General comments

The article introduces an interesting hybrid data assimilation scheme called Local DA that utilizes model space, observation space and multi-scale localization. It is believed that great efforts have been made to develop and implement this algorithm. However, the article is not well written and it is difficult to follow the algorithm of Local DA, and its advantage over the other existing methods are not convincing. I have concerns as follows:

1. Overall, the language is not concise and not professional. For instance, in Line 70-71, "observation-associated grids (for scalar observations) and/or columns (for observations that measure an integrated quantity of the atmosphere, such as precipitable water vapor (PWV))", the phrase "observation-associated grids or columns" is strange, can it be simply "observation space"? in Line 99-100 "To avoid issues associated with the quality control of observations when evaluating the performance of Local DA, we adopt observing system simulation experiments (OSSEs)", "when evaluating the performance of Local DA" can be removed. In Line 86-87, "To simplify this study, we leave this issue to be addressed in future work", it is not common to bring up the future work in the introduction. In my opinion, the language of article needs to be greatly improved.

2. It seems that Local DA utilizes model space, observation space and multi-scale localization. Is the model space localization different from the multi-scale localization? As I understand, the multi-scale localization is done in spectrum space (which wavelengths?), does it work independently from the other localizations?

3. Can authors provide the dimensions for each components of Local DA (e.g., C_{oo}, S_o and etc.)? In addition, as in Hunt (2007), an analytical solution for Eq(1) can be derived. It is argued that the analytical solution is not feasible if the size of C_{oo} is large? Considering that 100 iterations are required to converge, the iterative method is computationally expensive. Can authors provide theoretical computational expense in case of the analytical solution?

4. It seems that the deterministic run is updated. How is it initialized and how is the Kalman gain of the deterministic run calculated? How are ensemble members updated (equations for this)?

5. I do not fully understand why the main discussion focuses on results of the single-cycle data assimilation. It is well-known that some cycles are usually required to spin up the system. It is argued that the Local DA suffers from imbalance problem. However, most of data assimilation algorithms have more or less the same problem. Is it possible to shift the focus of the discussion to the cycling data assimilation?

6. Evaluation of results: 1) I understand DTE as a metric for the error growth rate in forecasts, is it appropriate to use it to validate the analysis error? 2) In Figure 3, is the analysis or background spread/rmse shown here? Why are the RMSEs of two data assimilation experiments ignored? 3) It would be also interesting to see the verification only for the typhoon region (e.g., accounting for grid points $\geq$ 10 dBZ). 4) The shock is often mentioned in the article, can authors provide a metric for imbalance? Combining it with the RMSEs may show more insights of the results. 5) I suggest that a metric with scale skills (e.g., Fractions Skill Score) can be used especially for the reflectivity forecast verification.