CS6404: Computational Data Assimilation

Project 1

Sam Donald

October 17, 2022

1 Project Outline

This project applies various versions of Kalman filters to the Lorenz-96 data assimilation problem, whereby restricted measurements are merged with prior knowledge over time resulting in improved estimates of the system state. The report is structured as follows, section 2 provides descriptions of the filters and evaluation plots, with the plots discussed within section 3. Additionally, the process of generating the specific Lorenz-96 trajectories is described in full within appendix A.

2 Applied Methods and Results

The following section outlines multiple filters and methods applied to solve the Lorenz-96 data assimilation problem. All filters are evaluated through a rank histogram (of the first state) and an RMSE plot over time.

2.1 EnKF

A baseline Ensemble Kalman filter (EnKF) was developed and is provided in the associated donald_EnKF.m file, with corresponding results outlined by fig. 1.

2.2 ETKF

An Ensemble Transform Kalman filter (ETKF) was developed and is provided in the associated donald_ETKF.m file, with corresponding results outlined by fig. 4. A symmetric square root transform was chosen, with the process of calculating the transform matrix $\boldsymbol{T}_i$ described by eqs. (1) to (3)

$$\dot{\boldsymbol{Z}}_i^{bT} \boldsymbol{R}_i^{-1} \dot{\boldsymbol{Z}}_i^{b} = \boldsymbol{Q}_i \boldsymbol{\Gamma}_i \boldsymbol{Q}_i^T \in \Re^{N_{ens} \times N_{ens}} \tag{1}$$

$$\boldsymbol{T}_i \boldsymbol{T}_i^T = (\boldsymbol{I}_{N_{ens}} + \dot{\boldsymbol{Z}}_i^{bT} \boldsymbol{R}_i^{-1} \dot{\boldsymbol{Z}}_i^{b})^{-1} = \boldsymbol{Q}_i (\boldsymbol{I}_{N_{ens}} + \boldsymbol{\Gamma}_i)^{-1} \boldsymbol{Q}_i^T \tag{2}$$

$$\boldsymbol{T}_i = \boldsymbol{Q}_i (\boldsymbol{I}_{N_{ens}} + \boldsymbol{\Gamma}_i)^{-1/2} \boldsymbol{Q}_i^T \tag{3}$$

2.3 Localization

2.3.1 EnKF Covariance Localization

Covariance localization was applied to the EnKF via the Gaspari-Cohn function (with decorrelation parameter $L = 5$). This function is used to capture variable correlation decreasing with distance, with corresponding results outlined by fig. 2.

2.3.2 LETKF

As the ETKF does not explicitly calculate the covariance matrix $\boldsymbol{P}$, a state-centric localization method is instead applied via the local ensemble Kalman filter (LETKF). Subsets of observations L_l are typically created based on proximity to the variable $x_{i;l}$, yet for simplicity states are instead grouped into bins of size 10 such that states 1-10, 11-20, 21-30 and 31-40 are paired, along with their associated observations. The standard ETKF processes is then applied to each ensemble subset, along with the inclusion of mean weights outlined by eqs. (4) and (5). Results of the LETKF are displayed within fig. 5

$$\overline{w}_i^a\{l\} = \dot{\boldsymbol{Z}}_i^b(L,:)^T(\dot{\boldsymbol{Z}}_i^b(L,:)\dot{\boldsymbol{Z}}_i^b(L,:)^T + \boldsymbol{R}_i(L,L))^{-1}(\boldsymbol{y}_i - H(\overline{x}_i^b))_L \tag{4}$$

$$\dot{\boldsymbol{X}}_i^a(l,:) = \dot{\boldsymbol{X}}_i^b(l,:)(\overline{w}_i^a\{l\} + \boldsymbol{T}_i\{l\}(:,e)), \qquad e = 1, ..., N_{ens} \tag{5}$$

Note that $\boldsymbol{T}_i\{l\}$ is calculated according to eqs. (1) to (3) using states l captured within subset L and their associated covariance $\boldsymbol{R}_i(L,L)$).

2.4 Covariance Inflation

Multiplicative covariance inflation is implemented for both the localized EnKF and LETKF through scaling of their respective ensemble covariance and ensemble of anomalies. This is accomplished through eq. (6) for the EnKF, and eq. (7) for the LETKF as a square root transform is used. Results of covariance inflation are displayed within fig. 6.

$$\boldsymbol{P_i^b} \leftarrow \boldsymbol{P_i^b}\alpha^2 \tag{6}$$

$$\dot{\boldsymbol{X}}_{\boldsymbol{i}}^{\boldsymbol{b}}\dot{\boldsymbol{X}}_{\boldsymbol{i}}^{\boldsymbol{b}T} = \boldsymbol{P_i^b} \implies \dot{\boldsymbol{X}}_{\boldsymbol{i}}^{\boldsymbol{b}} \leftarrow \dot{\boldsymbol{X}}_{\boldsymbol{i}}^{\boldsymbol{b}}\alpha \tag{7}$$

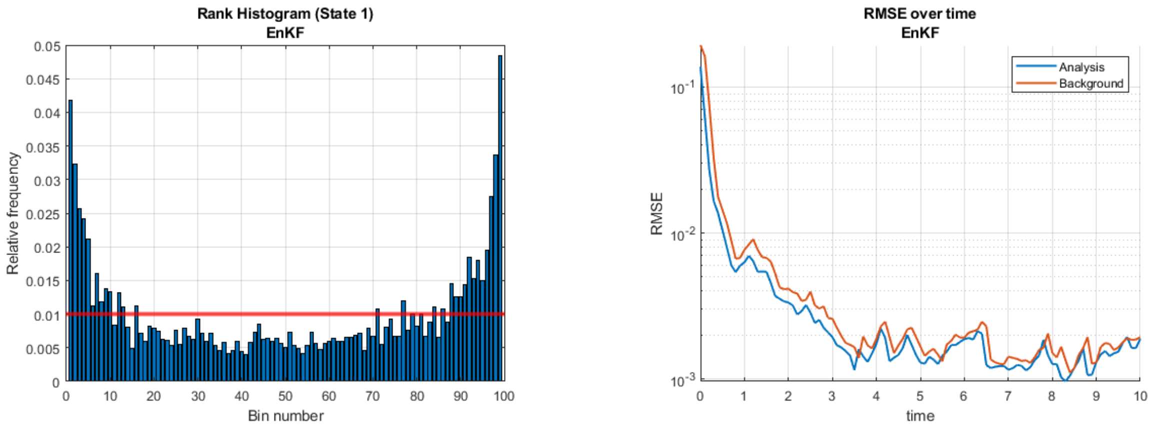

Figure 1: EnKF rank histogram (left) and RMSE progression (right)

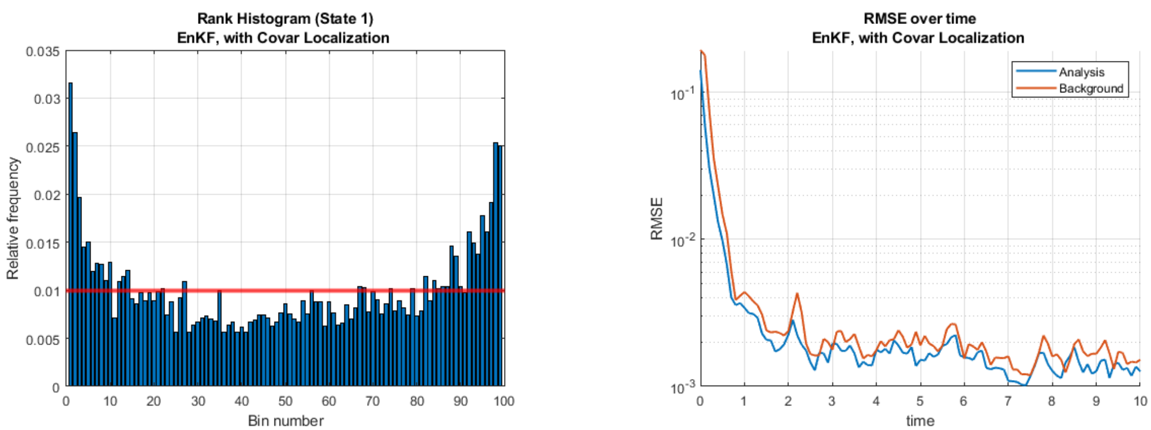

Figure 2: EnKF with localization, rank histogram (left) and RMSE progression (right)

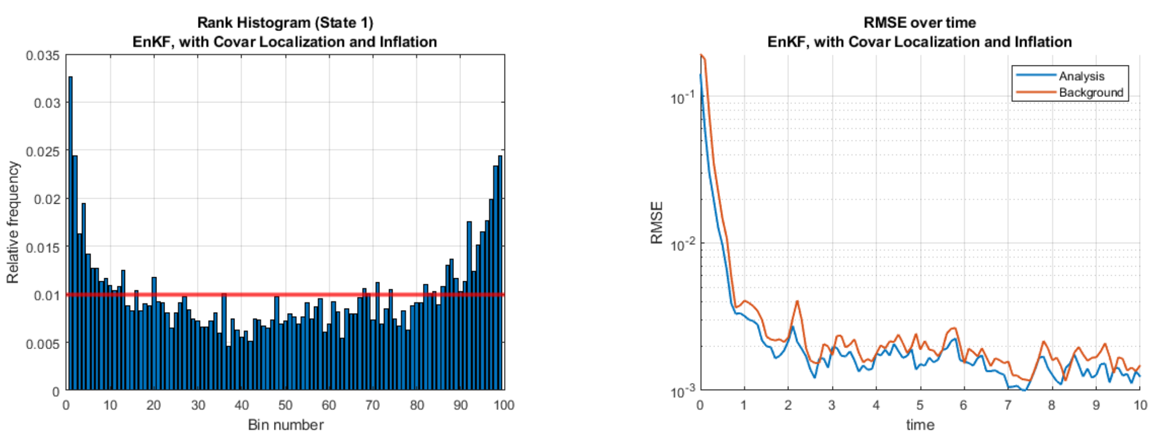

Figure 3: EnKF with localization and inflation, rank histogram (left) and RMSE progression (right)

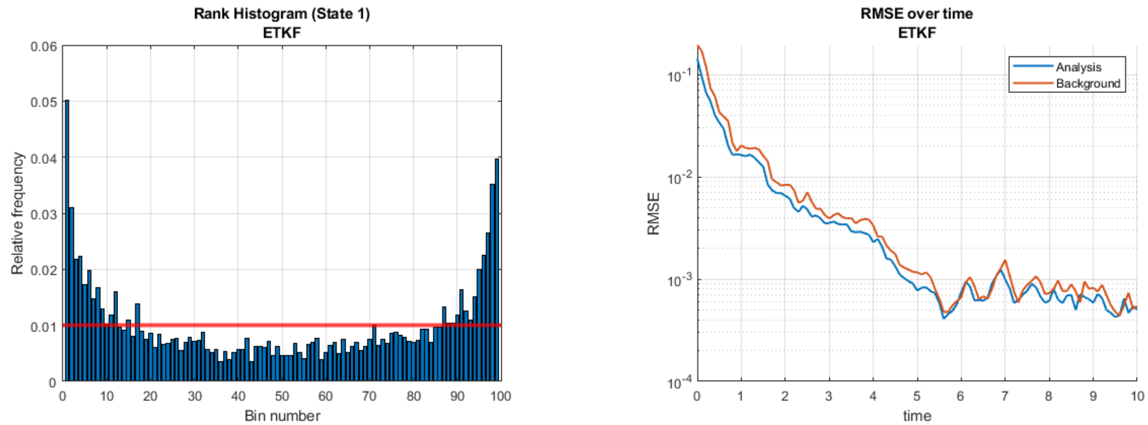

Figure 4: ETKF rank histogram (left) and RMSE progression (right)

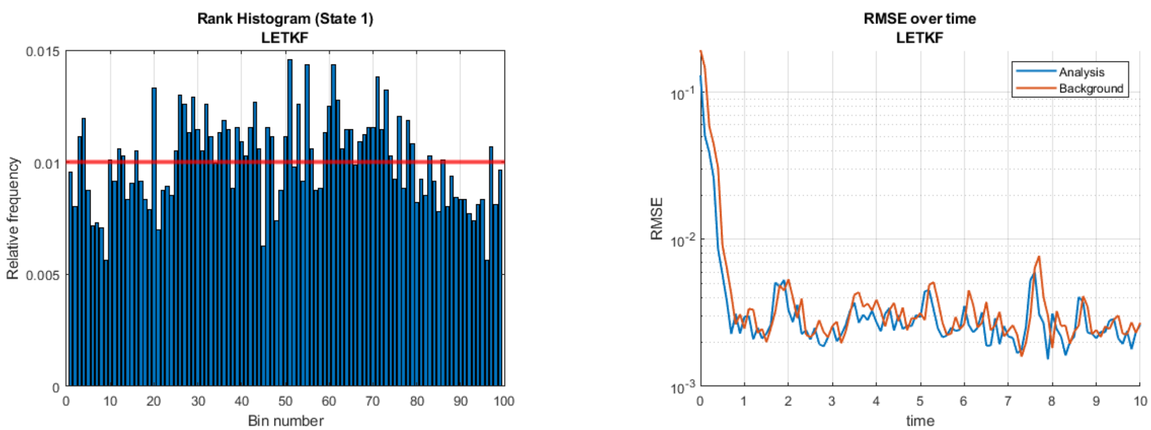

Figure 5: LETKF rank histogram (left) and RMSE progression (right)

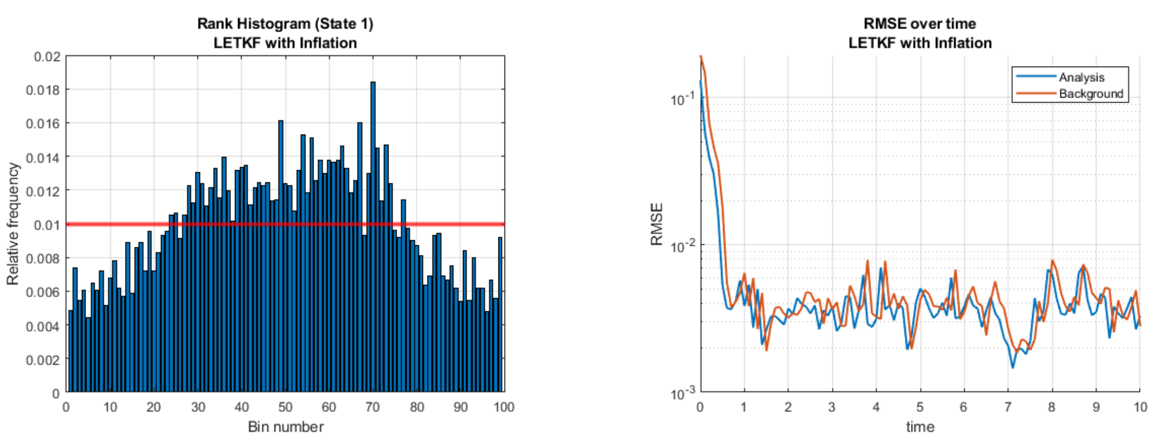

Figure 6: LETKF with inflation, rank histogram (left) and RMSE progression (right)

3 Discussion

3.1 Overall performance

All of the filters implemented successfully and consistently solve the posed data assimilation problem. The RMSE plots display stabilizing RMSE values for all filters at a value of 0.002 ± 0.001 after T_{max} time units. Additionally there is a clear discrepancy between the background and analysis RMSE values, with the difference attributed to the successful assimilation of observation data.

3.2 EnKF

With respect to the EnKF, the rank histogram for the baseline filter within fig. 1 is clearly concave and the ensemble therefore under-dispersive. The addition of covariance localization increases ensemble dispersion slightly and increases the rate at which RMSE decreases during the assimilation process fig. 2. Finally the addition of covariance inflation using $\alpha = 1.02$ has a very slight positive impact on the ensemble spread fig. 3, implying that the filter would likely benefiting from a higher value of α.

3.3 ETKF

With respect to the ETKF, both RMSE and the rank histogram fig. 4 are similar to that of the baseline EnKF fig. 1 as expected, with the noted difference of ETKF using significantly less memory as the covariance matrix P is not calculated. Applying localization in the form of an LETKF dramatically improves performance as demonstrated by the uniform rank histogram and rapidly decreasing RMSE fig. 5. Applying covariance inflation to the already uniform LETKF rank histogram has the expected negative impact of increasing ensemble spread, and making the ensemble over-dispersive.

3.4 Limitations

The results of these filters are directly related to paramaters outlined within appendix A. Some values have been arbitrarily chosen such as ensemble size and T_{max}, while others have not been optimized for such as the covariance inflation constant α. It is therefore not possible to make fair comparisons between the overall performance of the implemented filters without further analysis and experimentation.

Appendix A Lorenz-96 Trajectory Configuration

A.1 Model description

The Lorenz-96 model is a dynamical system commonly used to benchmark various data assimilation methods, in part due to its chaotic nature when a forcing constant $F = 8$ is applied. Trajectories generated by this system are used throughout this project, and are summarized by eq. (8).

$$\frac{dx_i}{dt} = (x_{i+1} - x_{i-2})x_{i-1} - x_i + F(t), \tag{8}$$

A.2 Model parameters

Parameter values used to generate the various Lorenz-96 trajectories are outlined within table 1.

Parameter	**Value**	**Description**
N_{state}	40	Number of states
N_{obvs}	20	Number of observed states
H	Odd numbered states	Observation operator
N_{ens}	50	Ensemble size
Δt	0.1	Time step size
T_{max}	10	Max simulation time
r	0.025	Observation noise
R	$r^2 I_{N_{obs}}$	Observation error matrix
L	5	Decorrelation constant
α	1.02	Inflation constant

Table 1: Descriptions of model parameters used within project

A.3 Trajectory generation process

A.3.1 Reference states

The initial conditions $\boldsymbol{x}_0 \in \Re^{N_{state} \times 1}$ are first drawn from $\mathcal{N}(0, 16)$, with the forward Lorenz-96 model (using Matlab's ode45 solver) applied for $5\Delta t$ time units resulting in $\mathbf{x}_{-1}^{true}$. The reference state $\mathbf{x}_0^{true}$ was then generate by propagating for a further Δt time units.

A.3.2 Background ensemble

An ensemble of background states $\mathbf{X}_0^b$ were generated by adding N_{ens} perturbations drawn from $\mathcal{N}(\sigma^2 \mathbf{I}_{N_{state}})$ to $\mathbf{x}_{-1}^{true}$ with $\sigma = 0.02$, before propagating for Δt time units. This ensemble was used to calculate a covariance matrix B_0. The condition number of B_0 for the given seed equaled [XXX] and satisfied the definition of small in accordance with [REF]. As such a modified background covariance was not generated as the problem is well-conditioned. For the following data assimilation problems, initial background states are sampled from $\mathcal{N}(\mathbf{x}_0^b, B_0)$, where x_0^b is selected from the states X_0^b with the smallest $||X_0^b - x_0^{true}||$

A.3.3 Observation trajectory

A reference trajectory was generated by running the forward model over the interval $[0, T_{max}]$ using initial conditions $\mathbf{x}_0^{true}$, with data saved every Δt time units. The observation operator was then applied to the reference trajectory (selecting the odd states), before adding N_{ens} samples of additive random noise drawn from $\mathcal{N}(0, R)$ resulting in the ensemble of observation trajectories.