

# PROJECT2:ENSEMBLE DATA ASSIMILATION WITH ML MODEL

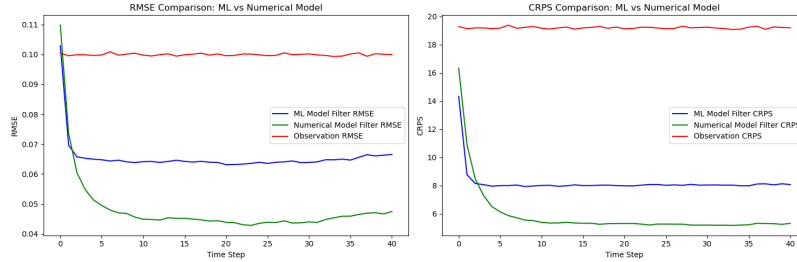
## CHU ZHUYIHENG 12455799

### 1. ENSEMBLE DATA ASSIMILATION WITH ML MODEL

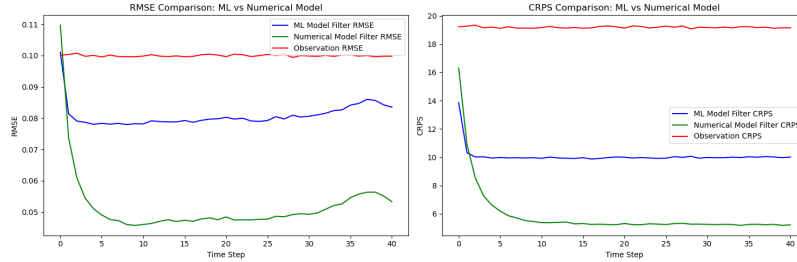
The script `test_addCRPS.py` has completed the task.

The following three figures show the results of the filtering performance comparison when the mean observation error is 0.1. The numerical ensemble size is set to 5, and the ML ensemble sizes are set to 20, 50, and 200, respectively.

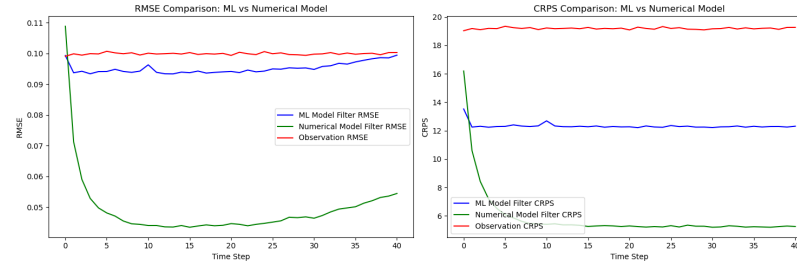
From the figures, it can be observed that the filtering effect corresponding to the numerical ensemble is significantly better. Interestingly, a larger ML ensemble size leads to increased error, contrary to my expectations.



(A) ML Ensemble Size = 20.



(B) ML Ensemble Size = 50.



(C) ML Ensemble Size = 200.

FIGURE 1. Comparison of filtering performance with numerical ensemble size = 5 and ML ensemble sizes = 20, 50, and 200.

However, the ensemble size cannot be too small. The following figure shows the results when the ensemble size is 2 and 10 respectively. It can be found that the error becomes larger than the observation error, which is obviously unacceptable.

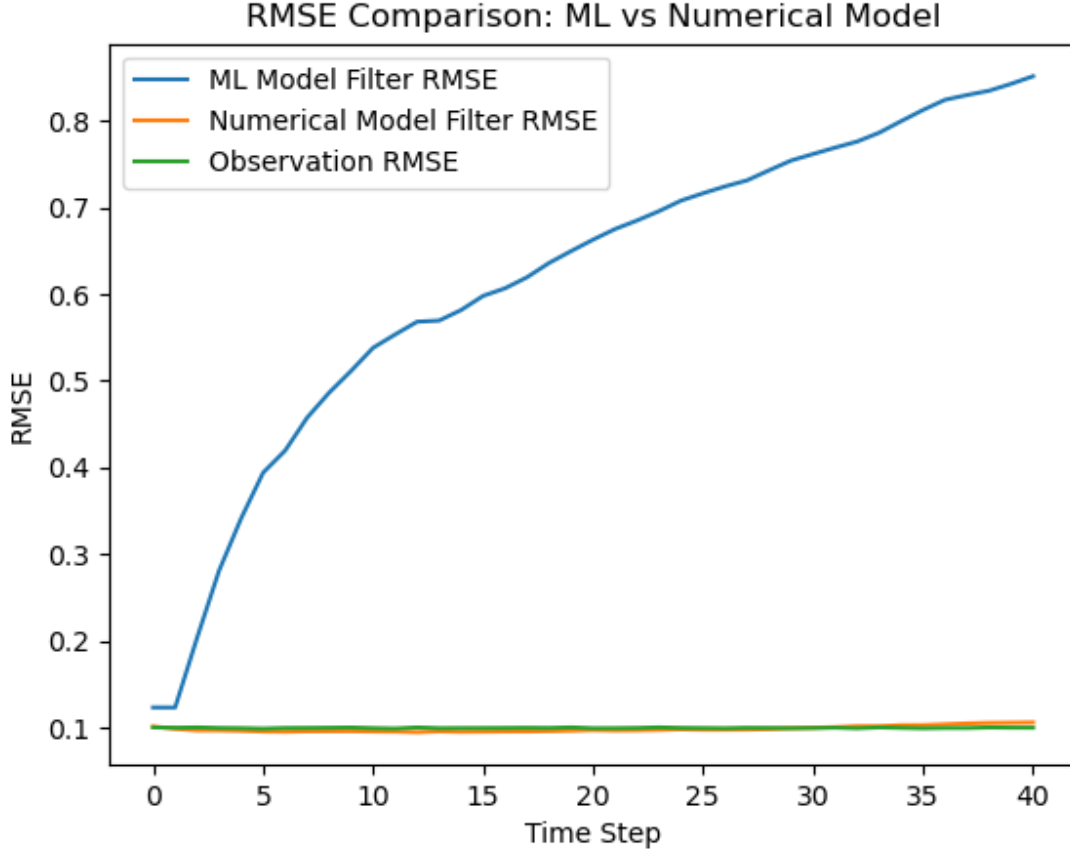


FIGURE 2. Numerical Ensemble Size = 2. ML Ensemble Size = 10. Get a big error

## 2. ATTEMPTS AT CUT-OFF RADIUS AND LOCALIZATION METHODS

The previous section basically covers the content required to achieve Project goals. Next, I will talk about the problems I encountered when doing the project and the solutions.

The problem is that the dimension of the problem state is too large. The covariance matrix that needs to be constructed is about 30k\*30k size. The computer memory is insufficient and the running time is very long.

A paper [1] I found pointed out that it is possible to set a cutoff radius and simply ignore the covariances beyond this radius. This method can make the covariance matrix sparse, solving the problems of insufficient memory and long running time.

I tried this method but it didn't work well.

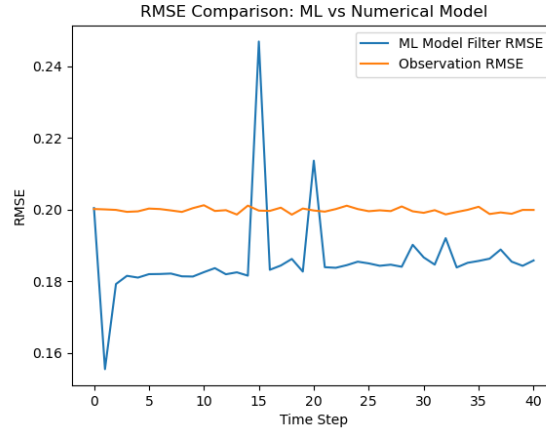


FIGURE 3. Cutoff radius = 2

Then I found another article [2] by the same author, which mentioned that the Schur product can be used instead of the cutoff radius method. This article uses a very complex compact support function for the Schur product. For the sake of convenience, I directly use the linear function of the distance.

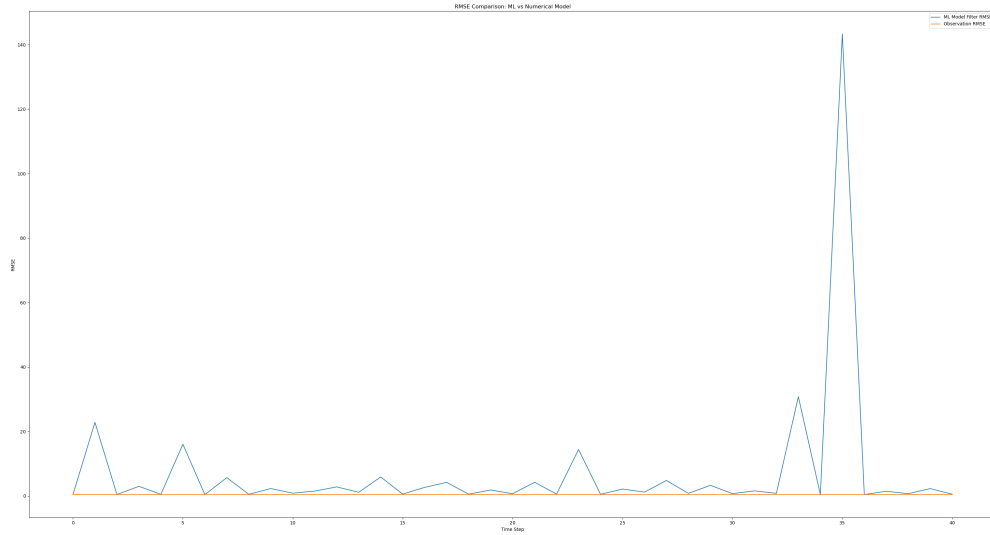
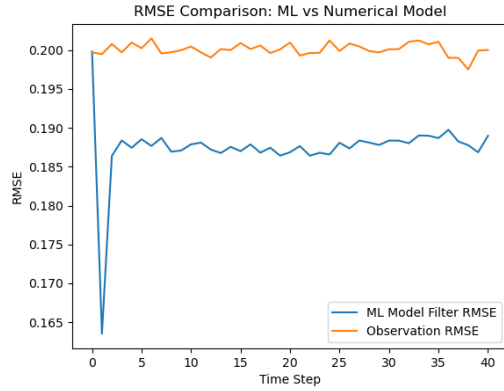


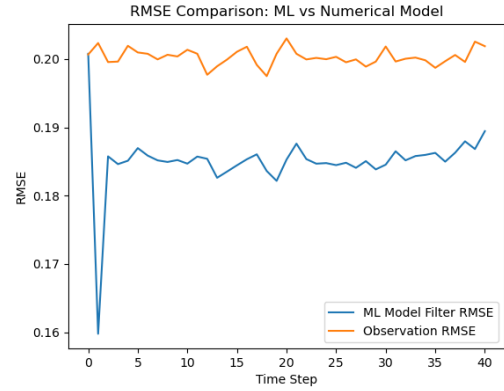
FIGURE 4. Cutoff radius = 2, Schur product method for localization

It can be found that the effect is very poor, but the article mentions that for the same set size, the error will first decrease and then increase as the cutoff radius increases.

Then I tried increasing the cutoff radius and got the following results:



(A) Cutoff radius = 3, Schur product method for localization



(B) Cutoff radius = 4, Schur product method for localization

It can be found that the error has been significantly improved. The results corresponding to the larger cutoff radius cannot be run due to insufficient computer performance.

The results in the previous section are obtained by running ENKF based on the optimal parameters here. (cutoff radius is 4, using Schur product method for localization)

#### REFERENCES

- [1] Houtekamer, P. L. and Mitchell, H. L. (1998). Data Assimilation Using an Ensemble Kalman Filter Technique. *Monthly Weather Review*, 126(3), 796–811. link.
- [2] Houtekamer, P. L. and Mitchell, H. L. (2001). A Sequential Ensemble Kalman Filter for Atmospheric Data Assimilation. *Monthly Weather Review*, 129(1), 123–137. link.