Reviewer 1

Summary/overall

I have carefully read and reviewed Park et al.'s preprint submission to EGUsphere titled "Development and Testing of Ensemble-Variational Data Assimilation Capabilities for Radar Data within JEDI coupled with FV3-LAM Model." Having read much of this group's prior literature (and, in fact, I think I reviewed Liu et al. 2022) I am quite familiar with the research and development activities covered herein. Overall the paper is fairly well-written and the science is sound. The paper is acceptable overall, but it has some flaws. One semimajor concern is that this particular manuscript doesn't advance the overall science much, as it seems to mostly be an extension of Liu et al. (2022) by introducing the same reflectivity operator developed previously for GSI into JEDI. Also, other than producing the forecasts using a different dynamical core, FV3-LAM, there is little novelty. Additionally, only a single case study is used to illustrate the impact of the new reflectivity observation operator, which to the authors' credit, is acknowledged, albeit late in the manuscript, and well after verification plots are presented. Regardless, it remains true that the assessment of the new operator is reliant on a single case; it would be preferable for many cases to have been run and evaluated so that the statistical robustness of any improvements coming from this work can be addressed. Finally I will note that the authors make some comments in the discussion section that are not strongly supported by the evidence provided. These statements need to be softened to avoid showing a confirmation bias. In spite of these weaknesses, however, I will recommend publication of this paper pending satisfactory improvements are made corresponding to the detailed specific comments below

Response to Summary/Overall Comments :

We thank the reviewer for the careful and constructive review, the positive assessment of the overall quality and soundness of the manuscript, and the recommendation for publication pending revisions.

Regarding the reviewer's concern about novelty and the relationship to Liu et al. (2022), we agree that the present work builds upon the reflectivity observation operator introduced in that study. However, this paper does not simply port the existing operator into JEDI. Instead, we introduce targeted modifications to the observation operator and conduct a systematic comparison between the original and modified formulations within the next-generation JEDI ensemble-variational framework coupled to the FV3-LAM model. The resulting impacts on hydrometeor analysis increments, analysis behavior, and subsequent forecasts are examined in detail. These aspects were not explored in Liu et al. (2022). We have revised the manuscript to more clearly articulate this contribution and to better distinguish the scope and novelty of the present study.

We also acknowledge the reviewer's concern regarding the reliance on a single case study. We agree that broader multi-case evaluation would be desirable for assessing statistical robustness. The primary objective of this paper, however, is to demonstrate the correctness and practical feasibility of the algorithm implementation within the JEDI framework, rather than to provide a comprehensive statistical evaluation across many events. From this development and technical perspective, a well-documented high-impact case and the current number of radar data assimilation cycles are sufficient to verify the behavior and consistency of the new capability. We have revised the manuscript to clarify this scope earlier in the paper and to explicitly acknowledge the limitations associated with using a single case, while highlighting the need for future multi-case testing.

We note that more comprehensive testing was constrained by recent reductions in NOAA funding. Nevertheless, the capabilities developed and evaluated in this study have been incorporated into the official JEDI repository. As such, this paper serves as an important technical documentation of these capabilities, enabling further testing, evaluation, and extension by the broader research and operational community.

Finally, in response to the reviewer's comments regarding confirmation bias, we have carefully reviewed the discussion and results sections and softened or revised statements where appropriate to ensure that conclusions are strictly supported by the presented evidence.

Specific comments (Listed primarily by level of concern rather than by appearance in the manuscript)

(1) The following list contains statements of questionable veracity or in which there is a sense of confirmation bias. These statements need to be either softened or altered to state only what is shown rather than what the authors hope is the case (some of which we cannot know for sure given only one case was tested).

• (a) Lines 280-282: I sense some confirmation bias. Some of this is not at all obvious to me (only the broadened moderate reflectivity behind the convective line is clear).

Response: We have revised the text to describe only the reflectivity differences directly evident in Fig. 5 and removed subjective or evaluative language to avoid overinterpretation based on a single case.

The relevant sentences now read as follows:

“The impacts of the modified reflectivity operator on the background fields are illustrated in Fig. 5. Compared to JEDI-Org (Fig. 5c), JEDI-New (Fig. 5d) exhibits locally stronger reflectivity along portions of the leading convective line and a broader area of moderate reflectivity in the post-line region. In contrast, JEDI-Org shows weaker reflectivity coverage in the trailing stratiform or weak-echo region in this case, leading to differences in the representation of the convective system.**”**

- (b) Line 309: “comparable to MRMS” is not supported by Fig. 6. The dark greens top out around 10 km in JEDI-new and 14 km in MRMS. And they're only marginally higher than in JEDI-org or GSI-new. Soften and/or include this additional detail. Be specific about what threshold you’re using to make such statements.

Response: We agree that the previous wording was too strong for the secondary tower. We removed the phrase “comparable to MRMS” and revised the text to describe only the features directly evident in Fig. 6, without making quantitative equivalence claims.
The relevant sentences now read as follows:
“The analysis fields indicate that radar DA recovers the previously missing western convective feature. In JEDI-New (Fig. 6h), the secondary tower is more evident than in the background fields and the NODA experiment, although differences remain relative to MRMS.”

- (c) Line 312: It looks to me like the vertical extent (i.e., updraft top) in all three DA experiments (Figs. 6f,g,h) is a bit short compared to that in the MRMS cross sections. If you're talking about the height of the top of the stratiform area, on the other hand, I don't think that is as different or important.

Response: We agree with the reviewer. The previous statement regarding the vertical extent has been removed, and the revised text now focuses on features that are directly evident in the reflectivity cross sections without making comparisons of updraft or echo-top height.
The relevant sentences now read as follows:
“GSI (Fig. 6f) and JEDI-Org (Fig. 6g) underestimate its strength. In the primary tower region near 95°W, JEDI-New also effectively reduces the midlevel overestimation seen in the background. By contrast, JEDI-Org retains an overly strong midlevel core.”

- (d) Lines 328-329: disagree/confirmation bias. The increment plot (Fig. 7c) looks pretty similar to that from GSI-new; there is certainly a coherent region of negative graupel increments associated with the eastern updraft tower.

Response: We agree that Fig. 7c shows a coherent region of negative graupel increments associated with the eastern convective tower. We revised the text to acknowledge this feature and to state more precisely that the negative increments in JEDI-Org are weaker and more localized, while JEDI-New and GSI-New exhibit stronger and more spatially coherent negative increments.

The relevant sentences now read as follows:

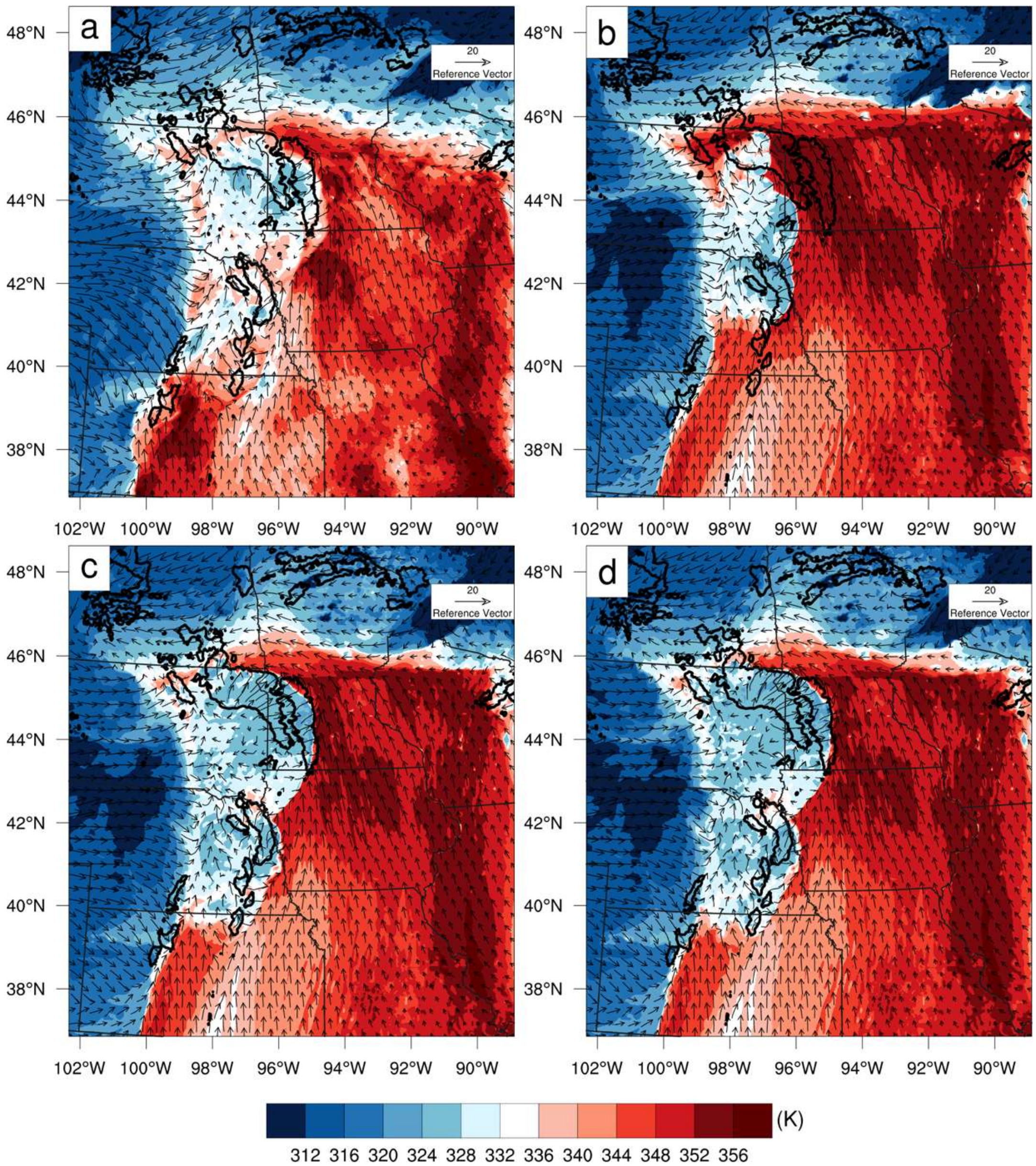
"For graupel (Figs. 7c, f, i), all three DA experiments show negative increments collocated with the eastern convective tower. The negative graupel increments are weaker and more localized in JEDI-Org (Fig. 7c), whereas JEDI-New (Fig. 7f) and GSI-New (Fig. 7i) exhibit stronger and more spatially coherent negative increments centered in the midlevels of the primary tower."

• (e) Lines 390-392, 393-394, and Figure 9: not at all clear from inspection. Make a difference figure instead. In general, at the horizontal extent of the plots in Fig. 9 it is difficult to discern these statements. I suggest either making Fig. 9 larger or adding a second plot specifically showing wind speeds. More broadly, I would suggest zooming in on this region to make it easier to see smaller-scale differences and confirm the statements that are being made in the text, which I don't find entirely supported by what I am able to see. Also, I see no value in including the vorticity contours in Fig. 9. Remove.

Response: We agree that the differences discussed in Fig. 9 are subtle at the domain scale shown. We revised the text to soften the language and to describe the differences qualitatively rather than making strong quantitative claims. Given the focus on overall cold pool structure and inflow patterns at the system scale, we retain the original domain but removed the vorticity contours to improve clarity.

The relevant sentences now read as follows:

" In the DA experiments, relatively stronger surface winds are present in the cold pool immediately behind the gust front compared to NODA, and the overall cold pool structure qualitatively resembles that in URMA (Figs. 9c and 9d)."; "Further south and east of the gust front, over Kansas, Missouri, and Iowa, the DA experiments exhibit stronger southerly inflow relative to NODA, with flow patterns that are qualitatively closer to URMA."



• (f) Line 454: "80%" – No; only NODA appears to have any 80% probabilities in it. Re write.

Response: We agree and thank the reviewer for catching this. The ">80%" threshold was incorrect. We revised the text to describe the high-probability corridor using a lower threshold consistent with Fig. 12 (e.g., >60%) and softened the wording accordingly.

The relevant sentences now read as follows:

"Compared to NODA and the reference forecasts, both GSI-New and JEDI-New produce higher neighborhood probabilities and improved spatial alignment with the observed severe reports. GSI-New shows a well-defined swath of probabilities exceeding 60%

extending from northern Iowa into southern Minnesota, collocated with numerous tornado and wind reports."

• (g) Line 463: Check the labels of Fig. 12, because this description does not match my impression from Fig. 12a. If Fig. 12a is the NODA forecast, then I'd say the NODA forecast tended to broadly capture the spatial distribution/envelope of severe weather reports, but it was too broad and a bit overdone in the gap in southwest MN and also had a bit of a southwestward displacement. Slide the whole region northeastward 50 km and it would look a lot better.

Response: We agree with the reviewer and revised the description of the NODA forecast to reflect that it broadly captures the spatial envelope of the severe reports but produces a probability swath that is overly broad and southwestward displaced relative to the observations.

The relevant sentences now read as follows:

" The NODA experiment broadly captures the spatial envelope of the observed severe weather reports, but the UH probabilities are too diffuse and exhibit a southwestward displacement relative to the report clusters, particularly over southwestern Minnesota."

(2) Line 199: 214 hPa seems too low in the troposphere for this application. Check this value. If this value is true, then there will be concerns about the validity of this model to capture the observed event if the model top is lower than storm tops (which is probably the case here).

We have employed the vertical coordinate tested in the RRFS system. This will have 65 vertical layers (66 interface levels) with the top pressure of 212.637 pascal. This is a typo and thus, we have revised the paragraph as follows.

"In the vertical, a terrain-following hybrid sigma-pressure coordinate with 65 vertical layers is employed, with the model top located near 2 hPa, consistent with RRFS_v1beta."

The below is an example of describing the vertical coordinate we have employed.

Yokota, S., Carley, J. R., Lei, T., Liu, S., Kleist, D. T., Wang, Y., & Wang, X.: Scale-and variable-dependent localization for 3DEnVar data assimilation in the rapid refresh forecast system. *Journal of Advances in Modeling Earth Systems*, *16*(11), e2023MS004098, https://doi.org/10.1029/2023MS004098, 2024.

(3) Lines 396-398: No! The convective system in this case never traversed that region. Also, the winds in that region are out of the west, very much in opposition to what they would be had the cold pool occurred in the area. This error has a different source. This must

be corrected. Hint: analyzed surface conditions by WPC (https://www.wpc.ncep.noaa.gov/html/sfc-zoom.php and select the proper date in the calendar) clearly show a cold front and low pressure system splitting the Dakotas east-to west at 00Z.

Response: We agree with the reviewer and revised the text to remove unsupported storm-scale explanations. The low θ_e is now interpreted in the context of the synoptic-scale frontal environment, consistent with the Weather Prediction Center surface analysis at 00 UTC. The wording was adjusted to avoid overinterpretation in the absence of direct evidence shown in the manuscript.

The relevant sentences now read as follows:

"The relatively low θ_e over northwest Nebraska and southern South Dakota in both NODA and the DA experiments is likely associated with the synoptic-scale frontal environment. This interpretation is consistent with the Weather Prediction Center surface analysis at 00 UTC, which depicts a cold front and an associated low-pressure system extending across the Dakotas, accompanied by westerly to northwesterly postfrontal flow. This suggests that the θ_e bias in this region is more closely related to errors in the large-scale air-mass structure, such as the frontal position or intensity, which are not directly constrained by radar DA. Assimilating additional conventional surface observations may help reduce these near-surface thermodynamic biases (e.g., Schenkman et al. 2011)."

(4) I noted a number of literature references in the body of the text that were not found in the references list. The authors need to carefully check that all stated references are provided in the reference list at the end of the manuscript, or to remove references from the text if a full citation will not be provided. Examples include Courtier et al. (1994) on line 49, Benjamin et al. (2016) on line 79, and Carley et al. (2023) on line 80.

Response: We thank the reviewer for pointing this out. We have carefully checked all citations in the manuscript and verified that every reference cited in the text is now included in the reference list. The reference list has been thoroughly reviewed to ensure consistency between in-text citations and full bibliographic entries.

(5) Lines 78 and 102: For one, I insist that the Dowell et al. (2022) reference is the appropriate modern reference for the version of HRRR used herein (compared to Benjamin et al. 2016, which really only covers the RAP, not the HRRR). In addition I would include Weygandt et al. (2022) (https://doi.org/10.1175/WAF-D-21-0142.1) as it discusses additional configurations of how radar reflectivity is used to spin-up HRRR forecasts. In addition, as covered in Weygandt et al., HRRR uses forced latent heating

during a 1-hour spinup cycle based on observed radar reflectivity as an additional means of indirectly assimilating radar data into its model atmosphere. This note should be included somewhere.

Response: We have revised the description of HRRR to use Dowell et al. (2022) as the primary reference and added Weygandt et al. (2022) to explicitly describe the radar-based latent heating spin-up procedure used in HRRR. The corresponding text in the Introduction and experiment description has been updated to reflect the current operational configuration more accurately.

The relevant sentences are now read as follows:

”...... indirect assimilation through cloud analysis (e.g., Banos et al. 2022; Dowell et al. 2022), which is used operationally in the High-Resolution Rapid Refresh (HRRR) system, together with additional radar-based latent heating during a one-hour spin-up cycle (Dowell et al. 2022; Weygandt et al. 2022).”

“The forecasts are further compared with those from the operational HRRR, which employs indirect assimilation of radar reflectivity through cloud analysis and radar-based latent heating during a one-hour spin-up cycle (Dowell et al. 2022; Weygandt et al. 2022), and an experimental RRFS configuration that uses the reflectivity-state-variable approach.”

(6) Equation 3: re-check this formula. I looked at the Thompson MP code from the WRF (v4.7), specifically, the calc_refl10cm code, and came up with

$$Z_{er} = \frac{10^{18}\times 720(\rho q_r)^2}{\pi^2 {\rho_r}^2 N_{tr}}, \qquad (3)$$

you have an extra ρ (air density; it should not be in the squared parentheses) and I don't know where the 10^{18} came from.

Response: We thank the reviewer for checking Eq. (3) against the WRF Thompson microphysics implementation (calc_refl10cm). We would like to clarify that the apparent differences arise from the definitions and units of the variables used in the manuscript versus those used in the WRF code, rather than from an inconsistency in the underlying formulation.

In the manuscript, we follow the formulation and unit conventions used in Tong and Xue (2005), in which the equivalent reflectivity factor is expressed in the conventional units of mm^6 m^{-3}. This introduces a factor of 10^{18}, which results from converting drop diameters from meters to millimeters in the Rayleigh-scattering expression $(10^3)^6 = 10^{18}$.

Regarding the air density ρ_a, the reviewer’s expression uses the total number concentration N_{tr} defined per unit volume (m^{-3}) as in the WRF code. In our manuscript,

N_{tr} is defined on a per-mass basis (kg^{-1}) following Tong and Xue (2005). Converting between these two definitions introduces a factor of air density via

$$N_{tr}^{(m^{-3})} = \rho_a \, N_{tr}^{(kg^{-1})},$$

which explains why ρ_a appears differently in the two expressions. When written using a consistent definition of N_{tr} and a consistent unit system, the manuscript expression is equivalent to the WRF Thompson formulation.

We have revised the manuscript and explicitly stated the units.

Tong, M., and M. Xue, 2005: Ensemble Kalman Filter Assimilation of Doppler Radar Data with a Compressible Nonhydrostatic Model: OSS Experiments. *Mon. Wea. Rev.*, **133**, 1789–1807, https://doi.org/10.1175/MWR2898.1.

(7) Lines 161-171: The authors need to substantially re-write this paragraph to better summarize the activity presented in Fig. 1. They need to avoid the repetitive nature of the text compared to previous paragraphs while still explaining that the test of the new formulations of snow and graupel reflectivity are a better fit to the diagnosed values than the previous formulations. For example, the text on lines 163-164 is a nearly verbatim repetition of that from lines 125-126, and that on lines 167-168 is a nearly verbatim repetition of that from lines 127-128. Line 170 features text that was already stated on lines 131 that adds nothing of value. Additional corrections in this paragraph include: a) 162 – include the version of the WRF that was run for this test so that reproducibility is guaranteed (even though the Thompson scheme hasn't changed much over the years). b) On line 168 change "diagnosed spread" to "spread of diagnosed reflectivity." c) The sentence beginning on line 170 also needs to be capitalized. d) Finally, you should explain the dramatic difference in the diagnosed reflectivities for snow between the old and new formulations (line ~166) by mentioning how clear Fig. 1 makes it that the original formulation for Zs was not built to handle the impact on reflectivity due to the various behaviors of snow (wet/dry etc.).

Response: We have substantially rewritten the paragraph describing Fig. 1 to remove repetitive material, focus on the comparison results, and explicitly explain the large improvement for snow in the updated formulation. The WRF version used for the diagnosis has been added, and the wording has been revised for clarity and precision as suggested.

The relevant paragraphs now read as follows:

"To assess the consistency of the reflectivity observation operator with the Thompson microphysics scheme, reflectivity values computed using the updated and original operators are compared against reflectivity diagnosed directly from WRF (v4.7) model output (Fig. 1). For snow, the diagnosed reflectivity exhibits a wide spread at a given mixing ratio, reflecting the complex treatment of snow microphysics in the Thompson scheme, including temperature-dependent behavior and transitions between dry and wet snow. The original operator, which does not explicitly account for these effects, shows poor agreement with the diagnosed values (r^2 = 0.165). In contrast, the updated operator introduces separate temperature-dependent relations for dry and wet snow, resulting in a much closer fit to the diagnosed reflectivity, particularly for the wet-snow branch, and substantially improves the agreement (r^2 = 0.996).

For graupel, the diagnosed reflectivity also displays considerable spread due to state-dependent variations in particle density and intercept parameter. By explicitly incorporating these temperature- and density-dependent effects, the updated graupel formulation better captures the spread of diagnosed reflectivity and improves the agreement relative to the original operator (r^2 = 0.968 versus 0.781). Overall, the updated relations provide compact power-law approximations to the Thompson microphysics while retaining its key physical dependencies and ensuring smooth behavior suitable for tangent-linear and adjoint implementations."

(8) Lines 300-301: There is still a similar looking reflectivity pattern in NODA, so why not adjust the location of the vertical cross section so you can show it in Fig. 6? I think it is shortsighted/disingenuous to force the comparison to look at fixed cross sections rather than highlighting the area of interest.

Response:

We thank the reviewer for this comment. The vertical cross section shown in Fig. 6 follows a fixed transect (as indicated in Fig. 5) that is applied consistently across all experiments to facilitate direct comparison of vertical storm structure. We agree that the NODA experiment still produces a similar reflectivity pattern in plan view, but its alignment with the fixed cross section differs from that in MRMS and the DA experiments due to positional differences in the analyzed system. We have therefore revised the text to clarify this point and softened the wording to avoid overstating the displacement, while retaining the fixed cross-section approach for consistency.

The relevant sentences are now read as follows:

"In contrast, the NODA analysis (Fig. 6e), which does not assimilate radar observations, shows a convective reflectivity structure that is less well aligned with the fixed cross section relative to MRMS, reflecting differences in the analyzed system position."

(9) Lines 240-241: It doesn't appear that you evaluate the ensemble forecasts, though, so it is not appropriate to mention this. If you did actually evaluate the probabilistic forecasts later in the manuscript, then the presentation comparing these results to the HRRR and RRFSp2 is incompatible since you're comparing an ensemble forecast to a deterministic one. Also, the ensemble element is not shown in Fig. 3.

Response: it should be "deterministic forecast" instead of "ensemble forecast".

(10) Lines 409-410: is that true? Does neighborhood radius = ΔX imply no neighborhood at all or is it really a 1-grid-square-sized radius around each grid point (which is still a neighborhood)? I find it difficult to believe that the forecasts were truly skillful at the grid scale all the way through 6 hours of forecast, even at 25 dBZ. How are you aggregating the data to produce this single plot covering 6 hours of forecast?

Response

We thank the reviewer for pointing this out. We agree that a neighborhood radius of 3 km does not imply point-wise (grid-scale) verification, but rather represents the smallest evaluated neighborhood in the FSS framework. We have therefore revised the text to avoid implying grid-point accuracy.

The sentence has been revised to:

"For the 25 dBZ threshold, both DA experiments exceed the minimum skillful threshold at the smallest evaluated neighborhood radius (3 km), suggesting improved representation of convective-scale features in a neighborhood sense when averaged over the 6-h verification period."

(11) (a) Lines 176-186: This event was a robust derecho and should be described as such. (b) Line 180: you're describing a synoptic scale setup but not including any graphics to show. Include a note that figures showing this synoptic scale setup are omitted/not shown for brevity.

Response

We thank the reviewer for this helpful suggestion. We agree that this event meets the definition of a derecho and have revised the text accordingly to explicitly describe it as a derecho-producing convective system.

We also agree that the synoptic-scale description would normally be supported by additional figures. To avoid expanding the scope of the paper, we have added a clarifying note indicating that figures illustrating the synoptic-scale environment are not shown for brevity.

The sentences have been revised to:

"Figure 2 shows the observed composite reflectivity from the Multi-Radar Multi-Sensor (MRMS; Smith et al. 2016) at 0000 UTC 13 May 2022. The event is a robust derecho-producing convective system, characterized by a well-organized, north–south-oriented convective line with a pronounced bowing segment over the Upper Midwest. A negatively tilted upper-level trough over the Midwest, combined with a low-level jet under an omega block pattern, facilitated northward transport of moisture and instability (synoptic-scale figures are not shown for brevity)."

(12) Line 198: this should be Fig 2. Also, mention in the Fig. 2 caption that the displayed domain is the FV3-LAM model domain.

Response： Thank you for pointing this out. We have corrected the figure reference in line 198 to Fig. 2. We have also updated the Fig. 2 caption to explicitly state that the displayed domain corresponds to the FV3-LAM model domain.

(13) Lines 235 and 369-370: There is no mention of assimilation of conventional or mesonet data until lines 369-370. I suggest moving this further up, to line 235. Also, does this mean that in NODA the model could effectively be run straight through the 6-hour spin-up (in which the other experiments are performing DA steps)?

Response

We thank the reviewer for pointing this out. We agree that the description of assimilated observations should be clarified earlier in the manuscript. We have revised the experiment description (around line 235) to explicitly state that only radar reflectivity and radial velocity are assimilated in the DA experiments, and that no data assimilation is performed in the NODA experiment.

We also clarify that the NODA experiment is a free model integration over the same 6-h period, whereas the other experiments perform hourly radar DA cycles.

(14) Use consistent naming for experiments. There are variations throughout: (a) Line 310: “GSI” is an incomplete experiment name (b) Figure 9 caption (c) Line 325: A lot of text used to say “GSI-New”

Response: We replaced “GSI” with “GSI-New”.

(15) Figure 11: not much data is actually plotted here. This could be a table instead.

Response: We thank the reviewer for this suggestion. While Fig. 11 summarizes a limited number of aggregated verification points, we chose to present the results using performance diagrams because they provide a compact and widely used framework to simultaneously illustrate the relationships among probability of detection, false alarm ratio, frequency bias, and critical success index. These multi-dimensional characteristics cannot be conveyed as effectively in tabular form. We therefore retain Fig. 11 to facilitate a direct and intuitive comparison of forecast performance across different systems.

(16) Lines 216-218: if these symbols aren't used later in the manuscript, there's no need to introduce them now. Also, the symbol for number concentration for rain is strange. Wouldn't you want to use Ntr as was used in eqns. (2) and (3)?

Response: The symbols that are not used elsewhere in the manuscript have been removed to improve clarity.

(17) (a) Line 456: a minor westward displacement where? Be specific. (b) Line 457: “two primary UH clusters” is inappropriate. You’re referring to storm report clusters; no UH is depicted in Fig. 2 (nor could it be).

Response: We agree with the reviewer. The phrase “minor westward displacement” was removed because the spatial offset between GSI-New and JEDI-New is not sufficiently clear to state robustly. We revised the sentence to describe the common probability swath and its correspondence with the clusters of severe storm reports. We also replaced “two primary UH clusters” with “two primary clusters of severe storm reports,” since Fig. 2 shows MRMS reflectivity rather than UH.

(18) Lines 252 and 481: “significantly” appears in both places referring to decreased RMSI in all experiments compared to NODA. Is this statistical significance or just a modifier word to enhance the impression of the impact?

Response: We thank the reviewer for pointing this out. We agree that the term “significantly” may be interpreted as implying statistical significance, which is not intended here given that the results are based on a single case without formal significance testing. We have therefore revised the wording to use more appropriate descriptive terms (“substantially”) to reflect the magnitude of the reduction without implying statistical significance.

(19) Figs. 4a,b: why is the RMSIZ = 12 line delineated in these panels?

This was intended to show differences between backgrounds and analysis more clearly but this was not required. Thus, we have removed the dashed reference line at Figure 4-a,b

(20) Lines 287 and 289: “hook echo” is not the correct term to describe this reflectivity pattern.

Response: We agree with the reviewer that the term “hook echo” is too specific in this context. We have revised the text to use a more descriptive and less interpretive phrase (“a curved, localized reflectivity enhancement in the southern portion of the convective line.

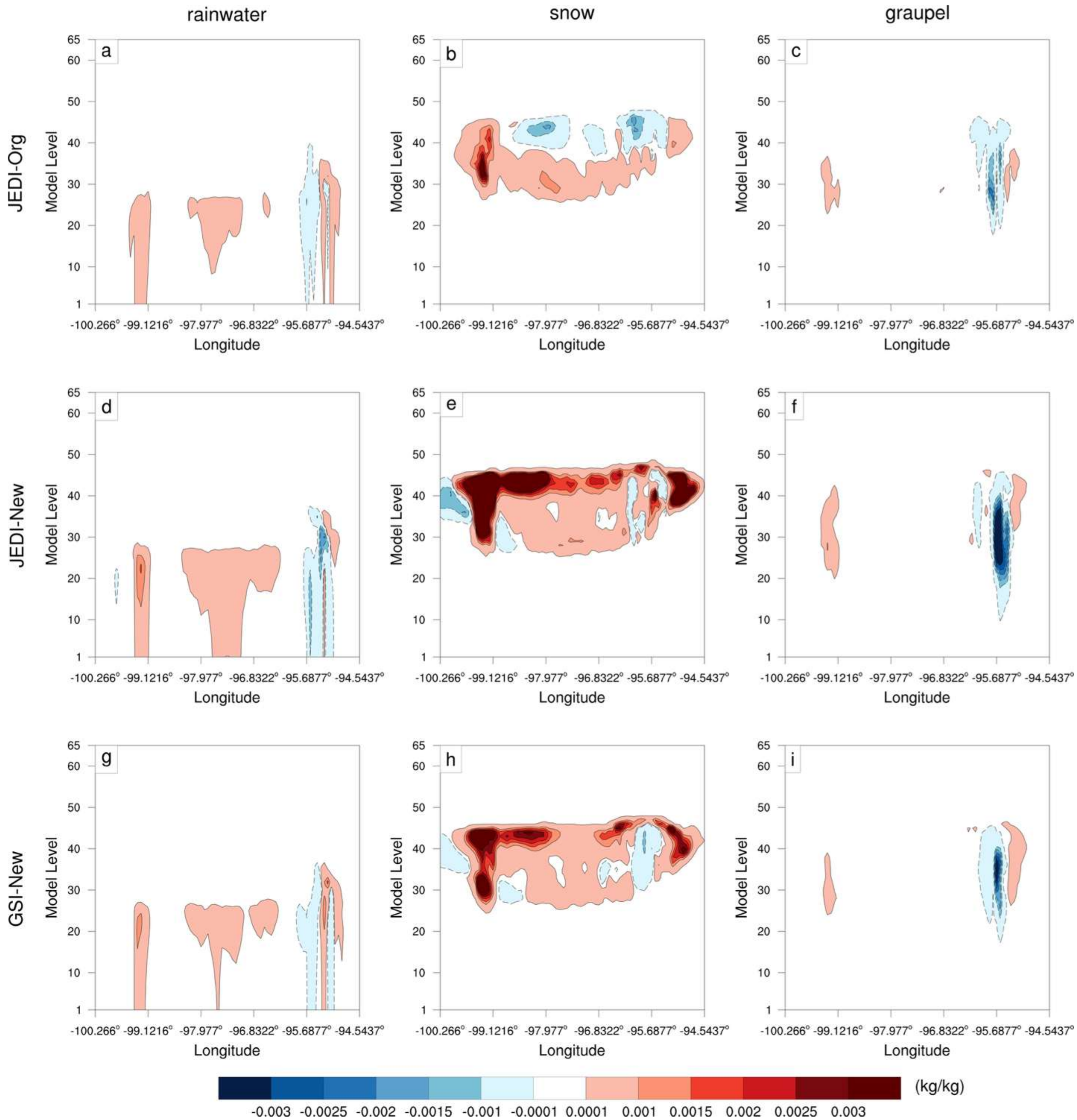
”) to avoid implying a definitive supercell or tornado-related signature.

(21) Line 304: The eastern updraft column in Fig. 6d is high-biased almost entirely throughout. I’d say “2-8” km is a better range than “5-8”.

Response: We have revised the vertical range to 2–8 km to better reflect the extent of the high bias in the eastern convective column.

(22) Figure 7: No units in figure or caption. Also, the very bottom of the image appears to be cutoff.

Response: Thank you for the comment. We have added the units for the hydrometeor mixing ratio increments in the caption of Fig. 7. We also checked the figure rendering in the revised PDF and adjusted the figure placement/margins to ensure that the bottom portion is fully visible.



(23) Line 34: Note that HRRRE is no longer an entity. It was terminated a few years ago as development of RRFS became dominant (and work with WRF ceased at the labs developing RRFS). It would be better to cite the RRFS Ensemble System, or REFS. However, there are no formal references for REFS. But you can use this link and possibly cite Matt Pyle as a personal communication reference: https://www.emc.ncep.noaa.gov/mmb/mpyle/refs/

Response: We thank the reviewer for the clarification. We agree that HRRRE is no longer an active development pathway within NOAA. We have revised the text to clarify that HRRRE is referenced as a past convection-allowing ensemble system and to better reflect the ongoing transition toward RRFS-based ensemble frameworks.

(24) Line 183: as of my check of SPC's severe reports on 22 December, there were seven tornado reports in South Dakota during 2300-0000 UTC, not six. I was able to confirm the remainder of the report counts, though.

Response: We have revised the manuscript to reflect seven tornado reports in South Dakota during this period.

(25) Line 48: Would it not be more appropriate to refer to "multi-loops" since there is theoretically no limit on how many outer loops can be used?

Response: We thank the reviewer for this comment. In the variational data assimilation literature, the term *double-loop* is standard and refers to the algorithmic structure consisting of an outer loop and an inner loop, as originally formulated by Courtier et al. (1994), rather than to the number of outer-loop iterations performed. While multiple outer-loop iterations can indeed be used in practice, the terminology *double-loop* remains the conventional and widely accepted description. We therefore retain this wording for consistency with the established literature.

(26) Lines 306-307: This feature was also a smaller-scale event that didn't last very long. So I don't think it says much either way about the forecast system that backgrounds didn't fully capture it. There appears to at least be a hint of it in some of the forecasts. Bottom line: I don't know if this miss by the forecasts is critical to the discussion. But you can leave this text unedited if you wish.

Response: We corrected the longitude range of the western secondary tower to be consistent with Fig. 6a and revised the wording to avoid an overly absolute statement.

The sentences have been revised to: "Additionally, the western secondary tower near 99°W that is evident in the MRMS cross section is not clearly reproduced in any of the background fields."

(27) Lines 222-227: I realize it is outside the scope of this research, but did you examine how much impact the Vr assimilation has in this case? Does it change things much compared to only assimilating Z? No changes to the paper are required.

Response:

We thank the reviewer for this insightful question. In this study, radar reflectivity and radial velocity were assimilated jointly, consistent with common operational practice, and our primary focus was on demonstrating and evaluating the implementation of the new reflectivity observation operator within the JEDI En3DVar framework. A dedicated sensitivity study to isolate the relative impacts of reflectivity-only versus combined reflectivity and radial-velocity assimilation was beyond the scope of this work. We agree that such an analysis would be valuable and plan to investigate the respective roles of Z and Vr assimilation in future studies.

Technical/writing corrections

-Line 33: insert “an” before “ensemble of analyses” -Remove excessive commas on lines 40 (“covariance matrix,”), 372, and 373

-Line 64: “employ” should be “employing.” Also remove the semicolon. No punctuation is necessary to separate the clauses.

-Line 67: eliminate the paragraph break. This content is still just as relevant to the previous paragraph.

-Line 73: delete “used”

-Line 74: GSI was introduced previously. Move the acronym definition and this reference to the first time it appears rather than here.

-Lines 176-178: This is a poorly structured sentence. Re-write.

-Line 207: remove “ensemble inflation”

-Line 211: delete “for” after “within”

-Lines 232 and 233: delete “initial” on line 232 and change “first” on line 233 to “subsequent”

-Line 254: “show” should be “shows” and “lower” is inappropriate. Use “reduction/decrease in”

-Figure 5 includes a lot of extraneous text. Delete “observation space diagnostics from” and “using the original and modified Z observation operators”

-Line 299: the displacement is southward, not northward.

-Line 324: should be level 25, not 15

-Lines 338-340: This sentence seems obvious and adds nothing. Re-write or delete.

-Line 351: insert “the” before “NODA”

-Line 354: insert “the” before “central CONUS”

-Line 368: “northern and northwestern” seems repetitive. Also, are you referring to the mostly non-convective reflectivity over the northwestern corner of the plotted domain?

-Lines 388-389, numerous corrections: insert “are” before “much smaller”, insert “in” before “URMA”, “guest” should be “gust”, and “is URMA” should be “in URMA”

-Line 408: Looks like 66 km to me, not 63.

-Lines 425-427: it appears a line of text was duplicated.

-Line 440: use “less” instead of “weaker”

-Line 451: “tornado” should be “tornadoes” and “strong wind” should be “severe wind gusts”

-Line 461: I suggest changing “larger” to “higher”

-Line 495: “a Z-state-variable” is a bit colloquial. Re-write. I don’t think “a” should be there to fit the parallelism setup earlier in the sentence.

Response:
We thank the reviewer for the detailed technical and writing suggestions. All listed editorial, grammatical, and stylistic corrections have been carefully addressed in the revised manuscript. This includes adjustments to articles, verb tense, punctuation, wording precision, figure captions, and terminology to improve clarity, consistency, and parallelism throughout the paper.

In addition, several sentences and passages were restructured, merged, or removed in response to earlier substantive comments. As a result, a small number of the line-specific suggestions listed here are no longer directly applicable in the revised version because the referenced text has been modified or deleted.

Reviewer 2

Recommendation: Major revisions

--- Overview

This study develops a new radar observation operator within a JEDI ensemble-variational data assimilation framework. The new operator is consistent with the forecast microphysics scheme and demonstrates notable advantages in a high-impact squall line event.

The authors provide an excellent evaluation of the new radar operator, including detailed comparisons with the still-operational GSI system, as well as operational forecasts from HRRR and the experimental RRFS. I appreciated the discussion of how radar increments project onto different hydrometeor species, along with the combination of subjective and objective verification statistics. The results from the new radar operator in JEDI are promising and represent an important step toward broader adoption of JEDI workflows.

I only have two major comments, which are summarized below, and followed by minor suggestions.

--- Major comments

1. The introduction text up to line 75 should be considerably shortened, as it does not directly relate to the research topic of this paper. Instead, the authors should focus the narrative on the subsequent text which pertains to the type of radar DA issues addressed in this manuscript.
   Response: We have substantially shortened and refocused the Introduction by removing general background material on ensemble and variational data assimilation and concentrating the discussion on issues directly relevant to radar reflectivity assimilation, observation operator nonlinearity, and microphysics consistency. The revised Introduction now more clearly motivates the development and evaluation of the reflectivity operator within the JEDI ensemble–variational framework.
   Response: We have substantially shortened and refocused the Introduction by removing general background material on ensemble and variational data assimilation and concentrating the discussion on issues directly relevant to radar reflectivity assimilation, observation operator nonlinearity, and microphysics consistency. The revised Introduction now more clearly motivates the development and evaluation of the reflectivity operator within the JEDI ensemble–variational framework.

2. The results in Fig. 4 are promising. However, unlike radial velocity, the reflectivity-based innovation statistics in Fig. 4 do not show signs of convergence toward stable values, indicating that six cycles may not be sufficient. In the DA literature, fair comparisons traditionally discard these spin-up periods during which DA diagnostics exhibit mean trends. Since the paper focuses on a single case, I believe extending the experiment to additional cycles is worth pursuing and would provide a more convincing demonstration of the benefits of the new radar operator developments in JEDI.
   Response: We thank the reviewer for this insightful comment regarding the reflectivity-based innovation statistics in Fig. 4. We agree that, unlike radial velocity, the reflectivity diagnostics do not fully converge to stationary values within the six data assimilation cycles shown, and that longer spin-up periods are often desirable in studies aimed at rigorous statistical comparison.
   The primary objective of this paper, however, is to demonstrate the correctness, stability, and practical feasibility of the newly developed radar reflectivity observation operator and its tangent-linear and adjoint implementations within the JEDI ensemble–variational framework. From this development and technical perspective, the selected number of radar data assimilation cycles is sufficient to verify the qualitative behavior of the innovation statistics, the relative performance between operator formulations, and the consistency of the system across cycles, rather than to establish fully equilibrated innovation statistics.
   We have revised the manuscript to clarify this scope earlier in the Introduction and Discussion, explicitly noting that the presented results are based on a single high-impact case and a limited number of cycles, and that extended cycling and multi-case evaluation would be valuable for assessing longer-term convergence and statistical robustness. Such additional testing was constrained by available computational and project resources. Nevertheless, the capabilities developed in this study have been incorporated into the official JEDI repository, and this paper serves as technical documentation to support future extended evaluations by the broader community.

--- Minor comments

Please check whether acronyms for ensemble-variational methods are used consistently throughout the paper. For example, lines 50–51 state that Wang and Wang (2017) developed an En3DVar system; the correct acronym should be 3DEnVar.

Response:
We thank the reviewer for pointing this out. In this manuscript, we intentionally use the term *En3DVar* to denote ensemble–variational methods in order to maintain consistency with our previous studies (e.g., Chen et al. 2021; Li et al. 2022; Liu et al. 2022), where the same terminology was adopted. While *3DEnVar* and *En3DVar* refer to the same class of ensemble–variational algorithms, we have ensured that the terminology is used consistently throughout the manuscript.

Line 114: This refers to the radar reflectivity factor, not simply reflectivity.
Response: We thank the reviewer for the clarification. The text has been revised to explicitly refer to the radar reflectivity factor and to distinguish between Ze and its logarithmic representation in dBZ.

Fig. 1 and surrounding discussion: The revised text should clarify how the WRF reflectivity values are calculated.

Response: We have revised the text to clarify that the reference reflectivity values in Fig. 1 are diagnosed using the native reflectivity calculation of the Thompson microphysics scheme in WRF (v4.7).

Section 3.2.2: It would be helpful to write out the full DA equations to make this paper self-contained.

Response: We thank the reviewer for this suggestion. The full mathematical formulation of the EnVar/En3DVar data assimilation framework is well documented in previous studies and, more specifically, for the JEDI framework in Liu et al. (2022).

The focus of the present study is not on the development of a new DA algorithm, but on the implementation and evaluation of a microphysics-consistent radar reflectivity observation operator within existing En3DVar systems in GSI and JEDI. To avoid redundancy and excessive length, we therefore do not reproduce the full DA equations here.

We have clarified this scope in the manuscript and added appropriate references (Liu et al. 2022) to guide interested readers to detailed descriptions of the underlying EnVar formulation.

Liu, Z., Snyder, C., Guerrette, J. J., Jung, B.-J., Ban, J., Vahl, S., Wu, Y., Trémolet, Y., Auligné, T., Ménétrier, B., Shlyaeva, A., Herbener, S., Liu, E., Holdaway, D., and Johnson, B. T.: Data assimilation for the Model for Prediction Across Scales – Atmosphere with the Joint Effort

for Data assimilation Integration (JEDI-MPAS 1.0.0): EnVar implementation and evaluation, Geosci. Model Dev., 15, 7859–7878, https://doi.org/10.5194/gmd-15-7859-2022, 2022.

Line 235: The NODA experiment does not assimilate radar data. What about other conventional observations? Please clarify in the revised manuscript.

Response: We have clarified that no conventional observations are assimilated in any of the experiments. The NODA experiment is a free forecast without any data assimilation, while the other experiments perform hourly radar-only data assimilation cycles. The revised text has been added in Section 3.2 to avoid ambiguity.

Lines 256–258: Can you briefly comment on what causes the positive biases?

Response: We agree with the reviewer's comment and have revised the manuscript to briefly clarify the source of the positive reflectivity biases in the NODA experiment. Specifically, we now note that these biases primarily arise from displacement and structural errors in the simulated convective system, rather than from a systematic overprediction of reflectivity intensity.

When referring to states in spatial maps, it would be helpful to label them directly on the maps. It would also be helpful to include title labels for each plot; for example, experiment names in Fig. 12.

Response: We agree with the reviewer that clearer geographic context would improve the readability of the maps. We have revised the figures to explicitly label both the experiment names (e.g., NODA, GSI-New, JEDI-New, HRRRv4, RRFS p2 CNTL) and the relevant U.S. state abbreviations (e.g., SD, MN, IA, MO) directly on the maps. These additions help readers more easily identify the referenced regions and experiments without altering the underlying results.

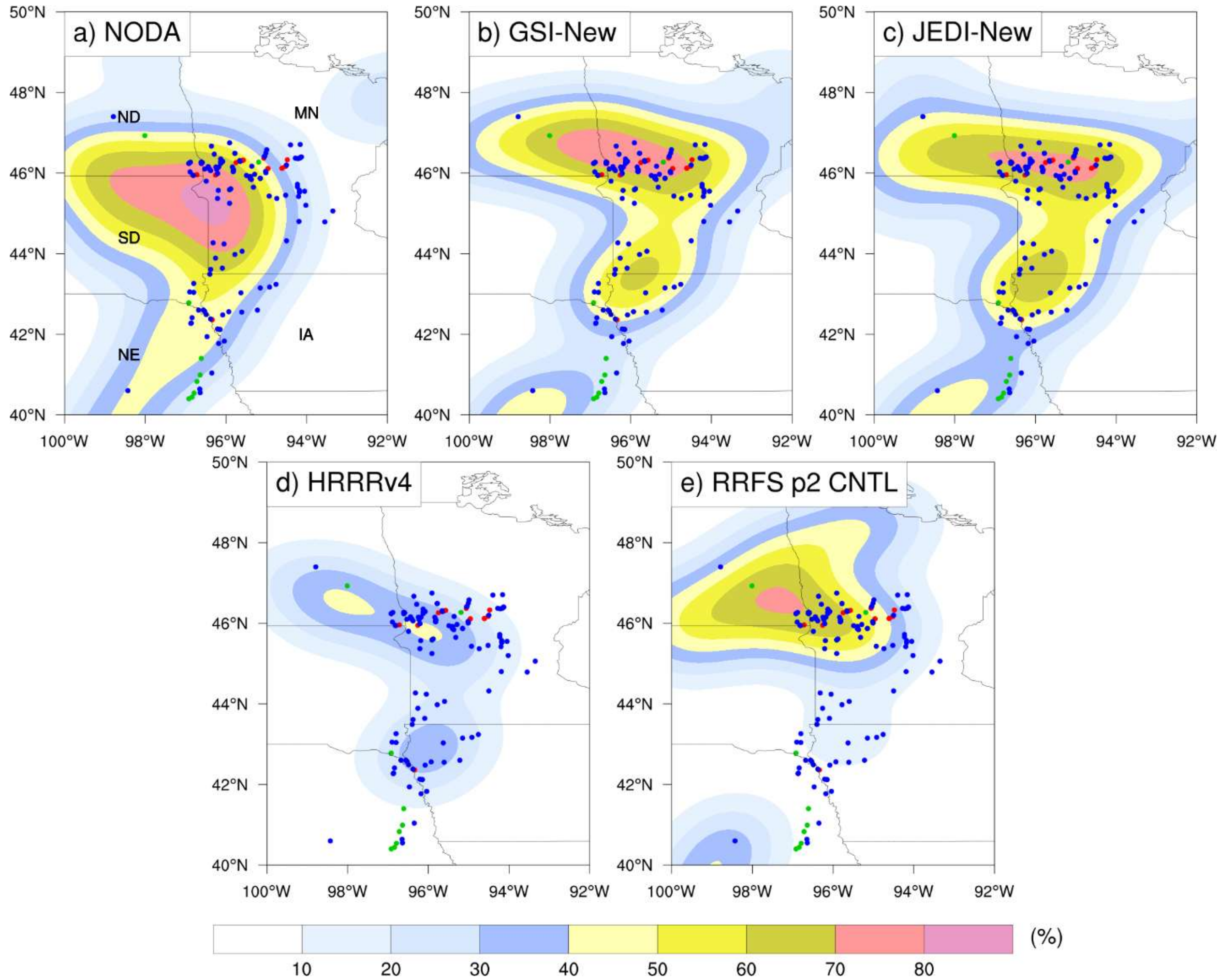


Fig. 9: In addition to underestimating surface temperatures, it should be noted that the DA experiments produce a stronger outflow. This behavior is consistent with cold pool dynamics (stronger cold pools correspond to stronger winds).

Response: We have revised the discussion of Fig. 9 to explicitly note that the stronger near-surface winds in the DA experiments are physically consistent with enhanced cold pool dynamics.

Line 463: "The NODA experiment … with weak, spatially inconsistent probabilities" — The NODA probabilities appear to be much higher than those of any other experiment; please reword this statement.

Response:

We agree with the reviewer. The original wording overstated the weakness of the NODA probabilities. We have revised the text to clarify that NODA broadly captures the spatial envelope of severe weather reports but produces overly diffuse UH probabilities with noticeable spatial displacement relative to the observed report clusters.