

Summary/overall

I have carefully read and reviewed Park et al.'s preprint submission to EGU sphere 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 semi-major 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.

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).
- (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.
- (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.
- (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.

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

(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).

(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.

(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.

(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.

(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

$$Z_{er} = \frac{720 * \rho_a * q_r^2}{(\pi \rho_w)^2 N_{tr}}$$

So 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.

(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 Z_s was not built to handle the impact on reflectivity due to the various behaviors of snow (wet/dry etc.).

(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.

(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.

(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?

(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.

(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.

(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)?

(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"

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

(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 N_r as was used in eqns. (2) and (3)?

(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).

(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?

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

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

(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".

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

(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/>

(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.

(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?

(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.

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

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.