Commendations to the authors on an interesting computation, and congratulations to the student on a PhD well earned.

However, this manuscript seems to be not well written for the professional literature, rather it is in the somewhat tedious style of a dissertation. The captive audience of a reader committee would make all their check marks: enough effort reported, quantitative analysis technique mastery, awareness of caveats, and the student made thoughtful connections to some prior papers in the literature. But earning the attention of readers and (hopefully) citation in the professional literature is a different enterprise, and a major revision should be undertaken for that purpose. However, I do not volunteer to review again - I had forgotten that I tend to dislike ACP, perhaps for its apparent incentives toward first-draft quality or length. I haven't studied the business model, but experiences seem mixed.

Main science points of the manuscript as I see them:

* Second-generation development of deep convection in a triggered, highly unstable sounding is a delicate process, which adds a burst of difference growth to an ensemble, and introduces a need for feature-centered rather than Eulerian measures.

* The main mechanisms of this delicate secondary development involve a combination of cold-pool gust front lifting plus the strength of thermally forced gravity wave displacements in the inversion layer of the lower free troposphere.
* Low-level shear enhances gust front lifting on the downshear side, adding asymmetry that is convenient for analysis purposes. It also has longer-term effects on the resulting systems, but those are not really evident here.

* Because the redevelopment process is delicate and amplifies differences, any small variation can generate an ensemble of second-generation outcomes. Here the low-level shear strength (distractively cast as an altitude difference of the linear shear layer's top which is smaller than the grid spacing!) happens to have been used. There is no indication that the results are monotonic with this shear strength perturbation: ENS-05 has the largest perturbation, but ENS-03 has the biggest outcome difference, and the final Abstract sentence suggests so too. If the perturbation size and outcome difference magnitude are uncorrelated, then emphasizing the role of shear is misleading: lucky resonances among cold pools and internal waves aloft seem to be the main unstable difference-makers, and anything might cause ensemble divergence.

* Secondary development or non-development affects both the gust front and deep convection+wave fields at later times, of course (Fig. 3), and has clear correlates in the top height, width, volume, and thus mass flux of the squall line at any given time (however it is measured; the outlines in Fig. 5 are as informative or more so than the colored tracer difference that predominates in the attention of the viewer).

* The time scale of deep difference development is about 40 minutes, at which time about half the 80-minute ensemble spread has been established already (Fig. 6).

* Outcomes (secondary develop/non-develop) appear to be coherent in the vertical at least up to midlevels (there is little curve crossing in Fig. 10... why not color all 10 lines, either arbitrarily or by the value of their initial perturbation z_top, then making reference and ENS-3 bolder for clarity?)

High-level style critique of manuscript:

Many of the key findings are quite buried behind overly detailed and sometimes defensive material about methods in a deep, murky dissertation-style text. This does not serve a professional reader's interests optimally, since at this point the student does not need to prove amount of effort or technical mastery, but rather a sharp eye and crisp tongue about what is important. The slow resurfacing at the end into Synthesis and then Conclusion sections (also rather long) does help to pull these key points and highlights out somewhat, but those key points could be even more polished into the Abstract for instance. There is no role for suspense (historical or at readtime) in scientific writing. But those wrapup sections (4 and 5) come after a long slog of sometimes unclear prose in the
late-middle (the very long sections 3.3 and 3.4).

Helpfully, section 3.2 leads it off, with simple differences between the most-different ensemble members. In this reader’s view sections 3.3 and 3.4 then add little scientifically, except a glimpse of the 10 members as spaghetti plots (why not colorize all 10 members so we can see if curves cross?). An unclear overanalysis in confusing statistical terms (terms like “source” and “error”, and “auto-correlation” for intra-ensemble rather than temporal-lag correlations) were unhelpful or confusing, and it all gave few clear insights that aren’t in the bullets above and in section 3.2. Might the paper or at least sections 3.3 and 3.4 be cut by aspirationally 50% with no loss (and a gain of clarity) on the reader’s part? Long stretches of text appear to describe figures not shown, without stating (not shown). Who needs these, who will remember them an hour after reading much less a month or year?

Detailed local comments:


Lines 19-20: The meaning of this result is that the perturbations chosen are in a non-essential field, but that even those differences grow (or explode). What does the word “intrinsic limit”, lifted from some over-realm of philosophy it seems, really add to this idea?

A general game: how many words could be trimmed or eliminated without loss of meaning? In journals with page charges there are incentives; perhaps a lack of incentive is what I have not liked about ACP (does it have length charges?)?

158: “interface height” — this is just a reference value in an analytic formula, which translates into shear strength on the 100m grid, right? This description was quite confusing.

197: “dam breaking” is never mentioned, not all readers may understand this allusion to a classic problem’s metaphor

248-250: what and where are we looking?
253: does your contouring routine treat \ sloped features different from / sloped features?

276: "difference" vs. in Fig 4 caption, sign is different (or ambiguous). Why not make it clear?

291: “will be” —> “were”

301: “reduced in the reference” seems backward to the idea of a reference vs. perturbation

302: “no less than 38.3(1)...only 23.5” — these seem like rather dramatic descriptions of excessively precise numbers that are not terribly different, and how was the second one even measured?

330: “argmax” is a useful word from the code world

Fig. 6: “autocorrelation” is often used to mean a series with time lags, perhaps say “intra-ensemble correlations at the indicated time with the minute-80 ensemble spread”

374: “Now we will continue” seems to presume the reader’s commitment, can topic phrases and topic sentences be part of a thought-thread rather than a presenter’s convenience-order?

377: “latter” —> “resulting” variable?

Fig. 7: “Correlation structure” of what field? (zonal wind, correlated with the base time series which should be given a symbol: W_ref or something like that)

389: “source” —> base time series, which could be given a symbol: W_ref or something like that for clarity?

392, 392-394, 399, 403: all not shown? This narrative timeline is perhaps too much detail and drama.
399: “trough” and “crest” — what do these mean? Is this the u field, does it even show vertical displacements at all?

Fig. 8: “Squall line structure” — what does this mean? Covariance of u with the (standardized?) W_ref or whatever is the base series?

408: “circulation” — what does it mean? We are looking at complicated multi-lobed structure of the u field.

429: the narrative gets kind of sprawling here, is this storytelling all necessary?

433: “compensating” — what does this word mean? It implies a big back-story in the authors’ minds about how things are related and constrained, makes me nervous.

442-4: huh, how does removing all buoyant gridcells remove “gravity wave contributions from saturated parcels”?

445: “a lot vertical transports” —> “lot of” — but arguably these aren’t really transports, merely oscillations.

455-6: “cloud fraction” — in the y direction?

Fig. 9: so crude, circular blobs and various arrows and low quality, makes a poor impression

461: Paragraph break suggested, “Most of the ...” is an important statement and should lead a paragraph, not be buried in a very long one.

472: “circulation describes in the earlier part of this paragraph”. First off, the paragraphs are unclear (please indent them). Also, text can always be rearranged to never have to use such a crude self-referential grammatical constructs. Repeating an earlier confusion, what is this “circulation”? It seems like a complex multi-lobed bag of features in the u field, some convective and some wave-related, is that right?
480-onward: “error” —> “difference”

Fig. 11: very useful, it shows how feature-relative differences are better than simple Eulerian differences. Why not color the lines, so we can see if they cross and interleave?

550: “differential propagation” — isn’t it simply the different strength, not actually a difference in ‘propagation’? Perhaps ’propagation of differences’?

587-592: “compensation...opening up space...” This makes me nervous about possible misconceptions regarding causality and how mass continuity works and is maintained.

From line 500 or so: the paper is getting rather long and verbose... can it be streamlined? If all the key results have been shown, why not gather them crisply and close?

Congratulations again on an interesting study and well-deserved student degree.