4DVAr with the Lorenz 1963 model

Dong Jian

Lorenz 1963 model

In this lab, the 4DVAR method is applied to the Lorenz 1963 equations, a simple dynamical model with chaotic behavior given by the following equations:

$$\frac{dx}{dt} = -\sigma(x - y),$$
$$\frac{dy}{dt} = \rho x - y - xz,$$
$$\frac{dz}{dt} = xy - \beta z,$$

where the model state (x, y, z) is time dependent, and three model parameters σ, ρ, β are 10, 8/3 and 28 respectively. These equations describe the evolution of the model state given the model parameters and the initial conditions.

4D-Var

Based on the Bayesian theorem, the 4D-Var method searches for a model solution minimizing the distance to both the observations (spread in time) and the (initial) background model sate, considering the statistics of their errors. The optimal analysis $\mathbf{x}_a$ minimizes the cost function $\mathrm{J}(\mathbf{x})$:

$$\begin{aligned}\mathrm{J}(\mathbf{x}) &= \mathrm{J}_b(\mathbf{x}) + \mathrm{J}_o(\mathbf{x}) \\ &= \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_{0\mathrm{b}})^{\mathrm{T}}\mathbf{B}_0{}^{-1}(\mathbf{x}_0 - \mathbf{x}_{0\mathrm{b}}) + \frac{1}{2}\sum_{\mathrm{i}=0}^{N}\left(\mathbf{y}_\mathrm{i} - \mathcal{H}_\mathrm{i}(\mathbf{x}_\mathrm{i})\right)^{\mathrm{T}}\mathbf{R}_\mathrm{i}^{-1}\left(\mathbf{y}_\mathrm{i} - \mathcal{H}_\mathrm{i}(\mathbf{x}_\mathrm{i})\right)\end{aligned}$$

with a strong constraint:

$$\mathrm{x}_{\mathrm{i}+1} = \mathcal{M}_\mathrm{i}(\mathrm{x}_\mathrm{i})$$

the model state x at time *t* is also a solution of the non-linear prognostic model, which implies the model is assumed perfect.

To find a minimum value, we investigate the sensitivity of the cost function to the initial state using iterative gradient descent method:

$$\nabla\mathrm{J}(\mathbf{x}_0) = -\lambda_0 + \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}_{0\mathrm{b}})$$
$$\lambda_0 = \mathbf{M}_0{}^{\mathrm{T}}\lambda_1 - \mathbf{H}_0{}^{\mathrm{T}}\mathbf{R}_0^{-1}(\mathcal{H}_o(\mathbf{x}_0) - \mathbf{y}_0)$$

For each iteration, first we run the non-linear model to calculate $\mathrm{J}(\mathbf{x})$ and then obtain its gradient using Lagrange multiplier method. And at last use descent algorithm to compute increment to initial state.

$$\mathbf{x}_0^{\mathrm{k}+1} = \mathbf{x}_0{}^{\mathrm{k}} + \alpha\nabla\mathrm{J}\left(\mathbf{x}_0{}^{\mathrm{k}}\right), \text{ where } \alpha \text{ is a step length}$$

Once the optimal initial conditions are reached, the model is run throughout the assimilation window to produce a forecast into the next window. This future forecast will provide the next assimilation windows background state.

1. **Is it better to have very few accurate observations or many observations which are less accurate?**

Two experiments are compared, experiment 1 has a few accurate observations without noise (freq=10; l_noise=0) and experiment 2 has many observations with some noises (freq=5; l_noise=1).

Figure 1 shows the solutions of model state of two cases. Even with few observations in experiment 1, the analysis curve is overlapped with truth curve. While in experiment 2, this agreement is valid only within simulation window, and the analysis can deviate far with truth by the end of forecast window.

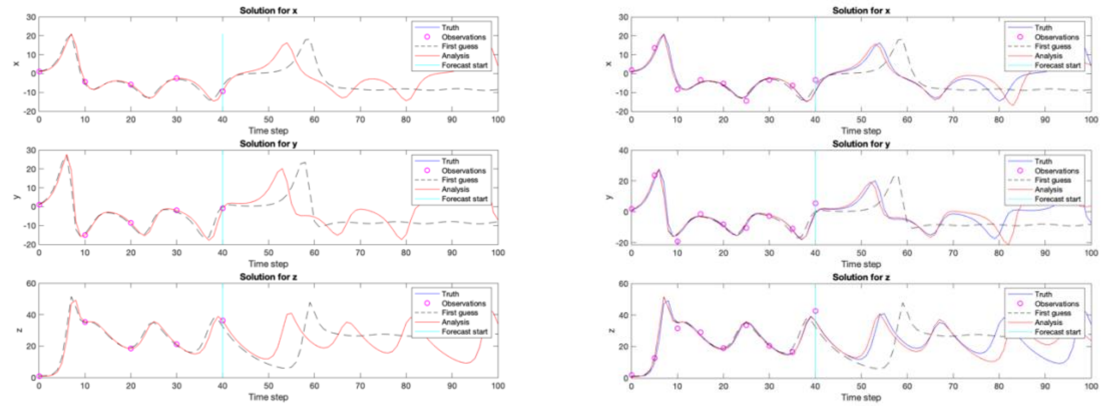

Figure 1 Solutions of experiment 1(left) and experiment 2(right), circle dots represent observations, blue curve is truth, red curve is analysis, vertical line indicates the time when the forecast starts.

Figure 2 plots the error (analysis – truth) of two cases. Within assimilation window, the errors of both cases remain almost zero. Near the 8th time step in experiment 2, one can notice an obvious stumble of the curve, this error may be relatively large compare error scale with that in experiment 1. The apparent differences start in the forecast period, the errors start to oscillate and increase in experiment 2 and become very large by the end of forecast. The errors in experiment 1 are still tiny and thus can be neglected. Both cases produce errors at around the end of forecast, while experiment 1 is relatively smaller and experiment 2 is much larger.

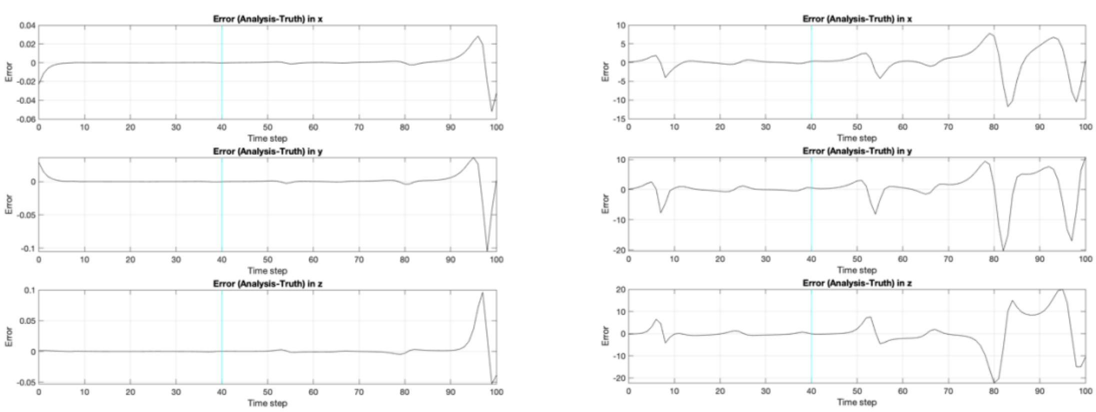

Figure 2 Errors of experiment 1(left) and experiment 2(right)

Figure 3 compares the convergence of cost function and gradient of two cases. Under consideration of experiment 1, the rate of cost function convergence is faster with only 13 iterations, while the latter uses 26 iterations. It means experiment 1 provides more accurate gradient of cost function and thus accelerate the convergence. In experiment 2, the cost function starts with a relative larger value and have abrupt changes during the minimization.

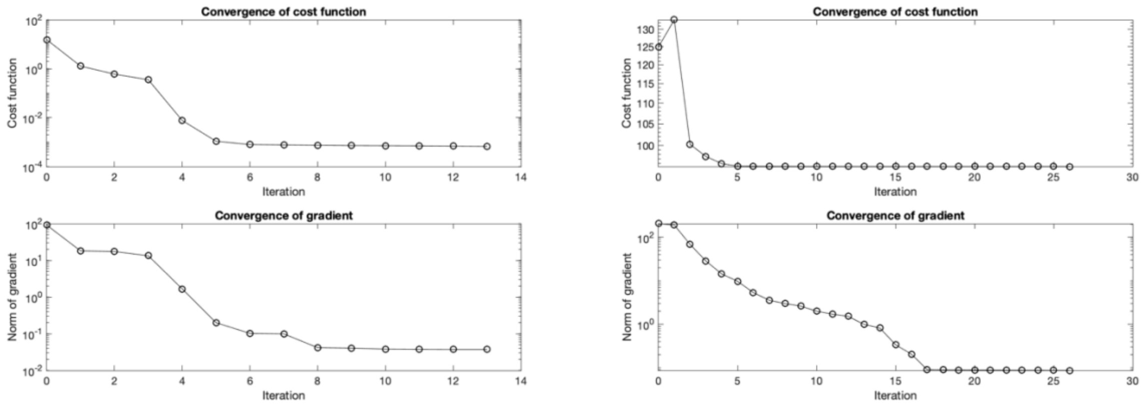

Figure 3 Convergences of experiment 1(left) and experiment 2(right)

Note due to the randomness of noise generator, experiment 2 can be repeated many times, even the plots of solution, error and convergence may be different each time, the conclusion is consistent with the result mentioned above. One can also set random generator as default, thus two experiments are comparable.

To conclude, the analysis of experiment 1 is more accurate and the cost function converges faster. Under above setting, one can say having very few accurate observations is better than many observations but less accurate.

2. **Is it better to have a long assimilation window with few observations or a short assimilation window with more observations? Does this depend on how much error there is on the observations?**

Experiment 3 has a long assimilation window (4 seconds) with few observations while experiment 4 has a short assimilation window (2 seconds) with more observations. Both cases do not have noise.

Figure 4 shows the solutions of model state of two cases. Experiment 3 runs assimilation for 80 time steps while experiment 2 runs 40 time steps. One can notice the analysis curve drifts off the truth curve before the end of assimilation window in experiment 3, while analysis overlaps with the truth in experiment 4. In the forecast period, analysis and truth are quite different in experiment 3, while they are still overlapping for experiment 4.

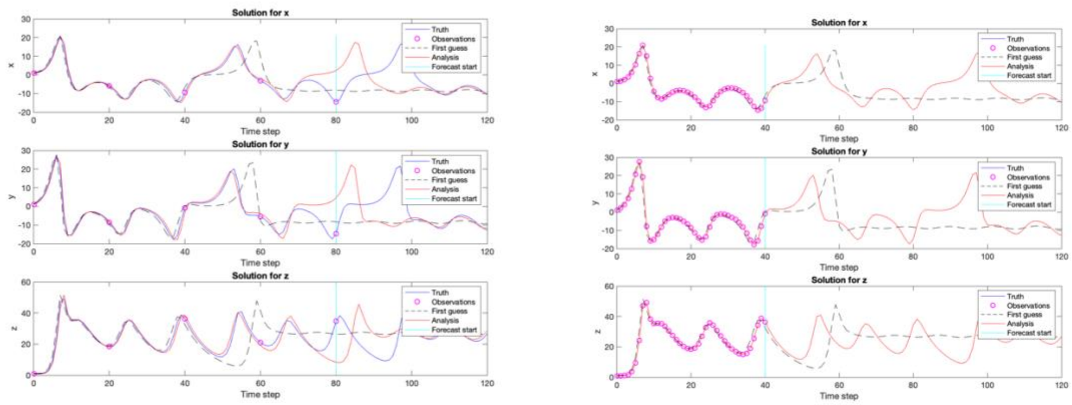

Figure 4 Solutions for experiment 3 (left) and experiment 4 (right)

Figure 5 plots the error (analysis – truth) of two cases. We find if assimilation window is too long, 4D-Var fails. The errors are unstable in the assimilation process and grow large in the forecast window. When the forecast starts, the first guess is far away from the observation, thus causing an inaccurate forecast subsequently. On the contrary, a short assimilation window produces an accurate analysis in both assimilation and forecast period.

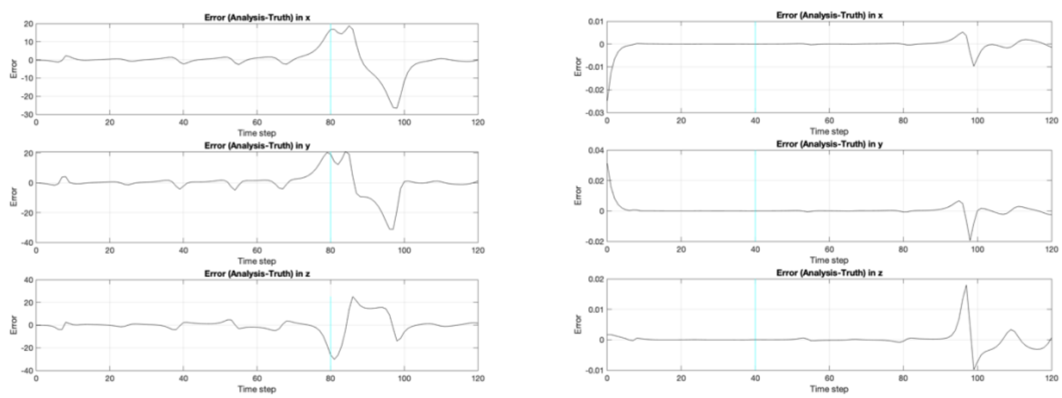

Figure 5 Errors of experiment 3 (left) and experiment 4 (right)

Figure 6 compares the convergences of cost function and gradient of two cases. A long assimilation window fails to converge in maximum number of iterations (exceeds 100), while a short assimilation window produces a much faster and stable convergence.

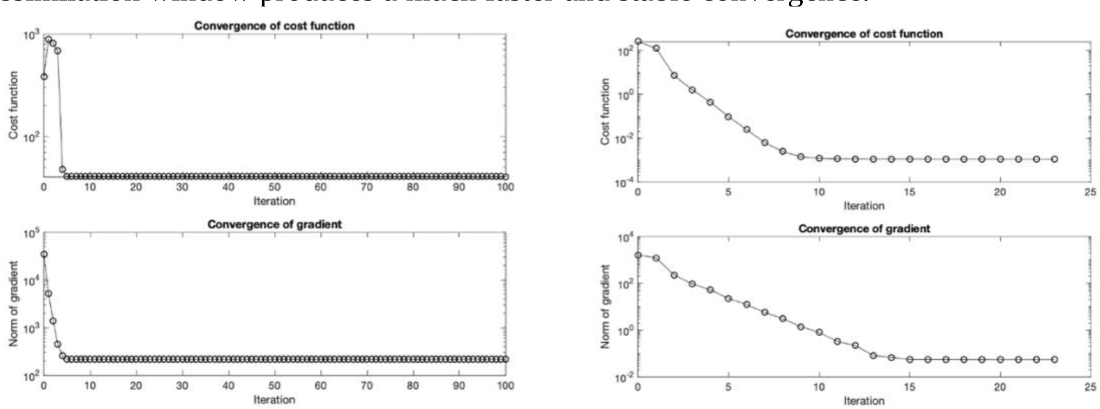

Figure 6 Convergences of experiment 3 (left) and experiment 4(right)

Under above setting without considering noise, a short assimilation window with more observation produces a better result. To exam the effects of noise, we activate random noise in both cases, repeat the two experiments and written as experiment 5 and experiment 6 respectively.

Figure 7 compares accuracy of the analysis when activating noise for experiment 5 and 6. The error range in experiment 5 is still comparable with experiment 3 without noise, which means a long assimilation window would produce large errors no matter whether noise exits. For experiment 6, the errors start to increase in the forecast period, smaller in the beginning and grow larger at the end, which indicates adding noise can produce error for later forecast and error increases with forecast time. If we only compare the beginning 2 seconds after assimilation, experiment 6 is still better than experiment 5. If we compare the whole assimilation and forecast period, both experiments fail.

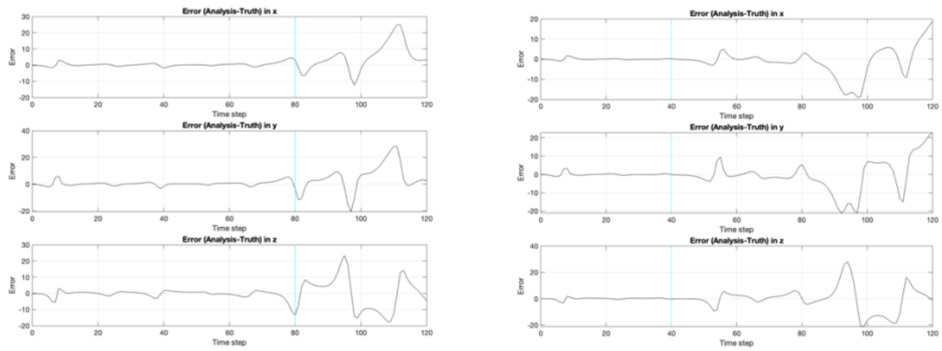

Figure 7 errors of experiment 5 with noise(left) and experiment 6 with noise(right)

Figure 8 compare the convergence of experiment 5 and 6. The convergence of cost function and gradient for experiment 5 is same with experiment 3 and fails to converge within maximum iteration, adding noise does not produce difference under the setting of experiment 3. Experiment 6 has a totally different convergence rate compared with experiment 4 without noise, the cost function starts with a higher value and needs more iteration to converge, the convergence rate of gradient is also slower. So, under the setting of experiment 4, adding noise do reduce efficient of minimization.

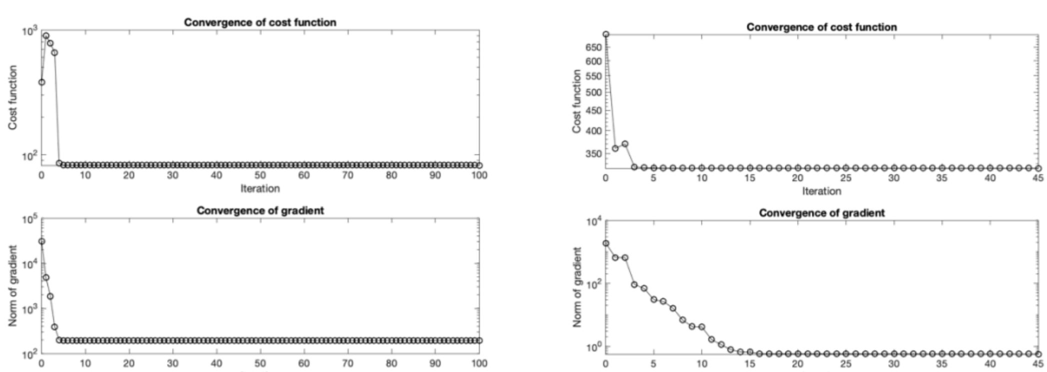

Figure 8 Convergences of experiment 5 (left) and experiment 6(right)

To conclude a short assimilation window with more observations is better than a long assimilation window. If a long assimilation window has failed the forecast, adding noise would not cause much influence. But adding noise do produce error and increase efficiency for a food forecast setting.

3. **How does the rate of convergence change if the background moved closer to or away from the truth?**

In experiment 7, the provided basic experiment is modified, the truth value at t=0 still is 1.0 for each state, consider no noise situation, the first guess is assigned with different values and count the iterations of convergence (see table 1).

Table 1 iterations for different first guesses without considering noise

First guess	0.6	0.8	1.001	1.005	1.01	1.05	1.2	1.3	1.4	1.5	1.6	1.7
iterations	32	17	fail	294	57	16	18	21	37	19	855	fail

* fail means iterations exceed maximum number of iterations which is over 5000

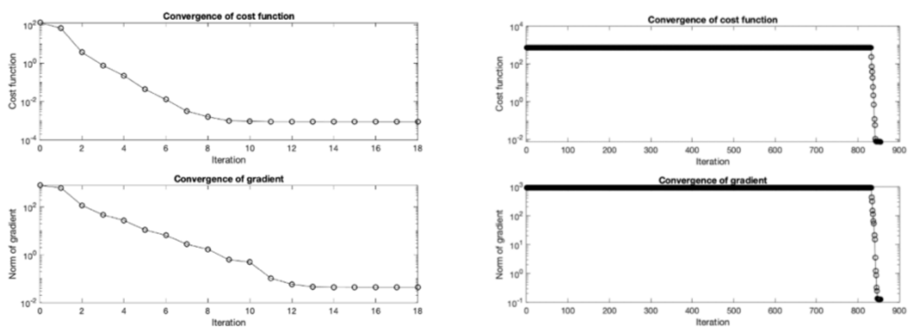

Figure 9 convergence of experiment 7 without considering noise, first guess = 1.2 (left), first guess = 1.6 (right)

We find if the first guess is close to the truth in a certain range, the rate of convergence is rational, however, if the first guess is too close (for the instance of 1.001), or the first guess is too far (for the instance of 1.7), it may fail to converge within the maximum number of iterations.

Next, the noise is switch on, set random generator as default to make them comparable. Both 1.2 and 1.6 produce a reasonable convergence. The convergence iteration would increase if the first guess moves closer or moves away, there is a reasonable value for the first guess which is neither too close nor too far.

Table 2 iterations for different first guesses considering noise

First guess	0.6	0.8	1.001	1.005	1.01	1.05	1.2	1.3	1.4	1.5	1.6	1.7
iterations	197	50	76	72	68	44	39	17	83	42	38	3257

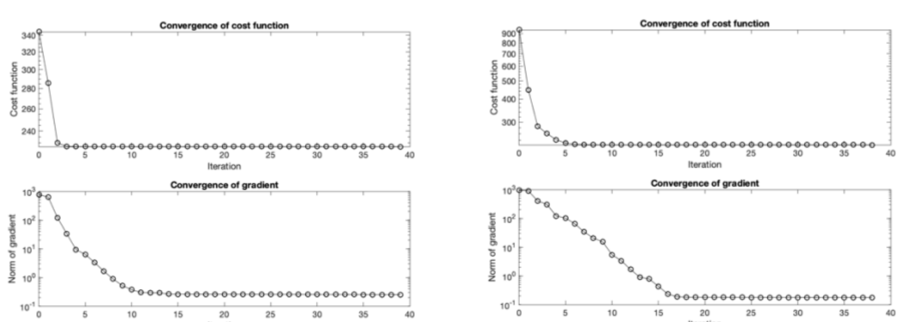

Figure 10 convergence of experiment 7 considering noise, first guess = 1.2 (left), first guess = 1.6 (right)

4. **Compare forecasts of different lengths for different conditions applied in the assimilation. This provides information about the Lorenz model predictability with respect to the selected assimilation setup.**

First, in the default setting, true value at t=0 is 1, the first guess is 1.2, switch off noise, simulate the truth over 12 seconds which we can use to provide "truth" value for later assimilations and for comparation.

Second, set assimilation window length T_1=2s, run the assimilation from T=0 to T=2s, and run the subsequent forecast starting from T=2s to T=12s.

Third, repeat the last step, starting assimilation from T=2s last to T=4s, using the truth from the first step and using analysis form the last step as the first guess, and run forecast for 8s.

Repeating the above step but start assimilation at the latter chunk and use information from previous analysis, thus creating several forecasts that overlap partly. Then we can compare the last chunk.

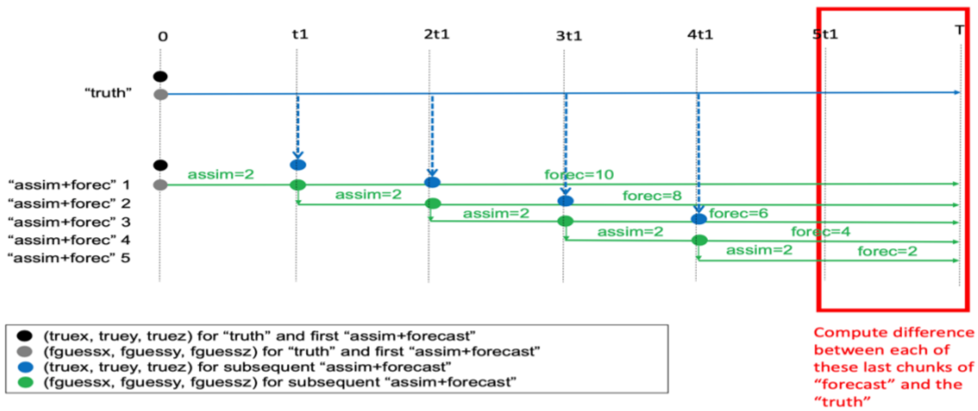

Figure 11 diagram for the experiment (Source:Nuno)

Figure 12 shows the errors of the last chunk for 5 different runs. "Forec1" run produces the least error (analysis minus the truth), except there is a slight shake near the end timestep. Because we use the analysis from previous experiment as the first guess, the error can be large. The deviation between the first guess and truth at the start of assimilation for the 5 runs are: 0.2, 1.09, 25.36, 17.45, 22.71 respectively. If the initial guess is not accurate, the forecast would not be accurate.

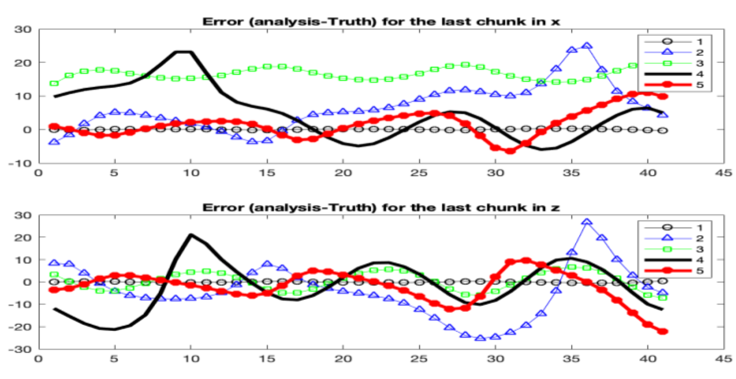

Figure 12 errors of the last chunk for 5 different runs

Next, we repeat the whole experiments, but the deviations between the first guess and truth are the same for the 5 experiments. In other words, the first guess is always 0.2 larger than the truth for all state variables and all experiment at initial.

Surprisingly, only forecast 2 and 3 produce large errors (Figure 13), probably because their prediction period is too long, and errors get projected for the last chunk. But forecast 1 has even longer prediction period and still have a good result, that is because the errors before the last chunk remain almost zero consistently, thus remain only very tiny error in the last chunk (Figure 14, left). Forecast 4 and forecast 5 have almost no errors because their forecast windows are quite short.

Therefore, one should stop forecast in time when there are some errors in the analysis period, otherwise, errors may get projected for later forecast.

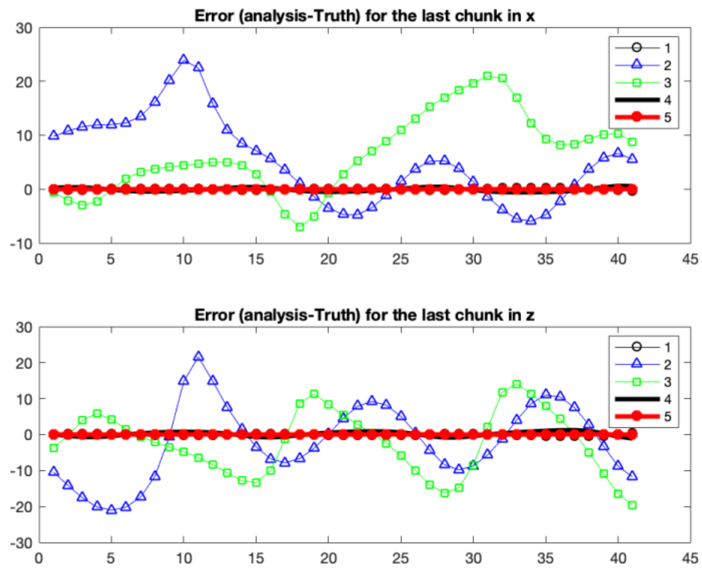

Figure 15 errors of the last chunk for 5 different runs

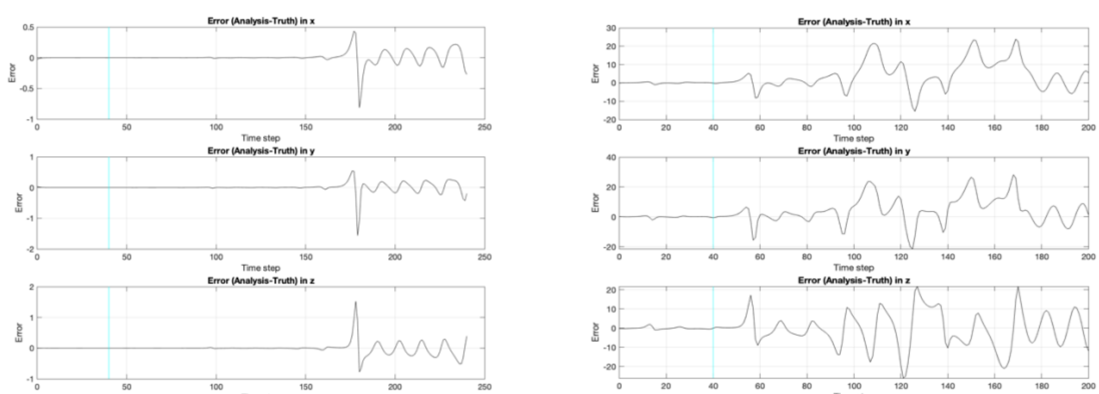

Figure 16 errors comparison for forecast 1(left) and forecast 2(right)