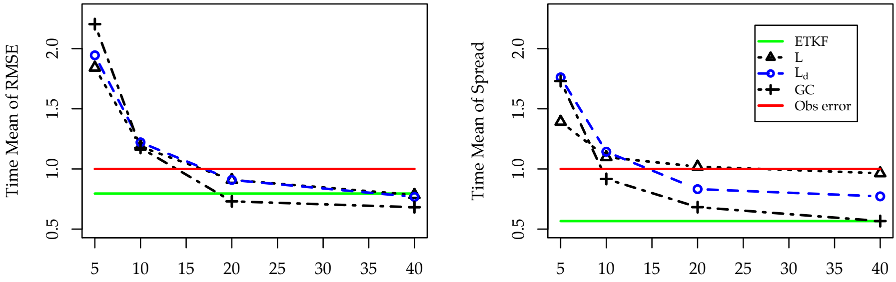
(d) $\mathcal{L}_d$.

Time mean RMSE normalized by the RMSE of the ETKF using 500 ensemble members shown for the serial EnKF using $\mathcal{L}$ and $\mathcal{L}_d$, plotted against inflation factor and K

Linear direct observations

Filtering results – Best tuned inflation values

10 observations, observation time step $n = 1$, comparison of 3 localization methods



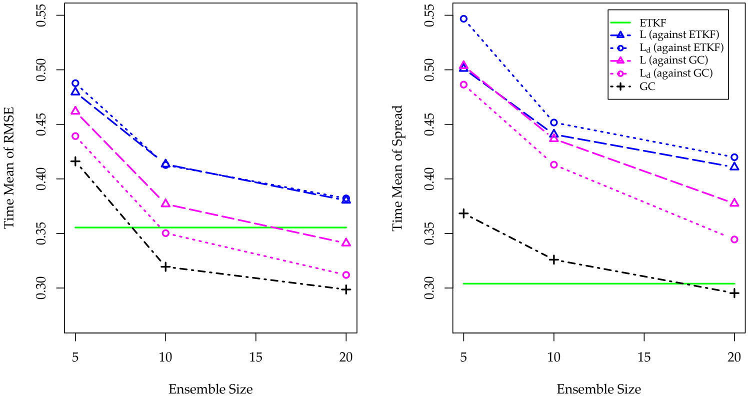
Time mean of RMSE (left) and spread (right) as a function of ensemble size

GC is more performant for $K > 10$ but is extensively tuned!

Linear direct observations

Regression using different products

20 observations, observation time step $n = 1$, all use best tuned inflation values

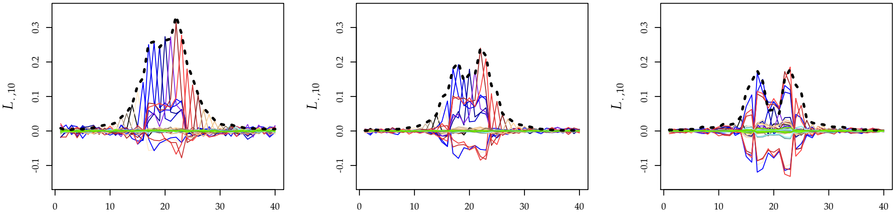


Two choices of *regression products*: (1) ETKF with 500 ensemble members and (2) serial EnKF with GC localization with optimal half-width.

Linear indirect observations

Maps $\mathcal{L}$ and $\mathcal{L}_d$ found by regressing onto ETKF L500

20 observations, 5 ensemble members

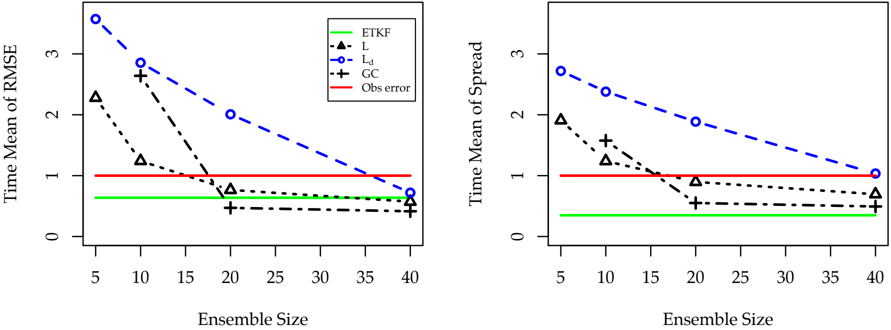


n = 1   n = 5   n = 10

Linear indirect observations

Filtering results with localization and best tuned inflation values

20 observations, observation time step $n = 5$



Time mean of RMSE (left) and spread (right)

Summary

- A data-driven method for improving the correlation estimation in serial ensemble Kalman filter is introduced.
- The method finds a linear map that transforms, at each assimilation cycle, the poorly estimated sample correlation into a sample of improved correlation.
- This map is obtained from an offline training procedure with almost no tuning as the solution of a linear regression problem that uses appropriate sample correlation statistics obtained from historical data assimilation product.
- In numerical tests with the Lorenz-96 model for ranges of cases of linear and nonlinear observation models, the proposed scheme improves the filter estimates, especially when ensemble size is small relative to the dimension of the state space.

Reference: M. De La Chevrotière, J. Harlim, *A data-driven method for improving the correlation estimation in serial ensemble Kalman filter* (2016), Mon. Wea. Rev., submitted.