ATM 652 Fall 2012 Project # 2
Due 5 November 2020

The purpose of this assignment is to give you a hands-on appreciation for the various data assimilation techniques that were described in class, namely 3DVAR, 4DVAR and the ensemble Kalman filter, using code originally developed by Greg Hakim at the University of Washington. As a reminder, the Lorenz 63 model is governed by the following equations:

$$\frac{dx}{dt} = -\sigma x + \sigma y \quad (1)$$

$$\frac{dy}{dt} = -xz + rx - y \quad (2)$$

$$\frac{dz}{dt} = xy - bz \quad (3)$$

where r, b and σ are all free parameters (typically set to 28.0, $\frac{8}{3}$, and 10.0, respectively). Go to the course webpage and download the basic python code for this experiment. You are free to use another program to complete this assignment; however, you need to make sure you code the forward integration correctly. For each part, compute the errors with respect to the Euclidian norm.

(1) First, we will consider 3DVAR, which is contained in the code L63_3dvar.py. Initially, we are going to look at observations that are taken at all three state variables. Before you can run 3DVAR, you will need to produce background error statistics by running L63_stats.py. Start with an assimilation interval of 1.0 time units and observation error variance of 0.01. Run the assimilation system for 200 time units and produce a 2.0 time unit forecast at each initialization time. Compute the average cost function and difference between the intermediate state and truth as a function of iteration. In addition, compute the analysis and forecast errors over all forecast times. These values will serve as the baseline for subsequent experiments.

(2) In general, data assimilation can be computationally expensive compared to actually running the model. One way to get the data assimilation system to run faster is to modify the value of α, which controls how big of an increment one tries to make in the gradient of the cost function. Try increasing or decreasing the value of α by a factor of 2 and compare the amount of time it takes to run the assimilation system relative to the control value. What does this result tell you about the importance of using the proper α?

(3) Now we are going to modify various aspects of the data assimilation system to determine the sensitivity of the analysis and forecasting system to these components. Start with the control experiment and increase and decrease the observation error variance $\mathbf{R}$ by a factor of four. How does this impact the analysis and background forecast error? Return to the control settings, but now run three experiments, where the observation error variance is increased for one of the state variable observations. Based on these experiments, which observation should have the lowest errors to obtain the most accurate analyses and forecasts? Note that you need to decrease the value of α when you decrease the observation error

(4) So far, we have only dealt with the case where the observations are state variables; however, this is not always the case for the real atmosphere. Run an experiment where only one state variable is observed (i.e., you need to modify $\mathbf{H}$). How does this change the results relative to the default case. Which state variable is the most important to observe (i.e., which of the single variable assimilation experiments has the lowest error)?

(5) Finally, try assimilating observations that are averages of different combinations of the state variables, which is akin to how satellite radiances sample the atmosphere. How does this result compare to the default? What does this imply about observations that are not state variables?

(6) 4DVAR has the advantage of assimilating observations over a period of time, which you can adjust. Run 4DVAR, but where you vary the time window over which observations are assimilated (fvarwin) or the frequency over which observations are assimilated within the window (nobstimes). How do these changes impact the analysis and forecast error? Finally, change fvarwin to 0.7, nobstimes to 16, and toff to 0.4, which has the effect of increasing the assimilation window, but keeping the temporal observation density constant. How does the error compare to the control? How does the time it takes to run 4DVAR compare to 3DVAR?

(7) Reset the 4DVAR settings back to the control, but now only assimilate a single state variable (similar to question 4) or an average of two state variables (similar to question 5). How does this result compare to what was obtained from 3DVAR? What does this result tell you about which state variables are (or are not) important to observe?

(8) The remainder of the assignment concerns using the ensemble Kalman filter (EnKF). Run the EnKF system in the default configuration with 10 ensemble members. How does the analysis and background forecast error compare with the default 3DVAR configuration? How does the 3DVAR and EnKF analysis error compare when only assimilating a single state variable observation, or an

observation that is a linear combination of all three state variables?

(9) The primary advantage of the EnKF is that the covariances are flow-dependent. Plot the covariance between x and y as a function of time. How does this compare to the climatological covariance computed with L63_stats.m?

(10) In the EnKF, the covariances are computed from the ensemble members; therefore, the results are going to be sensitive to sampling errors that come from using a small number of members. Try another experiment with 20 ensemble members and compare it to the 10 member results. What happens if you run this with 5 members? Another method of maintaining ensemble spread is through inflation. Run the EnKF with an inflation of 1.0 and compare the resulting analysis error to the control configuration. What does this tell you about the role of ensemble size and inflation?