This manuscript aims to show the ability of the DA experiments using modRSW to imitate some behaviors of an operational NWP so that one can utilize such inexpensive configurations to study DA in an operational-like environment. The topic is attractive, especially for researchers not in an operational center and who cannot run a real operational NWP due to computational resource limitations. This work is valuable, but the experiments are not well-designed concerning imitating the convective-scale cycling DA. My recommendation is a major revision before publication.

Major comments

1. The authors perform many experiments and give the experimental configurations relevant to the operational environment in terms of criteria such as grid spacing, ensemble size, localization, inflation, and RMSE. But what are the authors' suggestions on the improvement of localization and covariance inflation? If a new localization scheme or inflation scheme produces good results but cannot satisfy the criteria listed in this manuscript, should they be excluded from the operational NWP development? Showing the ability to imitate an operational-like environment is a good start, but further suggestions are also necessary. At least, the authors should tell readers what criteria should not be violated and what criteria can be adjusted when readers try to improve the DA performance in the simulated operational environment. Is a larger OID (>60%) not acceptable for imitating the convective-scale DA?

2. As far as I know, most operation centers use the variational DA algorithm or the variational-based hybrid algorithm. Only a few centers use the pure EnKF algorithm.
Giving a reason for choosing the pure EnKF algorithm to imitate the operational environment is necessary. I do not ask for conducting experiments with a variational DA algorithm, but a brief discussion on the selection should be helpful.

3. With respect to the convective-scale DA, the model and DA configurations are not so representative.

(1) The precipitation procedure in modRSW is more like a cumulus parameterization scheme that estimates the precipitation according to the large-scale thermodynamic environment (see the high correlation between $r$ and $h$ in Figure 6), while a feature of convective-scale NWP is using a microphysics scheme that explicitly simulates the physical procedures in the cloud. The difference in complexity is a gap between a convective-scale NWP and a synoptic-scale NWP. Heavy rain may occur with no large-scale forcing. So I think referring the DA experiments using modRSW as “convective-scale” DA is not proper.

(2) The observation density is too sparse for the convective-scale DA, especially in the case of assimilating radar and satellite data. The resolution of radar data is often 1 or 2 km. The lack of high-resolution observations is a flaw in the experiments aiming to imitate the convective-scale DA. Using multiscale observations is also a feature of convective-scale DA and is not considered or discussed in the manuscript.

(3) The precipitation $r$ in the manuscript is more like a simultaneous quantity, e.g., the precipitation rate in a time step. The accumulated precipitation, 3-h or 6-h, is used in real data assimilation. In this respect, the DA configuration in the manuscript does not imitate the real scenarios. This situation should be stated.

(4) In convective-scale DA, we must face that many model variables are not directly observed. This issue implies that some model variables must be updated through cross-variable covariance. It is better to show results without $r$ and $h$ observations.

(5) The operators of remote observations are often nonlinear, such as those of radar reflectivity and satellite observations. The inaccurate operator is also an issue in the convective-scale DA, but this issue is not discussed.

In general, the DA configurations in the manuscript are more suitable for a synoptic-scale DA study; many issues that the convective-scale DA has to face are not discussed. Since it is an idealized study, doing DA experiments with multiscale observations and without $r$ and $h$ observations should not be difficult. Observing the precipitation area with high-resolution $u$ observations should result in a much smaller RMSE, similar to radar data assimilation.
4. The text is not well organized. The description of the model, DA algorithm, and experimental design should be more concise. Too many details distract readers from focusing on the main idea of this paper. For example, I don’t think section 4.3.5 is necessary. There is no need to list the formulas of RMSE and ensemble spread. In addition, section 6 is more like a summary, not a conclusion.

Minor comments:

L320-323: With respect to imitating LFC with $H_c$, I have some reservations.

L407: Is $K$ in Equation (19) identical to $K_e$ in Eq. (6b)? If so, use $K_e$ please. If not, what is the difference?

L421: If a new DA method or a new configuration has a much larger influence (OID) in convective-scale DA, what is the authors’ suggestion?

L606: “the limitations of the EnKF” What are the limitations?

L611: Should “Fig. 11” be “Figure 11” at the beginning of a sentence?

L616: It seems that “).” is missed after (Fig. 10.

L623: “This shows the impact of a well-calibrated $P'_e$ matrix” How do authors define a well-calibrated $P$?