Review: EGUSphere 2012-2148

Recommendation: Reject

This article uses a newly developed storm updraft tracking algorithm to track long-lived cells in an ensemble simulation of a day when a lot of hailstorms were experienced in Switzerland. The tracked cells are compared to radar characteristics, and mean cell and environmental characteristics are calculated. Finally, Lagrangian tracers are used within the cells to examine storm inflow over time.

I spent a lot of timing thinking about this paper when conducting my review. There is clearly a lot of quality work that was done for this project. Yet, I was concerned as each new conclusion was drawn: I kept feeling that more analysis was needed to support said conclusion. In the end, I'm not certain I'm fully comfortable with any of the drawn conclusions without inclusion of more analysis. But how to square that with how much work went into this paper?

After much thought, I believe the problem is the paper is trying to do too many things. There could be several papers' worth. I list each below and add what additional analysis would need to be included.

1. Methodology paper explaining the cell tracking algorithm and example applications
   - Add: Comparison of results with existing tracking algorithms. There are several existing tracking algorithms that have dynamical time tracking and splitting/merging (PyFlexTRKR as one example). Why was development of this algorithm necessary?

2. Examination of tracked cell characteristics in ensemble and comparison to radar observations, exploring the physical processes behind the differences.
   - Add: Track the cells in model data using reflectivity, not updraft speed, so the radar and model tracks can be compared 1-to-1.
   - Why does the ensemble underproduce long-lived cells, and very large/very small cells? What are the microphysical and/or dynamical characteristics of the cells it tends to get right, and those it misses?
   - Why does the ensemble produce a lot of cells in eastern France that weren't observed via radar? Why does it largely miss the cells in northern Italy (assuming that's not a radar gap issue)?

3. Use of model simulation to understand physical processes ongoing in the 28 June 2021 hailstorm in Switzerland. This storm is the only one with extensive surface hail observations reported in Kopp et al.; surface hail verification is highly necessary for reasons described below.
   - Add: Full verification of the modeled hail swath and radar reflectivity presentation. This is particularly important because HAILCAST is a model diagnostic (see major comment #1 below) and not explicitly produced by hail-related physical processes in the model. Kopp et al. is an excellent resource of information about the hail swath produced by this storm. How does the modelled HAILCAST hail swath compare? Is the model simulation representative of what happened? Does the convective mode of the storm and its structure appear similar in the model and radar observations? Only if so am I comfortable with using the subsequent model analysis to determine the underlying physical processes important to hail production. If the simulation is not representative of observations, the conclusions drawn about those physical processes could be incorrect.
   - Trimming the analysis to one cell instead of many will also solve the problem of many fine-scale storm (and storm-surrounding) structures being smoothed and averaged out in the environmental field analysis. This smoothing is particularly

concerning if it is being conducted across multiple convective modes and/or storm types.
    - Trimming subjects that fit under paper ideas 1 and 2 above (cell tracking algorithm, cell characteristics) will also allow for connection of this analysis to previous studies in the literature that examine the impact of storm-scale environmental flows on hailstorm development (e.g., Dennis and Kumjian 2017; Kumjian and Lombardo 2020; Lin and Kumjian 2022). The Lagrangian parcel technique the authors are using offer a novel way of exploring if the results of these idealistic studies can be seen or applied in a real-data simulation.

I would be happy to review a revised version of this paper (or multiple papers) that follows one or more of these suggestions. As the paper stands, however, I’m not comfortable with its acceptance.

Major comments: (These assume a goal of a paper like idea #3 listed above)

- Lines 65- 67: Given the fine scale features that are important to both convective updrafts (Bryan et al. 2003), I'm concerned that a convective-permitting resolution simulation alone won't be enough to address the objectives the authors have laid out re: storm structures being resolved in the simulations, particularly updrafts and downdrafts. How do the authors plan to address this issues? Analyzing results in the context of existing idealized convective-resolving simulations and noting the limitations of the coarser convective-permitting simulation is one possibility.
- Section 2.2: It isn't clear to me why an on-the-fly feature tracking algorithm like those described in Lines 131 -133 can't just be used here; it seems like it would have saved a lot of development trouble- why wasn't it?
  That being said, the whole paragraph from Lines 131-147 could be trimmed down to almost a single sentence saying that tracking requirements include the ability to handle coarser time resolution (hence dynamical tracking) and splitting/merging, which inspired the authors to create their own. The tracking algorithm isn't the point of the paper.
- The introduction is a mix of too much and yet not enough information. The purpose of ensembles (e.g., Lines 39-47) is already well-established and doesn't need to be explained. However, the authors do need to include discussion of existing research into how airflow around a storm impacts the storm and hail trajectories within it. There is some (very) small discussion on lines 72-74, but this should be expanded. Note the Prein and Heymsfield study used reanalysis data, not convective-resolving simulations.
- Sections 2. 2, 3.2: In a study like this one, where the authors are hoping to understand physical processes using a model simulation, it is important to determine how (or even if) the simulation is representative of the observations.
    - If thinking of paper idea #2: Verification is obviously difficult to do for something as transitory as deep convection, but a comparison of tracked cell lifespans and areas to observed values seemed promising. However, I am quite confused why the simulation data would be tracked via vertical velocity, a field with no observational equivalent. Why not track both the observed and simulated reflectivity, and then all the cell track information can be compared me-to-one?
    - Again, for paper idea #2 Lines 201-202: I don't know that I'd consider these runs to be a good representation given what is shown. Fig. 3 makes it look like the model runs way over-produced convective cells (although that's partially due to comparing all ensemble members to radar. Can the individual members be colored separately?)

However, if the goal is to successfully reproduce the hail swath in red, that looks like it was done by many members.

- I had a lot of trouble following if the results in Sections 4 and 5 were from just one cell or all tracked cells. The text seemed to switch back and forth frequently. I also had trouble discerning if the averaging was being conducted in time, along-track, cross-track, over multiple cells, and/or over multiple ensemble members. That's a lot of potential variability to keep straight! Honestly, given the importance of fine-scale features in hail production, I don't see what is gained by averaging the results across multiple tracked cells together (hence why I recommend focusing on just one cell). I'd like to see some standard deviation information included in Fig. 5 to provide needed context from the averages over time and/or ensemble members.

- Line 329-336: I would be careful with this analysis. HAILCAST is only a diagnostic tool, and its output shouldn't be used to explore scientific processes behind hailfall generation. Its 1d nature means it won't be able to take advantage of increased updraft width as important for hail growth, which has been established as important for a while (e.g., Nelson 1983, 1987). Thus any correlation between environmental parameters and HAILCAST hail size will be suspect. If the authors chose to carefully compare the HAILCAST size to observed hail reports, and found them to strongly correlate, then I would be more comfortable with relating HAILCAST hail size with environmental model variables.

Minor comments:

- Section 2.1.1 The version of HAILCAST you are using here is best termed "CAM-HAILCAST", or convective-allowing model-HAILCAST, to differentiate it from the HAILCAST of Brimelow and Jewell & Brimelow. Brimelow-HAILCAST used a steady-state cloud model connected to a hail growth model, as you describe in lines 113-115. CAM-HAILCAST, used herein, embeds a pseudo-Lagrangian 1D hail growth model into a convection allowing model, and should be attributed to Adams-Selin and Ziegler (2016) and Adams-Selin et al. (2019).
  Additionally, while Brimelow-HAILCAST is a 1d model, CAM-HAILCAST is "pseudo-Lagrangian" as it parameterizes hailstone horizontal motion across the updraft by adding a time- dependent updraft multiplier term (essentially a cosine curve).
- Line 119: Multiple initial embryo sizes and at multiple temperatures.
- Line 115: The CAM-HAILCAST outputs are the max, mean, and standard deviation of the different hail sizes produced by the different embryos, not of the entire storm. It would be worth including a small correction.
- Line 120: If only the hail sizes from the 10-mm embryo are used, where do the d_{hail} distributions come from in Fig. 7?
- Line 165: Minor quibble, but since we don't have observations of the hail swaths, I wouldn't say "covered" by hail. Perhaps event with reports from the most towns, or something like that.
- Line 176: "pressure distribution... was flat." What does this mean?
- Lines 178-179: Interesting that the hodograph is almost a straight line. How unidirectional was the shear? That's unusual for hailstorms, at least in the U.S. How do you expect this could have affected the storm morphology?

- Line 180: I don't know that I'd consider that profile to have a high level of moisture above (colder than) 0°C.
- Lines/80-185, Fig. 1b: Why is this profile chosen? How representative is it of the wider area over which hailstorms initiated?
- Lines 196-197, Fig. 2b: Interesting result. What would you say this means physically?
- Line 214: Is this hail diameter produced from just one ensemble member, or all of them? How do the hail sizes produced compare across storm lifetime, ensemble member, and to observed reports?
- Line 213, Sections 4 and 5: Is this analysis from just one storm, a several of them? Lines 218- 222 make it sound like several storms, but 225 says just one.
- Line 235: It is interesting that less intense phase correlates with a temporary increase in terrain height. I'm curious what the wind profile is doing at that time. Can you calculate a storm inflow parameter, based on the low-level winds and storm motion, to see how that is changing? Also, how does the storm structure change during this time? Is it (still) linear? How wide are the updrafts? There are plenty of places to examine for a reason why the max updraft is weaker.
- Fig. 4: I would replace pressure on the y axis of (a) and (b) with temperature, as it is more relevant to hail growth. You could also add a horizontal line where T = 0C to help guide the reader's eye.
- Line 239: 0-6km wind shear should be shown. Perhaps divide Fig. 4 into two figures. That would also give you more room to expand (a) and (b) horizontally. A colorbar with more distinction in it would help too.
- Lines 246- 249: I am wary of conducting an analysis on storms that were not observed on radar in real life. Could you at least also require the modeled storm tracks to have a similar, nearby observed radar track, and remove from the analysis those that don't? I'm also wary given the differences in maximum cell areas shown in Fig. 2b between modeled and observed. The model could have gotten the convective mode of some of the cells wrong, as it doesn't have enough small cells or very large cells compared to the radar values.
- Line 247: Are only storms that reach a W of at least 25 m/s considered, or only the timesteps where W is > 25 m/s? If the latter, how many time steps are included from each cell? Are there enough to be representative?
- Line 251, Fig. 5a: I wouldn't consider a figure with only two contours for each field (and averaged over numerous cases) to "reveal intricate cloud structures."
- Fig. 5a: Is "total hydrometeor mixing ratio" only the precipitating hydrometeors? If so, I would note that.
- Fig. 5: Are these vertical cross-sections averaged in the across-track direction at all?
- Lines 257-258: While I agree with you about the cloud ice, precipitation (assuming that's the color fill in Fig. 5b) forms much earlier and at warmer temperatures. You could say it reaches a maximum around -38°C.
- Lines 257-264: You're right that little cloud ice is being produced before -38°C, but that's not all the frozen liquid fields. I recommend reviewing the snow and graupel mixing ratio fields, which I expect will extend much lower.
- Lines 265- 266: But remember, averaging in time will act to smooth out fine-scale features, making them appear more like larger-scale subsidence. Standard deviation

would be helpful here as a start. It could also simply be that multiple storm modes (and downdraft locations) are being averaged together and cancel each other out (as explained in Lines 266-270.) I would remove the large-scale process supposition from the text unless more evidence can be provided.

- Line 273: There's a cold pool due to evaporative cooling in a convective downdraft. That’s not likely to be caused by large-scale downward motion.
- Fig 6: Without any scale information in the circles, it's hard to tell if features across subfigures are collocated. On the other hand, I understand the desire to avoid cluttering the figures. I'd add a dot representing storm track center (and thus updraft center, right?) at least. Possibly also interior x and/or y axes.
- Lines 274 - 278: It's hard to tell, but I don't think the area of surface convergence (a) and updraft core (c) overlap.
- Section 4.2: Convective morphology will be important when calculating area-mean composites. Are all the storms cellular, or some linear? If they have internal rotation, are they rotating the same way?
- Line 282-238: It would be very interesting to know how much this updraft core width varies from cell to cell.
- Line 288-289: Yes, good point, but would rephrase to "horizontal advection of hailstones to other grid columns."
- Line 305: I thought only timesteps where updraft velocity was > 25 m/s were included?
- Line 309: "From this...": this what?
- Fig. 7: Please include some grid lines within the plot so the reader can more easily estimate magnitudes. Height of the maximum vertical velocity (or rather, temperature) would also be interesting to include.
- Lines 320-323: I'd argue this result is one of the key takeaways from this figure discussion, but it isn't shown! It should be (perhaps added to (d) ?)
- Lines 323-324, Fig. 7f: It isn't clear to me what the physical meaning behind this subfigure is.
- Line 344: Why z = 5000 m?
- Section 5: This section is back to a single storm track, correct? Just one ensemble member, and if so, which?
- Line 349: What time is its highest intensity?
- Line 345: Is this criteria used to ensure the air parcel is lifted within the updraft? How effective is it, and now necessary is it? How do parcels end up in the updraft core if not through the updraft base- entrainment? It would be interesting to see if the amount of environmental entrainment into the updraft changes over the course of the mature period.
- Fig 9: Minor quibble, but I'd reverse the direction of the colorbar so 750 hPa is at the top.
- Lines 350- 354: Really interesting work. What can you say abut the vertical distribution of the inflow trajectories over time? Does that result agree with the mean values calculated in lines 319- 324? The idea that storm-relative inflow becomes broader and less coherent as a storm moves toward dissipation is an intriguing one that should be called out for future research. A comparison to previous studies of storm-relative inflow of similar convective modes would be good here too.

- Fig. 10: Now back to all cells?
  Also, are these values averaged over all times, with t= 0 when the trajectory enters the updraft? Or are they relative to just the most intense time of each cell? I'm concerned if the increasingly broadening inflow shown in Fig. 9, for example, is being averaged all together in Fig. 10.
- Line 358: I see no plot of latent heat release (it could be calculated from the model data if you want it, perhaps?)
- Lines 383-384: Good result. The rain falling into the updraft could be a result of the "coarser" (not convective-resolving) nature of a 1.1-km grid: the updraft and downdraft are not sufficiently resolved so sometimes portions of them occur in the same grid cell.
- Line 417: not "convective-resolving" (typically considered to be on the order of ~100 m, Bryan et al. 2003) but instead convective- permitting.
- Line 432: But the research results gleaned herein will still be subject to any internal biases inherent in the hail diagnostic.