Data Assimilation with a Climatologically Augmented Local Ensemble Transform Kalman Filter

Matthew Kretschmer$^1$ Brian Hunt$^2$ Edward Ott$^{1,3}$

1 Department of Physics
University of Maryland, College Park

2 Department of Mathematics
University of Maryland, College Park

3 Department of Electrical Engineering
University of Maryland, College Park

Weather Chaos Group Meeting
Outline

1. Motivation
2. The climatologically augmented Local Ensemble Transform Kalman Filter (caLETKF)
3. Numerical Experiments
4. Results and Conclusions
Can there be hybrid-like data assimilation in a pure ensemble framework?

- Hybrid methods combine static and flow-dependent background-error covariance matrices.
  - This helps boost the rank of the background-error covariance estimate.
- The method we propose -
  - Was inspired by Principal Component Analysis.
  - Works by performing the analysis on an enlarged (augmented) ensemble.
- Is it possible to avoid ensemble collapse?
- Note - We focus here on applying these ideas to the LETKF, but believe that these ideas may be readily applied to other EnKF formulations.
Perturbations enhance the ensemble in directions which are not represented, or are not represented with enough variance.

The ensemble’s rank is increased during the analysis.

**T - Additional Ens. Perturbation**
Enlarging the Ensemble at Analysis Time

- Forecast the dynamic ensemble between analysis times.
- A $k_d$-member dynamic ensemble can be expressed as:
  - A mean: $\bar{x}_d$
  - A matrix of perturbations: $\hat{X}_d = [\hat{x}_d^1 | \ldots | \hat{x}_d^{k_d}]$
- Generate additional $k_c$ constant, “climatological” ensemble perturbations $\hat{X}_c = [\hat{x}_c^1 | \ldots | \hat{x}_c^{k_c}]$
- Perform analysis on the ensemble with mean $\bar{x}_d$ and perturbations $\hat{X} = [\hat{X}_d \hat{X}_c]$
- Forecast $k_d$ of the $k_d + k_c$ analysis ensemble members to the next analysis time.
Choosing the Analysis Ensemble Members

- How should the $k_d$-member analysis dynamic ensemble be chosen?
- We construct the dynamic analysis ensemble from the “first” $k_d$ ensemble members.
- The ensemble mean and perturbations are updated using:
  1. Updated mean: $\bar{x}^a = \bar{x}^b + \hat{X}^b \bar{w}^a$.
  2. Updated perturbations: $\hat{X}^a = \hat{X}^b W^a$.
- The dynamic analysis perturbations are chosen from the first $k_d$ analysis perturbations (columns of $\hat{X}^a$), and re-centered.
- The $i$th analysis dynamic ensemble member is given by $x^{a,i} = \bar{x}^a + \hat{x}^{a,i}$.
Generating the Climatological Perturbations $\hat{X}_c$

- Derive climatological perturbations from a static background error covariance matrix $\mathbf{B}$
- $\mathbf{B}$ can be eigendecomposed as: $\mathbf{B} = \mathbf{A}\mathbf{A}^T$, with $\mathbf{A} = \mathbf{VD}^{1/2}$
  - $\mathbf{V}$'s columns hold orthonormal eigenvectors of $\mathbf{B}$
  - $\mathbf{D}$ is diagonal and holds the eigenvalues of $\mathbf{B}$.
- Select the $k_c$ columns of $\mathbf{A}$ that correspond to the largest eigenvalues of $\mathbf{B}$.
- Subtract mean of these columns to form $\hat{X}_c$
- Eigendecomposition only needs to happen once!
Numerical Experiments with the Lorenz 240-Variable Model

Model dynamics are given by \[ \frac{dZ_n}{dt} = [Z, Z]_{8,n} - Z_n + 15 \]

Static-in-time, homogeneous observation network: 12 observations assimilated each cycle, \( R = I_{12 \times 12} \)

Compare caLETKF results to LETKF results
  - Localization radius - 20 model grid points.

Background covariance matrix \( B \) generated with NMC method \(^1\).

Used constant covariance inflation \(^2\)
  - caLETKF used 2.75% inflation.
  - LETKF used 3% inflation

\(^1\) Parrish and Derber (1992)
\(^2\) Anderson and Anderson (1999)
\(^3\) Lorenz (2005)
The effect of the caLETKF on analysis accuracy

- Plot analysis accuracy as a function of dynamic ensemble size $k_d$.
- caLETKF converges faster than the LETKF by $\approx 10$ ensemble members.
What if we want to decrease forecasting costs for a given system?

Total ensemble stays constant, $k_d + k_c = 30$.

caLETKF with $k_d = 15$ rivals LETKF with $k_d = 30$. 
The caLETKF at constant forecasting cost

- How many climatological ensemble members will help at a given dynamic ensemble size?
- The dynamic ensemble (and hence forecasting “costs”) stays constant.
The (ensemble) forecast accuracy of the caLETKF

- LETKF and caLETKF only forecast $k_d = 20$ ensemble members.
- More accurate analysis ensemble $\rightarrow$ more accurate ensemble forecasts.
- Suggests a gain of nearly a day in forecast lead times.

$k_d = 20, k_c = 10$

| Lead Time (model hours) | LETKF | caLETKF |
|-------------------------|-------|---------|
| 0                       |       |         |
| 10                      |       |         |
| 20                      |       |         |
| 30                      |       |         |
| 40                      |       |         |
| 50                      |       |         |
| 60                      |       |         |
| 70                      |       |         |

Forecast Mean RMSE

February 2, 2015 12 / 18
Summary

- New method incorporates climatological information into LETKF, and allows for smaller dynamic ensemble size without loss of accuracy.
- This approach may appeal to users of pure ensemble frameworks without resources for large ensembles or existing variational codes (to implement traditional hybrids).

Open Questions

- Can the climatological perturbations alleviate sampling error and allow broader localization?
- Can the caLETKF effectively treat model error?
- Will such dramatic benefits be observed in a more complex model?
Thank you
Anderson, J. L. and Anderson, S. L., 1999. A monte carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.* **127**, 2741–2758.

Lorenz, E. N., 2005. Designing chaotic models. *J. Atmos. Sci.* **62**, 1574–1587.

Parrish, D. F. and Derber, J. C., 1992. The National Meteorological Center’s spectral statistical-interpolation analysis system. *Mon. Wea. Rev.* **120**, 1747–1763.
Implicit Background Covariances

- Traditional LETKF implicitly takes the background covariance to be \( P_d^b = \frac{1}{k_d - 1} \hat{X}_d \hat{X}_d^T \).

- Here in the caLETKF, the ensemble perturbations are given by \( \hat{X} = [\hat{X}_d \ \hat{X}_c] \).

- As a result, the implicit background covariance seen by the algorithm is actually a sum of dynamic and static contributions:
  \[
P^b = \frac{1}{k_d + k_c - 1} [(k_d - 1)P_d^b + k_c P_c^b]
  \]

- Choice here of \( k_c \) made because these aren’t random samples that we’re taking, necessarily. (We model climatological background-error covariance as \( P_c^b = \frac{1}{k_c} \hat{X}_c \hat{X}_c^T \)).

- To account for deficiencies in estimating \( B \), and also to maintain the contributions of each column of \( \hat{X}_c \), we scale these columns by a tuning factor \( \sqrt{\alpha} \).
How correlated are the climatological perturbations $\hat{x}_c$?

A plot of the “angle” between the first ensemble perturbation and each of the next 15 climatological perturbations.

The angle $\theta$ that is plotted is found via

$$\theta_{1,j} = \cos^{-1}\left(\frac{\hat{x}_c^1 \cdot \hat{x}_c^j}{\|\hat{x}_c^1\| \|\hat{x}_c^j\|}\right)$$
Side note: Subtracting the mean (2/2)

- How physically different are the perturbations look after subtracting off their mean??
- Red curve shows a column of $A$, black shows corresponding $\hat{x}_c$