METEO 597 Assignment 3: EnKF for Lorenz-96 system (Due by April 2)

Using the Lorenz-96 OSSE framework from the last assignment, we will implement an ensemble Kalman filter (EnKF) and study its behavior.

Create a nature run (truth) with 10000 time steps (dt = 0.005). The observations are located on every other grid point, i = 1, 3, 5, …, 39, and observation errors are uncorrelated and error variance is $\sigma_o = 1$. Generate synthetic observations from the truth. Create a background state different from truth initial condition with RMSE = 1. Then, generate an initial ensemble by drawing random white noises with $\sigma_b = 1$ at each grid point for each member, and adding these noises to the background state (ensemble mean).

The CNTL experiment uses N = 80 members, assimilate observations (every other grid point) every m = 10 time steps (cycling period is 0.05 time units) for the 10000-time-steps test period. In the following questions, use CNTL configuration and change certain filter parameter(s) to run sensitivity experiments as instructed. Save the truth, observations, and the initial ensemble, and load these data when performing each experiment.

1. Implement a deterministic EnKF (choose your favorite variant, EnSRF or LETKF) to perform CNTL experiment. Compare the EnKF analysis ensemble mean with the analysis from 3DVar/OI using a $\mathbf{B}$ matrix with l = 0.5. Plot a saw tooth graph of their domain-averaged RMSEs. Does EnKF perform better than 3DVar? Why? (Hint: plot the correlation structure from time-averaged $\mathbf{P^b}$ and compare to $\mathbf{B}$).

2. Implement the Gaspari and Cohn (1999) fifth-order polynomial localization function for EnKF (see Lecture Notes page 61 for formula). Radius of influence (ROI) is defined as the cutoff localization distance (number of grid points). Use only N = 10 members, test a range of ROIs and find the best-performing one. (Hint: plot the time-averaged analysis RMSE as a function of ROI). How does your results change when N = 40 members are used? Explain the influence of ensemble size and localization on filter performance.

3. Consider the situation when observations are only available every m time steps (i.e. cycling period is $m\,dt$. In your cycling code, only perform EnKF update when observations are available). How does filter behavior change when observation time frequency increases? Test the cases m = 1, 10, and 100, and plot saw tooth graph of RMSE and ensemble spread. Interpret your results.

4. Implement the relaxation-to-prior-perturbation method (see Lecture Notes page 65) as covariance inflation. The truth model uses forcing parameter F = 8. Use an imperfect model with F = 7.8 in the forecast steps during cycling data assimilation. Find the best relaxation coefficient α by trial and error. (Hint: calculate time-averaged analysis RMSE and Consistency Ratio, and plot them as a function of α). When model error increases, say F = 6, how does the best α change?

Bonus question: Perform a simultaneous state and parameter estimation (SSPE). Initialize the ensemble members with wrong parameters $F = 6\pm2$. Add update equations for the parameter in your EnKF algorithm. Plot the evolution of F for each member over time, do they converge to the true value F = 8?