Response to Referee:

The authors appreciate the referee's insightful comments and constructive suggestions, which helped us significantly improve the quality of this manuscript. Below are itemized, point-by-point responses to the referee' comments.

Review of Impact of Multiple Radar Wind Profilers Data Assimilation on Convective Scale Short-Term Rainfall Forecasts: OSSE Studies over the Beijing-Tianjin-Hebei region

General Comments:

Thank you for addressing all of my comments from the first round of reviews! I think the manuscript is getting really close to being publication-ready. I only have a handful of minor comments, and most of these just involve rewording some sections of the manuscript. Thanks again for all the work you have put into this manuscript.

Specific Comments:

1. Figure 2: It would be helpful to state that rain gauges that did not measure any precipitation are not included. This explains why there are different numbers of dots in the two panels.

   Response: Thanks for pointing out this. The claim has been added in the figure caption as "The rain gauges that did not measure any precipitation are not included here" (line 221).

2. Lines 320–321: I appreciate that the authors changed "convection-resolving" to "convection-allowing", but I don't think the Potvin and Flora (2015) reference is needed here. I just provided that as an example as to why having grid spacing > 1 km would be considered "convection-allowing". The term "convection-allowing" is widespread enough that I don't think a reference is necessary.

   Response: Corrected as suggested.

3. Lines 465–467: Should this statement only be for ETS? An examination of Fig. 12 shows several examples where SR is higher for NoDA compared to the DA experiments (e.g., when using a 40-dBZ threshold, the NoDA analysis has the highest SR). There are also examples where POD is better for NoDA compared to CTL (20-dBZ threshold, 6-hr forecast) and where CSI is better for NoDA compared to CTL (40-dBZ threshold, 3-hr forecast).

   Response: Yes. The claims have been reworded as "Overall, all DA experiments consistently outperform NoDA at all thresholds, exhibiting higher ETS values, except for the CTL (FH and FH_RD) experiment during the 0-4 h (4-5 h) forecast

period at the threshold of 40 dBZ. For most thresholds and forecast lead times, the assimilation experiments generate higher POD, SR and CSI scores compared to the NoDA experiments (with the exception of a few instances, primarily at the 40-dBZ threshold)" (lines 466-471).

4. Lines 470–471: For this statement about bias, is FH_RD being compared to CTL? Based on Fig. 12, it seems like the bias is the same or worse for FH_RD compared to CTL, not better.

   Response: In this statement, FH_RD is compared to all other assimilation experiments, not just CTL. Yes, the BIAS values of FH_RD are generally comparable to those of the other assimilation experiments and, in some cases, slightly worse. The claims have been reworded as "However, the BIAS values of the FH_RD experiment is comparable to that of other DA experiments and are sometimes slightly worse (Fig. 12a-d)" (lines 474-476).

5. Lines 497–501: Should the beginning of this sentence say "...inferior performance of FH_RD compared to FH"? The rest of the sentence focuses on why FH_RD performs worse than FH and does not mention RD at all (unless I am missing something here).

   Response: Thanks for pointing out this. This statement is intended to explain the previous sentence "The RD experiment outperforms all the other experiments in the 1-, 3-, and 4-h forecasts at the threshold of 10 mm". The expression has been reworded as: "One possible reason for the superior performance of RD compared to FH_RD and FH at higher thresholds is that, the heavy rainfall coverage forecasted by the RD experiment is the closest to the truth, while FH_RD exhibits a slight southward displacement error, and FH shows a northward displacement error (Fig. 11 vs. Fig. 10a-c)" (lines 502-508).

6. Lines 542–543: It is tough for me to evaluate this claim using Fig. 15, but based on Fig. 16 and 17, FH_RD does not seem to clearly have lower bias than NoDA.

   Response: This claim has been removed.

7. Figure 15: I appreciate that the authors added the NoDA experiments to this figure. Owing to all the points, it is now tricky to compare the NoDA and FH_RD experiments, especially because some of the colors are really similar. Would it be possible to increase the clarity of this figure? One option could be to split the figure in two, with one figure for each of the new cases (one for the 28 June 2023 case and one for the 12 July 2023 case). This option would not require a third figure for the 21 July 2023 case because those results have already been presented.

   Response: Good point! Per your kind suggestion, this figure has been split into Fig. 15 and 16 for the case 28 June and 12 July 2023, and we also reworded the corresponding text to reflect the change.

8. Lines 600–601: Can the beginning of this sentence be a bit more specific? E.g., "Some possible reasons why FH outperforms RD for shorter forecast lengths but RD outperforms FH for longer forecast lengths are…"

   Response: Corrected as suggested.

9. Line 617: The word "smaller" might be better than "minimal". "Minimal" makes it sound like RWP DA has almost no impact in the CTL and FH_RD_H3 experiments. RWP DA still has a considerable impact in CTL and FH_RD_H3 compared to NoDA, it is just that the impact is smaller than the other experiments.

   Response: Corrected as suggested.

10. Line 703: The line between "identical twin" and "fraternal twin" can be a bit blurry and can sometimes be ambiguous. I would recommend being more specific and saying something along the lines of "the same modeling system is used for the truth run and forecast system".

    Response: Corrected as suggested.

Technical Corrections:

1. Line 136: "NWP" was already defined on line 49, so there is no need to define it again.

   Response: Corrected as suggested.

2. Line 203: Replace "the constraint" with "that constraint".

   Response: Corrected as suggested.

3. Line 237–238: The way this sentence is currently worded makes it sound like the synthetic wind profile only consists of a single height. Maybe it can be changed to be something like "The heights where the winds are measured (H) at each simulated RWP site are as follows:"

   Response: Corrected as suggested.

4. Line 263: "setupa" should be "setup, a".

   Response: Corrected as suggested.

5. Line 278: Add "a" before "radar wind profiler".

Response: Corrected as suggested.

6. Line 408: “suppress” should be “suppresses”.

   Response: Corrected as suggested.

7. Line 440: Replace “It is due” to “This is due”.

   Response: Corrected as suggested.

8. Line 470: Can a transition be added before “the improvement in BIAS…”?

   Response: Corrected as suggested.

9. Line 533: “observed from” might not be the best word choice here. Maybe just saying “in” would be better?

   Response: Corrected as suggested.

10. Line 541: “SAR” should be “SR”.

    Response: Corrected as suggested.