Domaine : MSXS

BE DATA ASSIMILATION

Auteurs :

M. Ali Houssenaly
M. Maël Dacher

27 février 2026

Question 1

Q1.1 – Influence of σ_o :
A larger observation error standard deviation σ_o means that observations are less reliable, so the analysis remains closer to the background state. Conversely, a smaller σ_o increases the confidence in observations, and the analysis moves closer to the observations. [6pt]

Q1.2 – B diagonal :
(i) At observed grid points : Increasing the background variance σ_b^2 increases the Kalman gain, giving more weight to the observations and pulling the analysis closer to them.
(ii) At unobserved grid points : Since B is diagonal, there are no spatial correlations. Therefore, observations do not influence unobserved points and the analysis remains equal to the background. [6pt]

Q1.3 – B SOAR :
(i) At observed grid points : Same behavior as with diagonal B ; a larger σ_b^2 increases the weight of observations in the analysis.
(ii) At unobserved grid points : Due to spatial correlations in B, observational information is propagated to neighboring grid points, so the analysis is updated even where no observation is available. [6pt]

Q1.4 – SOAR parameters :
(i) Increasing the correlation length L spreads the influence of observations over a larger spatial region, while decreasing L localizes their impact.
(ii) Increasing σ_b increases the overall background uncertainty, thus increasing the weight given to observations across the domain. Decreasing σ_b keeps the analysis closer to the background.

Question 2

The EnKF code is working : we find $RMSE = 0.565$ and reduced errors compared to the background.

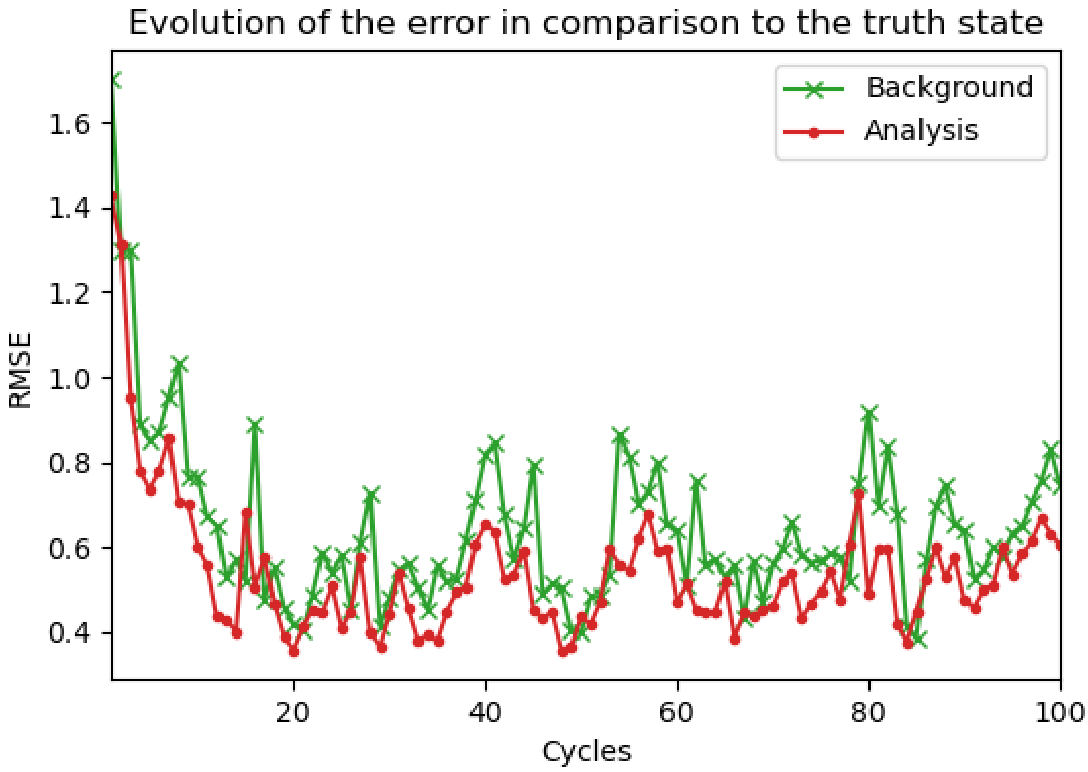

FIGURE 1 – Evolution of the RMSE of the state estimation

Question 3

The observation error standard deviation σ_R controls the trust given to observations in the EnKF.

For small values of σ_R, the observations are considered very accurate. As a result, the Kalman gain increases and the analysis closely follows the observations. This can lead to overfitting noisy observations and possibly to filter divergence if the ensemble spread becomes underestimated.

For large values of σ_R, observations are considered unreliable, so the Kalman gain decreases and the analysis remains close to the background forecast. In this case, the filter behaves almost like a free model run with weak observational correction, leading to larger estimation errors.

Question 4

The observation and background error variances are kept at their default values and we change the ensemble size N_e.

From the experiments, the smallest ensemble size for which the RMSE no longer exhibits a growing trend is found to be $N_{min} = 77$.

Below this threshold, sampling errors are too large and lead to unstable filter behavior. Above this value, the ensemble provides a sufficiently accurate approximation of the background covariance, ensuring stable assimilation.

Here are the error plots for $N_e = 76$ and $N_e = 77$.

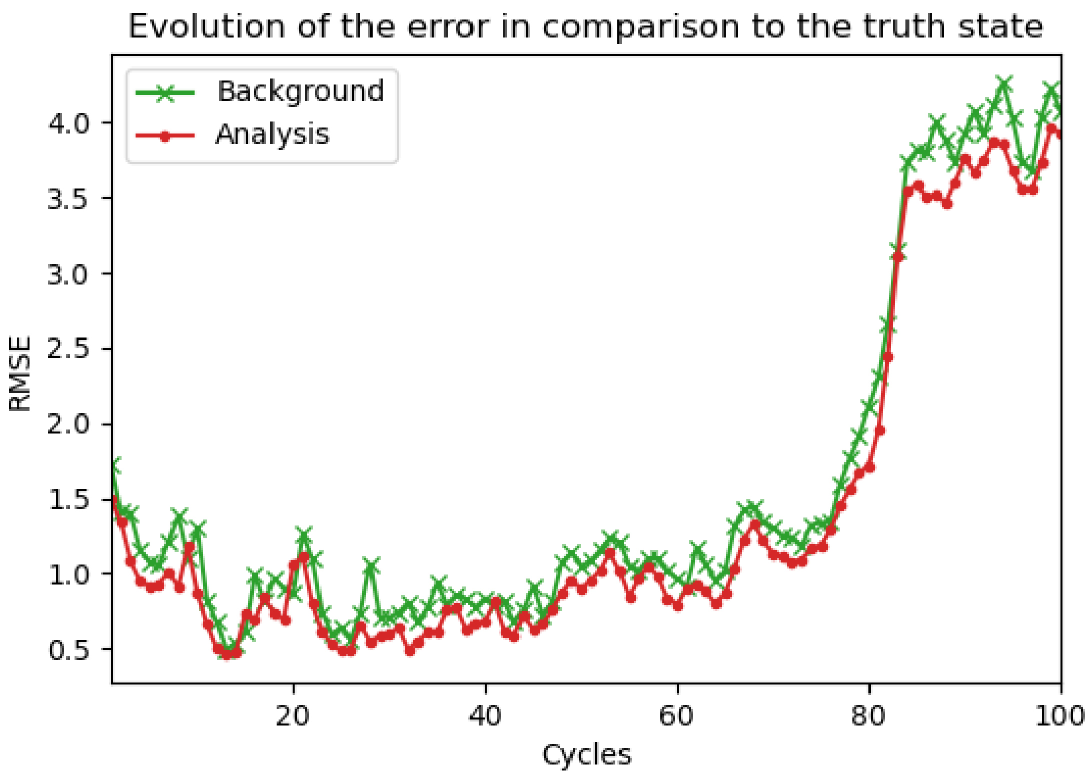

FIGURE 2 – Evolution of the RMSE of the state estimation for $N_e = 76$

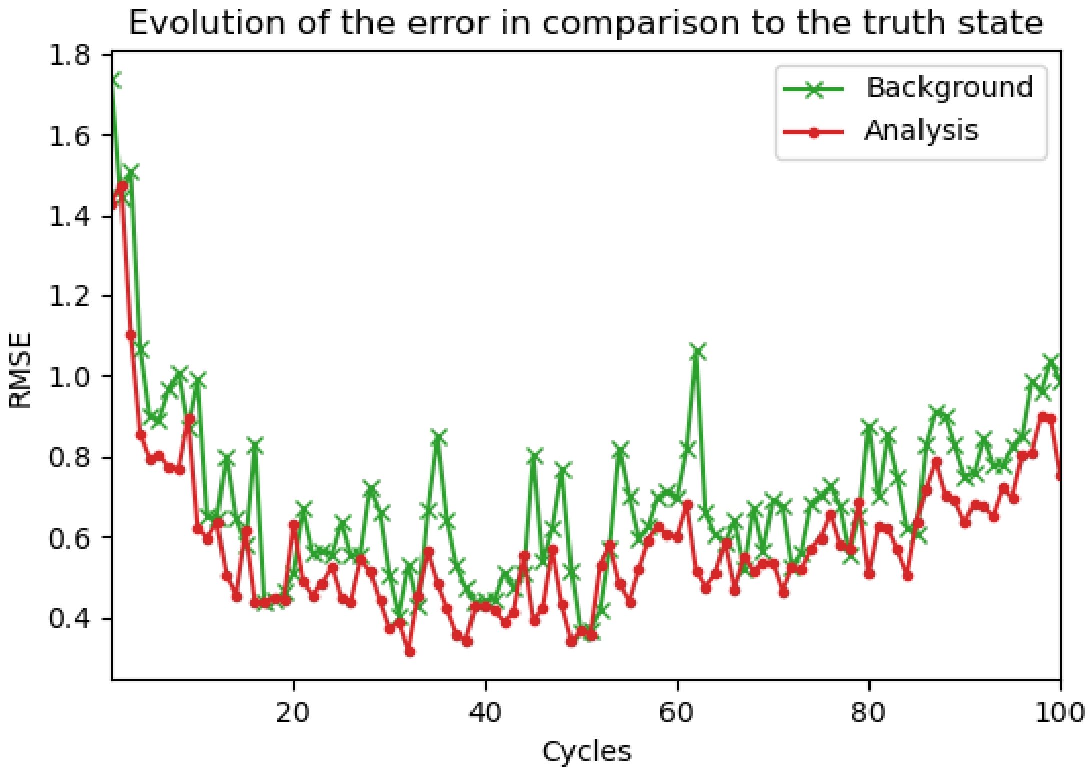

FIGURE 3 – Evolution of the RMSE of the state estimation for $N_e = 77$

Question 5

Yes, the EnKF is subject to ensemble collapse. Ensemble collapse occurs when the ensemble spread becomes too small, meaning that the ensemble members become overly similar and underestimate the true uncertainty of the state.

By examining the figure "Evolution of the local ensemble spread" for different ensemble sizes, we observe two distinct behaviors :

For $N = 10$, the ensemble spread rapidly decreases and becomes very small over time. This indicates a strong ensemble collapse : the filter becomes overconfident, underestimates the forecast error covariance, and may eventually diverge because observations are not given sufficient weight.

For $N = 100$, the ensemble spread remains significantly larger and more stable over time. The ensemble better represents the uncertainty of the state, preventing collapse and allowing the filter to maintain a balanced assimilation between model and observations.

Question 6

When enabling the inflation parameter in the code, the ensemble spread is maintained at a higher level, which improves the representation of uncertainty. As a result, the filter becomes more stable and less prone to divergence for small ensemble sizes.

Consequently, inflation can reduce the minimum viable ensemble size required for stable assimilation. We found the new $N_{min,inflation} = 28$.

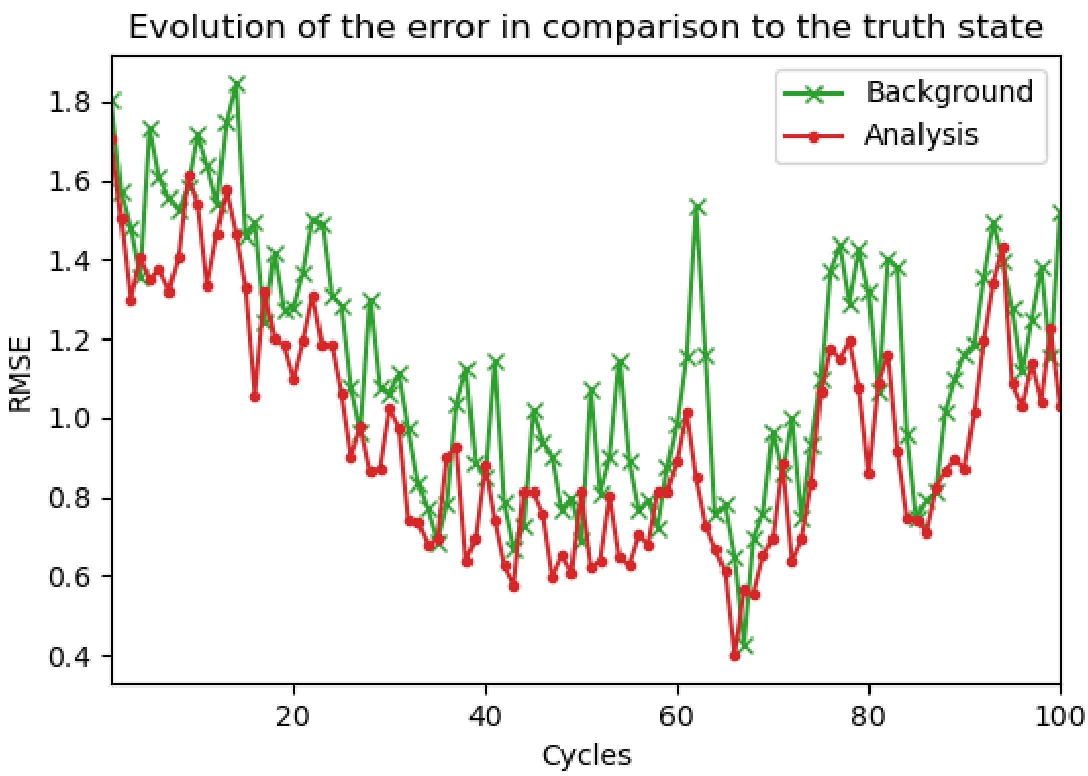

FIGURE 4 – Evolution of the RMSE of the state estimation for $N_e = 27$ with inflation

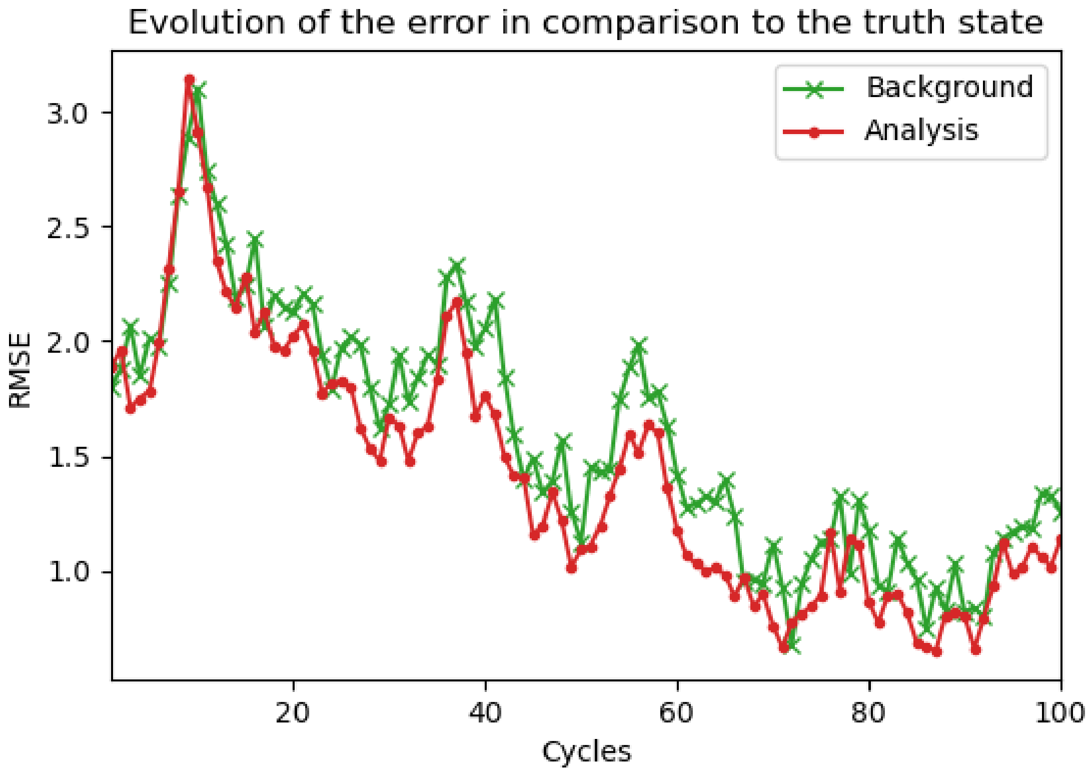

FIGURE 5 – Evolution of the RMSE of the state estimation for $N_e = 28$ with inflation