The paper tested different GSI assimilation settings with FV3 LAM runs, and proved that the system can predict the severe convective squall line case over Oklahoma on 4 May, 2020. The results are useful to most NCEP forecast data users. The paper has some unclear or incomplete reasoning but will likely be a significant contribution with revision and clarification.

General comments:

For most experiments in this paper, only the comparison results were shown, and the specific causes of the results were not analyzed.

For example, the differences between PSEUDO and 75EnBEC experiments were huge, but the authors have not given many diagnostics. Why more observations through GSI will cause overestimated convection?

Another example, the VLOC of 1 layer should capture finer vertical features of low atmosphere but the result showed that the positive impact is above 650 hPa and negative impacts are below 800hpa, why?

If the RRFS aims to replace the NCEP operational suite of regional and convective scale modeling systems in the next upgrade, it would be best to show the result from RAP as a baseline for all these tests.

Specific comments:

In Figure 2. RRFS cycling configuration diagram, the cold start is at 0 utc and the warm start seems to be from 1 utc to 6 utc. But in Figure 5 and relative context, the cold start is at 0 utc and 12 utc, what is the exact cold start interval?

L127 LAM appeared first time, should be limited area modeling (LAM) capability

In Figure 9, no "Matched pair counts used for RMSE and bias computation at each cycle" were found in the photos.