Experiments with hybrid (ensemble-variational) methods in L63.

NCEO Intensive Course on Data Assimilation

University of Reading, 2022

1 Objective

In this exercise, you will experiment with hybrid DA methods in the Lorenz 1996 model with $N_x = 12$ variables. This size allows for relatively easy visualisation while requiring localisation of the ensemble covariances. You will explore 2 types of hybrid methods:

- First, those that use a hybrid background error covariance $\mathbf{B}_h \in \mathcal{R}^{N_x \times N_x}$ in a traditional variational minimisation framework. Here you will use 4DVar-LETKF.
- Second, those that avoid computing tangent linear and adjoint models and instead use 4-dimensional (cross-time) ensemble covariances to communicate the impact of observations to the initital time. Here, you will use SC-4DEnsemble Var (SC denotes strong constraint).

The only file you have to modify and run is *ControlHybrids.py*. This file has been divided in cells. You can highlight the instructions with the cursor and then press F9. Cells allow to run separate parts of the code without having to run the whole document.

2 Review of variational and Kalman-based methods

1. Run the nature trajectory for the experiment with $t_{max} = 14.$.
2. Generate the synthetic observations with $periodobs = 2$, $r = 1$, and $gridobs =' 1010'$.
3. Generate a climatological background error covariance to use in the variational experiments.
4. Perform DA experiments with 3DVar and 4DVar with 'anawin=2', i.e. 2 observations per window. You will get plot of the trajectories, as well as RMSE plots. What can you say about the performance of the two methods?
5. Now, we move into ensemble data assimilation. Generate plots for the localisation matrix in both the state space ($\mathbf{L}_{xx}$) and in the mixed state/observation space ($\mathbf{L}_{xy}$). For comparison, use the options:

loctype	lambda
0	0.1
0	2
0	10
1	0.1
1	2
1	10

Can you interpret these plots? After experimenting, let us settle for: $lam = 2$ and $loctype = 1$.

6. Perform DA assimilation using LETKF with $M = 20$ and plots the resulting trajectories. The code has adaptive inflation implemented, so you do not have to worry about this parameter at all. What can you say about the RMSE plots?

7. One of the purposes of hybrid DA is to combine covariance information from a static yet full-rank source (the climatological $\mathbf{B}_c$) used in the VAR methods, with the flow-dependent yet low-rank information coming from a sample of trajectories (the $\mathbf{P}^b(t)$ obtained by ensemble methods). Compare the climatological $\mathbf{B}_c$) with that obtained by the LETKF (computed from the background ensemble) at different times. The raw and localised versions are ploted for different times instants. In this case you can modify the variables *nsample*, which is the number of instants in which you want to display the $\mathbf{P}_b(t)$. How would this change if you increase or decrease the number of ensemble members?

3 Hybrid DA part 1

It is now time to start doing real hybrid DA. The first method we will try is 4DVar-LETKF. Recall that this method uses:

$$P_h^b = \beta_1 B_c + \beta_2 P^b \tag{1}$$

within a regular SC-4DVar cost function. Cell 6 contains the general settings to do this, it imports the necessary routines and it creates the localisation matrices needed. You can play with the following variables:
- *loch*: 0 if you do not want localisation in the ensemble part of the covariance, 1 if you do.

Cell 6.1 runs the assimilation, displays the trajectories obtained by this assimilation. You can modify the following variables:
- M: ensemble size.
- *obsperwin*: observations in the assimilation window.
- *beta*: the coefficients for the static (first number) and the dynamic (second number) part of the covariance matrix.

4 Exploring 4D covariances

Now we move into more complicated hybrid DA methods. We will use SC-4DEnVar. Remember that this method avoids using the tangent linear and adjoint models by computing 4-dimensional ensemble covariances. Let us start by comparing the error evolution coming from 2 sources: (a) evolving $\mathbf{B}_c$ using the tangent linear $\mathbf{M}^{0\to t}$ and adjoint $(\mathbf{M}^{0\to t})^{\mathbf{T}}$ models, and by evolving an ensemble run with different initial conditions (sampled from a normal distribution centered on the truth with covariance $\mathbf{B}_c$). This

is done in cell 7.1. You can vary the parameters:
- M: ensemble size.
- lags: number of time steps for which you want to compute the covariance.

This cell will plot three rows of covariances. Can you tell what is being plotted in each row?

5 Hybrid DA part 2

The final section, found in cell 7.2, runs SC-4DEnVar and computes the analysis RMSE of this method with respect to the truth. In this case you have to generate the localisation matrix (with the same options as before), and you can vary the next variables:
- lam: the localisation half width.
- M: the number of ensemble members.
- $locenvar$: whether you want localisation or not.

We use a fixed (in time) localisation. Remember this can be problematic when localising cross-time covariances in long assimilation windows. Can you think of a way to test this?