Data assimilation and inverse problems – Homework Set 4

4D-Var and ensemble 4D-Var for L'96

Build 4D-Var, E4DVar and EDA systems for the L'96 model. You can follow the steps below to build these systems.

1. Start with a 4D-Var system (no cycling). Make a "background" covariance matrix by using a long simulation (a few thousand time units) and compute a covariance matrix from this simulation. You should end up with a 40×40 matrix that has a banded structure (and "ears"). The background mean could be the mean of your simulation (a 40-dimensional vector). Use Gauss-Newton, and the linearized model, to solve the 4D-Var problem that you obtain by collecting an observation of every other variable (a 20-dimensional vector) after 0.2 time units. You can use an Euler discretization to make your life easy (I suggest a time step of $\Delta t = 0.01$ or smaller). Assume $R = I$. Compute RMSE of the background state and of the analysis state at time $t = 0$ and at time $t = 0.2$. Make plots of the Gaussian prior, $p(x_0)$, and of a Gaussian approximation of the posterior, $p(x_0|y) \propto p(x_0)p(y|x_0)$. Plot the data and the true state in the same figure. Make another plot of the same type, that shows (approximate) prior and posterior at time $t = 0.2$, i.e. plot approximations of $p(x_T)$ and $p(x_T|y)$.

2. Re-use the code from (1) and code from your previous homework to build an E4DVar system. Use an EnKF with ensemble size $N_e = 20$ to update the background covariance. Localization and inflation may require some tuning (use the HPC). Use a spun up ensemble, generated by running an EnKF to initialize the E4DVar system. Assume that there are 0.2 time units between observations, that you observe every other variable, and that $R = I$. What RMSE and spread do you get? (Average over at least 1000 cycles.) Compare with an EnKF.

3. Re-use the code from (1) to build an EDA system. Use $N_e = 20$, localization and inflation may require some tuning (use the HPC). Use a spun up ensemble, generated by running an EnKF to initialize the EDA system. Assume that there are 0.2 time units between observations, that you observe every other variable, and that $R = I$. What RMSE and spread do you get? (Average over at least 1000 cycles.) Compare with an EnKF and E4DVar.