Reviewer 1

1. Was there analysis (akin to Figures 4-7) on changes in the hail growth zone specifically? If not, paired with a higher updraft threshold (see comment 2), I'm wondering if this would illuminate some of the divergence seen in the hail results specifically regarding the increase in diameter. I believe it may be beneficial to at least see if changes in the convective core differ from those seen when looking at the whole of the storm.

   We zoomed in on the hail growth zone by examining how higher updraft thresholds influence hail sizes. As shown in Figure S8 and Table S2, being further and further inside the convective core does not substantially change the seeding impact on hail size; the signal remains weak and less systematic than the effects seen for ice and graupel. Since moving into the hail growth zone reduces the number of available grid points without substantially changing the seeding impact on hail size (Figure S8 and Table S2), and as 1 m/s is sufficient at our 1.1 km resolution to identify the core (LeMone and Zipser, 1980), we retain the current tracking thresholds in the manuscript and include this figure in the supplementary material.

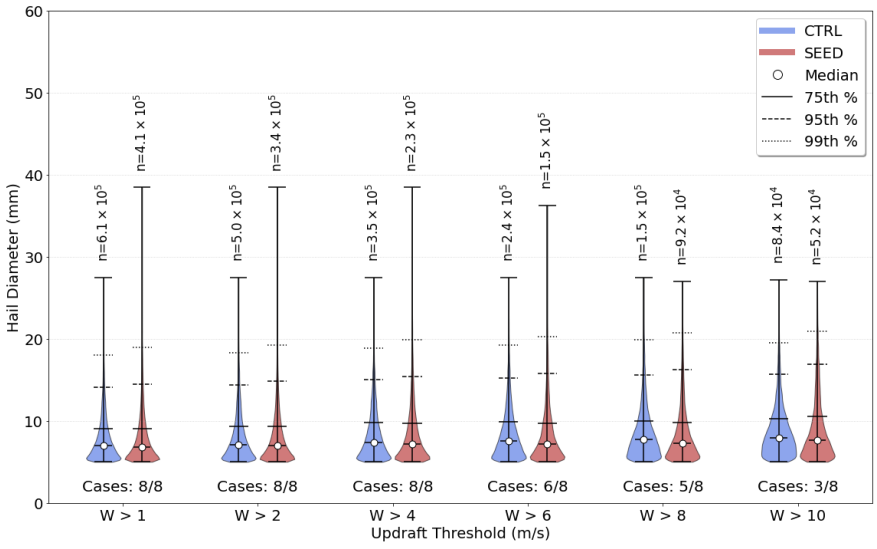

Figure S8) Violin plots comparing the distribution of hail diameters (mm) between control (CTRL, blue) and seeded (SEED, red) simulations across various updraft velocity thresholds (m/s). The outer bars of the violin plots represent the maximum value. Data are aggregated across all simulated cases, ensemble members, and grid points for the first 1.2 hours of tracking. White dots indicate the median hail size, n denotes the total number of data points for each distribution, and the labels at the bottom indicate the fraction of cases that reached each updraft threshold and produced hail.

		CTRL					SEED				
		n	Median	75th	95th	99th	n	Median	75th	95th	99th
W Threshold	Case		(mm)	(mm)	(mm)	(mm)		(mm)	(mm)	(mm)	(mm)
W > 1 m/s	01 Jul. 2020	$2.0x10^5$	7.2	8.9	12.8	16.8	$9.3x10^4$	7.1	9.4	15.8	20.4
	06 Jul. 2019	$8.1x10^4$	9.2	12.1	15.6	17.0	$5.0x10^4$	8.7	11.4	15.6	17.8
	12 Jul. 2019	$3.3x10^4$	5.9	6.9	10.1	14.5	$3.5x10^4$	6.3	7.9	13.5	18.5
	06 Aug. 2019	$4.3x10^4$	6.1	7.1	9.4	11.9	$3.9x10^4$	6.4	7.9	11.8	15.6
	17 Aug. 2020	$3.7x10^4$	6.0	7.1	9.6	12.8	$2.5x10^4$	6.1	7.7	13.7	19.8
	28 Jul. 2020	$1.4x10^5$	7.9	10.7	16.7	21.2	$9.5x10^4$	7.7	9.8	14.9	20.7
	22 Jul. 2020	$7.0x10^4$	6.1	7.1	9.9	13.4	$6.6x10^4$	6.0	7.1	10.7	15.6
	02 Jul. 2020	$8.1x10^3$	5.5	6.0	8.1	11.6	$9.1x10^3$	6.5	8.5	13.8	18.5
W > 10 m/s	01 Jul. 2020	$4.9x10^4$	8.2	10.0	14.1	18.2	$2.9x10^4$	8.2	11.6	17.7	21.7
	28 Jul. 2020	$2.8x10^4$	8.3	11.8	17.7	21.2	$1.7x10^4$	7.6	9.8	16.1	19.9
	22 Jul. 2020	$6.8x10^3$	6.3	7.2	9.3	12.2	$5.0x10^3$	5.8	6.8	10.4	14.8

Table S2) In contrast to the aggregated data shown in Figure S8, this table provides a case-by-case breakdown for updraft velocity thresholds of W > 1 m/s (baseline) and the highest one, W > 10 m/s. Cases are listed in descending order based on their CAPE. For each case and threshold, the total number of data points (n), the median hail size, and the 75th, 95th, and 99th percentiles are presented.

2. Analysis of hail was conducted using the 1 m s-1, and while Figure 12 shows that changing thresholds don't alter the main conclusion (i.e., changes to hail remain inconclusive), using a 1 m s-1 threshold doesn't seem appropriate for hail given that updrafts on the order of 10 m s-1 are generally considered the minimum needed for hail growth (e.g., Knight and Knight 2001). As such, basing the conclusions primarily on this— again even though it doesn't seem to make a difference in this study— is somewhat misleading, given that when looking at larger hail, the processes discussed take place in stronger updrafts (convective core). That said, the analysis with the 1 m s-1 threshold is still incredibly valuable, providing insights on the graupel to hail transition and tracking the result of smaller, numerous graupel on hail (and as seen in Table 1, the storms used in this study have weaker maximum updraft speeds). I think acknowledging that this updraft speed is generally insufficient to support hail growth and referencing section 3.5 where this is discussed in more detail would be fine.

Knight, C. A., &Knight, N. C. (2001). Hailstorms. Severe Convective Storms, Meteorological Monographs, 28, 223-248.

Yes, you are right. We addressed the sensitivity to updraft speed in our response to comment 1. We have now acknowledged this and cited Knight and Knight (2001).

"These thresholds play a significant role in determining which features are tracked and, consequently, influence the results. That is why, in this study, we devote a dedicated section to examining how these thresholds influence the mass of ice, graupel, and hail. In all other sections, the tracking thresholds are fixed at an AgI concentration of 150 $\mathrm{L}^{-1}$and a maximum updraft velocity in the grid column exceeding 1 $\mathrm{m\,s}^{-1}$. While literature suggests that updrafts exceeding 10 $\mathrm{m\,s}^{-1}$ are generally required for sustained hail growth (Knight and Knight, 2001), we employ the tracking threshold of 1 $\mathrm{m\,s}^{-1}$. This choice is motivated by three factors: first, sensitivity tests show that results on hail size

remain consistent regardless of the updraft threshold used (see Fig. S8). Second, given our 1.1 km horizontal resolution, a threshold of 1 $\mathrm{m\,s^{-1}}$ is sufficient to identify the core (LeMone and Zipser, 1980); third, a more restrictive threshold would drastically reduce the number of available grid points".

3. Lines 156-160 imply that differing terrain will be evaluated in addition to environmental conditions however this is not the case. Would suggest re-wording.

   Our intention here was to highlight the diversity of the selected cases, which makes our overall dataset more robust. We have re-worded these lines to clarify that terrain differences provide a representative background rather than a variable for explicit evaluation. Please see the revised text.

   "These storms occur in settings from moderately complex terrain to high-altitude regions under diverse synoptic conditions (Fig. S1). This variability provides a representative background that allows us to evaluate the robustness of seeding effects across diverse environmental and topographic contexts. Specifically, seven cases occurred over moderate topography, while one case (28 July 2020) was located over high-altitude terrain."

4. Section 3.2: the inclusion of this section comparing modeled and observed rain rate is unclear and does not seem necessary, especially given the study is evaluating changes in hailstones. I could only see this section be somewhat relevant if changes in the rainfall rate/accumulations as a result of seeding (and to see if the melting of smaller, rimed hydrometeors potentially led to increases) were also investigated.

   Section 3.2 serves to validate the model against observations, providing confidence in the subsequent seeding-impact analysis. We evaluated the impact of seeding on precipitation rate in Papaevangelou et al. (2025) and also in this manuscript in Figure S6. In order to make this comparison more relevant to this study, we added a comparison of lightning which often accompanies hail. Please see comment 7 from reviewer 2.

5. Section 3.3 & 3.4: All analysis is implied to be conducted in the bounding boxes seen in Fig. 1 (constructed using the tracking threshold of 1 m s-1 and AgI concentration of 150 L-1) however I think this needs to be emphasized better.

   We have now emphasized this in Lines 149-150.

6. Line 200: changes in updraft strength were analyzed (as seen in Fig. 4) but were changes in the updraft area (in the context of Dennis and Kumjian 2017) investigated at all?

Dennis, E. J., and M. R. Kumjian, 2017: The Impact of Vertical Wind Shear on Hail Growth in Simulated Supercells. J. Atmos. Sci., 74, 641–663, https://doi.org/10.1175/JAS-D-16-0066.1.

Yes, we also investigated the changes in the updraft area. Figure 5 and 6 (Manuscript) suggest that in our cases, seeding mainly affects updraft intensity, rather than its horizontal extent (area). We have included Figure 6 in the manuscript and added the following text:

"Figure 6 shows the updraft area distributions for the >1 m/s and >10 m/s thresholds in the CTRL and SEED simulations. For the >1 m/s threshold, seeding slightly alters the shape of the distribution while maintaining its bimodal structure. The median area is 640 km² for the CTRL and 661 km² for the SEED simulations; however, these changes are not statistically significant according to the Mann-Whitney U test (p = 0.45). This non-parametric test was selected because the pooling of the data results in unpaired distributions of unequal sizes, precluding a direct one-to-one comparison. Conversely, for the >10 m/s updraft threshold, seeding modifies the shape of the distribution by slightly increasing the peak and reducing the probability density for updraft areas below 70 km². Although the median area shifts from 101 km² (CTRL) to 96 km² (SEED), these differences are also not statistically significant (p = 0.22). Consequently, while we observe updraft invigoration in our simulations, there are no significant changes in the overall updraft area in response to seeding."

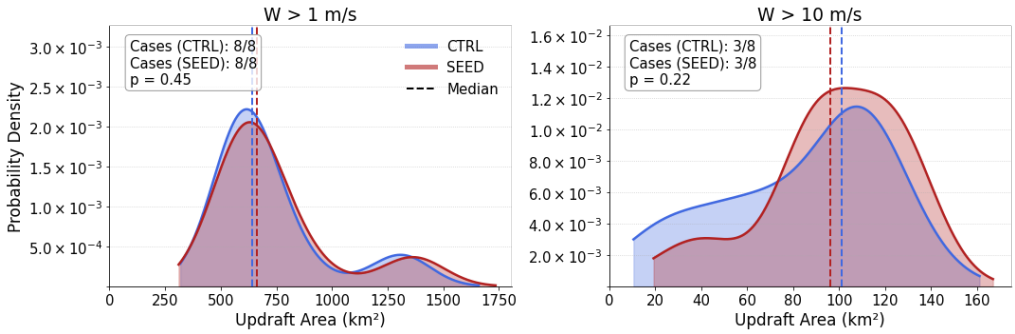

Figure 6 (Manuscript): Probability density functions (PDFs) of the updraft areas for the control (CTRL, blue) and seeded (SEED, red) simulations. These distributions aggregate convective cores across 10 ensemble members from up to 8 simulated cases, depending on the threshold. Vertical dashed lines denote the median value of each distribution. The number of cases contributing to each panel is indicated, along with the p-value derived from the Mann-Whitney U test.

7. Figure S1: a bounding box outlining the study domain (seen in Figure 1) would be beneficial to include.

   This change has been made.

8. Is there a potential that storms down-wind of those seeded could be influenced?

Investigating the downwind effects falls outside the scope of the present study. A comprehensive analysis of downwind effects would require a different tracking approach over larger spatial and temporal scales, which will be a focus of our group's future research.

Reviewer 2

1. The description of the seeding procedure is not sufficiently clear and needs substantial clarification (lines 130-138).

   There are a few ambiguous terminologies:

   >Please define "target cell" (used for maximum-updraft identification). How is it different from the seeding box? And also clarify whether the target cell and the seeding location move over time as the storm evolves.

   The target cell is defined as the convective storm identified by the maximum updraft at the cumulus stage**.** It is the convective storm cell that we seed.

   The seeding box, conversely, is a fixed-size geometric domain (5.5 X 5.5 km) centered on the target cell's updraft maximum.

   Regarding the movement: Both entities evolve over time. The target cell at this stage moves according to the atmospheric flow, and consequently, the seeding location is updated at each time step (t1, t2, t3) to ensure the seeding material is always injected into the updraft of the evolving storm. The final 'Seeding Area' shown in the schematic is the union of these boxes, covering the storm's trajectory during the 10-minute intervals (Fig.1). We have updated the desciption of seeding process and added this schematic in the manuscript.

   “In this study, we seed hail-producing convective cells during their cumulus stage, the initial phase of convective storms characterized by the presence of updrafts and the absence of downdrafts. With a model output frequency of 5 min and an AgI flare burn time of about 10 min, up to three output times occurred within a single seeding event. For each of these times, we identified the grid column with the maximum updraft in the cell we want to seed during its cumulus stage and defined a fixed 5.5 km × 5.5 km box centred on that location (Fig. 1). The seeding area was taken to include all three boxes from the different time steps (Fig. 1). We recognize that this area is likely larger than real-world flare plumes, which is a limitation of our approach. Across ensemble members, the position of the cells we seeded during the cumulus stage varied little, so the seeding area changed only slightly, with differences of no more than ±1 grid point. Seeding was applied for 10 min to five model levels immediately below and at the cloud base at a constant rate of ~0.033 cm^{-3} per

second (Papaevangelou et al. 2025). As a result of temporal accumulation within this fixed volume, the AgI concentration reached a peak value of 20 cm^{-3} at the end of the seeding interval."

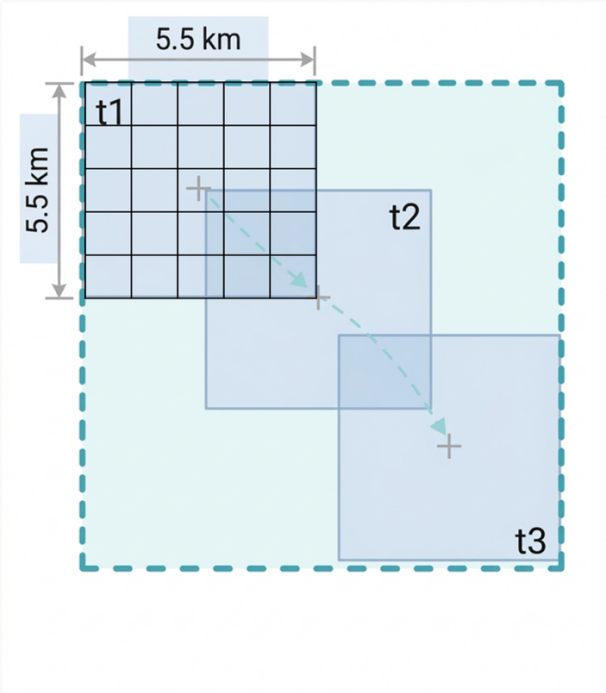

Figure 1 (manuscript): The schematic illustrates the spatiotemporal identification and treatment of a convective target cell during its cumulus stage. The process involves three discrete time intervals (t1, t2, t3) spanning a total of 10 minutes. For each interval, the maximum updraft is identified (indicated by the '+' markers), and a fixed seeding box of 5.5 X 5.5 km is centered on this location. For visualization, the t1 seeding box includes a 5 X 5 grid representing the model's resolution. The final total seeding area (outer dashed perimeter) represents the union of these individual boxes, ensuring the seeding material is consistently injected at the cloud base of the active updraft.

2. What is defined as "cumulus stage"? Accordingly, when is the actual seeding period for each storm?

   According to the US National Weather Service (NWS), the cumulus stage is defined as the initial phase of a convective storm's life cycle, characterized by the presence of updrafts and the absence of downdrafts.

   In this study, the actual seeding period is uniform across all cases and ensemble members. It takes place strictly during the cumulus stage—defined by the presence of updrafts and the absence of downdrafts—and lasts for a total of 10 minutes.

3. The temporal release process of AgI seeding is unclear and needs clarification. How is AgI released during the 10-minute window (constant rate or time-varying)? What is the release rate? The manuscript states that t"he seeding area was taken to include all three boxes from the different time steps"; does this imply discrete release at output times rather than continuous release? Correspondingly, the manuscript reports a peak AgI concentration at the end

of seeding, but additional release-process details are needed to interpret this number physically (is it due to accumulation over time or increasing release rate, etc).

Please see our answer to comment 1. The AgI seeding is performed as a continuous release at a constant rate throughout the 10-minute window.

The reported peak of 20 cm^{-3}at the end of the period is the result of temporal accumulation within this fixed volume. Given the 10-minute (600 s) duration and the starting concentration of 0 cm^{-3}, the constant release rate is calculated at 2 cm^{-3} per minute (~ 0.033 cm^{-3} per second).

4. The statement that seeding was "mirrored" from the 2025 study appears inconsistent with the current wording. In the current paper, s"eeding was applied for 10 min to five model levels immediately below and at the cloud base", while in their 2025 paper, a"t 1200 UTC, seeding is carried out across five distinct model levels underneath the cloud. ... Considering the observed reduction in cloud-base height between 1200 and 1210 UTC, seeding after 1200 UTC is conducted in model levels at or above the cloud base.

   We changed the wording and now say that the same seeding strategy as in the 2025 study was used.

5. The seeding box is a 5x5x5 grid boxes (5.5km x 5.5km horizontal with 5 vertical model levels), which is very large. Operationally, aircraft release is close to a point-source. What is the justification of using this big box instead of a single grid box? This set up may over-broaden the AgI dispersion. The manuscript acknowledges this limitation, but this could be improved (though not perfectly) by using one grid box instead. Have the authors conduct any sensitivity of the seeding impact to the box size? Given the chaotic nature of the convective storms, this treatment may substantially affect the spatial distribution of the AgI and thus the seeding signal.

   The physical reasoning behind the chosen setup is consistent with real-life operations: an aircraft circling below cloud base at approximately 200 km h^{-1} during a 10-min flare burn would traverse several grid points at the model's 1.1-km resolution. Therefore, our approach serves as a much more realistic spatial approximation of this continuous flight track than a single point source.

   Furthermore, given that seeding is conducted at the cloud base, the seeding material relies entirely on the presence of an updraft to be transported aloft; otherwise, it fails to reach the ice nucleation region. Consequently, only a fraction of the released AgI that is actively captured by the updraft successfully reaches this critical altitude. This is consistent with the findings of Miller et al. (2025), who reported that under most seeding conditions, the AgI activation

ratio ranges from only 0.1% to 1%. This natural depletion process ensures that the amount of seeding material transported aloft is comparable to what an aircraft would successfully release directly into a convective storm's updraft during real-world operations.

Therefore, conducting sensitivity tests for smaller box sizes (such as a single grid point) was not appropriate for this study. However, exploring finer model resolutions and smaller seeding areas remains a topic for future work.

We have updated also the text in the manucript:

"The campaign adopted the updraft seeding method—first introduced in 1948 and now widely used in operational weather modification efforts worldwide—targeting convective clouds during their cumulus stage and seeding from an aircraft that circles below the cloud base."

6. Actual time-window of the simulation and seeding start/end time are missing. The paper stated all simulations start at 0 UTC to cover early day storms, but no information of the actual storm time was given: how many early day storms are selected? What's the simulation length for each storm: is it same for all storms or ending differently based on the actual storm time? What are the identified start/end seeding time for each storm?

To clarify, all simulations were initialized at 00:00 UTC and ran for 18 hours across all eight selected cases.

We added a table with the start/end times of the storms, as well as the seeding periods in the Supplementary Material (Table S1).

Case	Start	End
01 jul 2020	15:35	17:10
06 Jul 2019	12:45	13:50
12 Jul 2019	10:50	12:55
06 Aug 2019	09:55	13:00
17 Aug 2020	13:45	15:30
28 Jul 2020	13:45	14:50
22 Jul 2020	14:45	17:30
02 Jul 2020	11:20	13:50

Table S1) The seeding period and the start and end times for the 8 simulated storm cases (in UTC). Seeding was applied for 10 minutes beginning at the storm start time.

7. Figure 2 does not show good model-observation agreement for precipitation evolution, what can it tell us, to what degree can we trust the results if it does

not represent the precipitation trend and magnitude in reality? How would that potentially affect the interpretation of the simulated hail intensity and therefore seeding impacts? This type of discussion would be critical before jumping into the simulated seeding impact interpretation.

Our model underestimates precipitation, and this could indeed potentially affect hail intensity if, for example, the reduced precipitation stems from weaker storm dynamics, with weaker updrafts and altered microphysics. However, the Lightning Potential Index (LPI) evaluates the potential for charge separation by combining updraft speeds with cloud hydrometeor data (graupel, ice crystals, snow, supercooled droplets, and raindrops). As demonstrated in Figure 3, the simulated LPI aligns with the LINET observations, specifically in that both occur within the exact same time period and spatial domain. This agreement gives us confidence that the atmospheric environment we seed in our model is comparable to the observed storm. Consequently, we have now added Figure 3 and the accompanying text in the manuscript.

“In this section, we compare the modeled precipitation with the CombiPrecip dataset from MeteoSwiss (2017) and the Lightning Potential Index (LPI) (e.g., Brülisauer et al., 2025) with the low-frequency (VLF/LF) international lightning detection network LINET (e.g., Nag et al., 2015). CombiPrecip provides hourly precipitation fields derived from a geostatistical combination of rain-gauge measurements and radar estimates. The dataset extends 100–150 km beyond the Swiss border, covering neighboring countries, which makes it well-suited for our analysis. Another advantage is its spatial resolution of 1 km, which matches that of our model output. As shown in Figure 3 (left panel), despite the limited number of data points and considerable spread, it is evident that COSMO tends to underestimate precipitation. This finding is consistent with Shrestha et al. (2022), who used a similar model setup. Furthermore, Figure 3 (right panel) compares the modeled LPI with the LINET dataset. The LPI evaluates the potential for charge separation by combining updraft speeds with cloud hydrometeor data (graupel, ice crystals, snow, supercooled droplets, and raindrops), whereas LINET monitors lightning activity across most of Europe, including Switzerland and Germany. Notably, the 5-minute temporal resolution of the LINET data matches the temporal resolution of our model output. Although the two datasets utilize different units, this comparison demonstrates that the observed lightning activity occurred within the same domain and during the same time periods as the simulated convective storms, albeit with spatial spread. Thus, the simulated storms and their broader environment are broadly comparable to the observed storms, despite the underestimation of precipitation.”

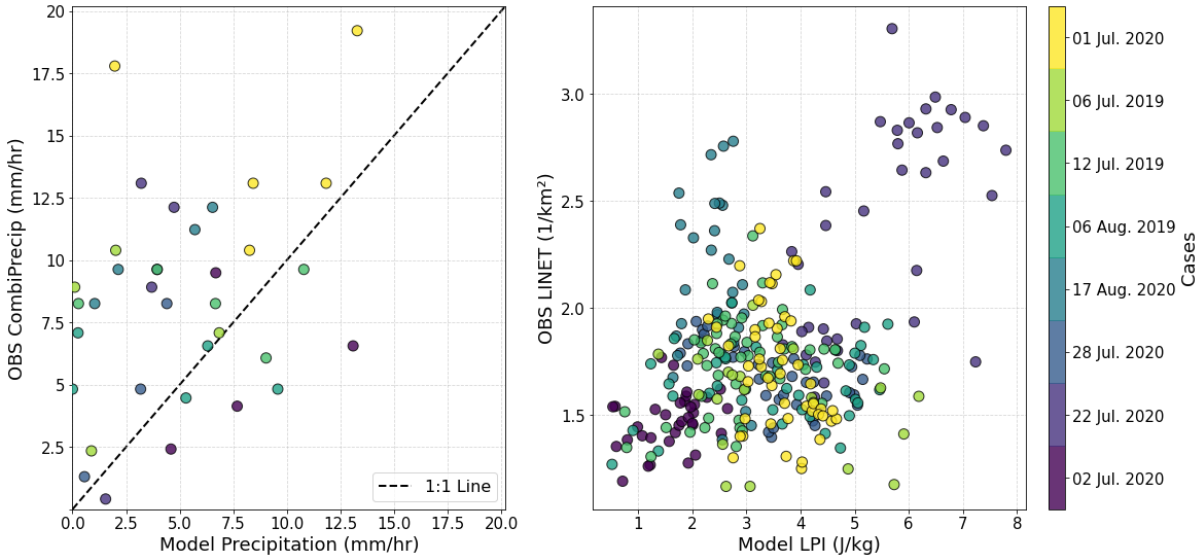

Figure 3 (manuscript): Scatterplots illustrating the comparison of model output with observational data for the studied cases. (Left): Comparison of model precipitation (mm/hr) with the CombiPrecip dataset (mm/hr). (Right): Model-simulated LPI (J/kg) vs. LINET lightning density (km^{-2}). The analysis for each case is performed over a domain centered on the simulated convective storm, extending 1° west/east and 1° south/north from the storm center. Cases are color-coded according to their CAPE values (ranked from highest: yellow to lowest: dark purple).

8. In Figure 7, most hail does not reach the surface in CTRL (very low number concentration << 1/m^3), although hail was observed in these events. So are these simulated number physically realistic and expected for the selected cases? In other words, is there a typical number concentration or threshold to define a hail event? Since the purpose is to suppress hail, it makes sense to realistically simulate hail reaching the surface. I would also suggest to use log scale in the x-axis to better illustrate the magnitude since hail at the surface is the most important metrics.

In our simulations, a hail event is defined by a concentration threshold higher than $10^{-3}\mathrm{m}^{-3}$, a threshold selected based on Auer (1972), as stated in line 109. We fully agree with the reviewer's observation that relatively few hail particles reach the surface in the CTRL simulations. This underestimation is primarily due to the specific microphysics scheme's known tendency to overestimate hail melting as hydrometeors fall below the melting layer (U. Blahak, personal communication, 2021), which remains a persistent challenge in convective modeling. Specifically, Rasmussen and Heymsfield (1987) showed that below the freezing level, meltwater is gradually shed from the hailstone to maintain equilibrium as its ice core shrinks. In contrast, microphysics schemes typically assume instantaneous meltwater shedding. This assumption causes the simulated hailstone to lose mass, increasing its atmospheric residence time in the warm layer and resulting in an overestimation of the total melting.

While excessive melting artificially reduces absolute hail size at the surface, the seeding induced freezing occurs well above the melting layer. Furthermore, since both CTRL and SEED simulations utilize the same microphysics parameterization, they are subject to identical melting rates. Therefore, by focusing on the differences between the CTRL and SEED simulations, we reduce this low-level melting bias.

Following your suggestion, we have updated Figure 7 (now Figure 9) to use a logarithmic scale on the x-axis to better illustrate these variations.

9. Figure 8 caption requires clearer metric definition:

Does values reflect ensemble mean of the domain-time average for each case? Given the nature of convective storm, that would average out too many physically meaningful structures and stage-dependent signals. Domain average will smear out the local signals, calculate the values based on sub regions (such as your target cells) will help identify seeding signals reponding to local CAPE/wind shear.

In this figure, we did not use a domain average. Instead, we applied a spatial mask targeting only the tracked convective cells to accurately capture localized storm signals. This is also mentioned now in the caption of this figure.

Averaging over the entire simulated period mask out the different cloud seeding response to storm at different stage (cumulus, development stage, mature stage, as well as dissipating stage), and for hail formation, usually the mature stage will be the focus.

We apologize for the imprecise wording in the caption of our manuscript. In fact, our analysis specifically isolates the mature stage through a two-step process: first, by tracking storms based on updraft presence (which inherently excludes inactive or dissipating periods), and second, by extracting the 99th percentile of the spatial mean hail size over the storm's active lifespan. This approach allows us to capture the storm's maximum hail production footprint. We have now revised the figure caption to accurately describe this methodology.

“Scatter plots illustrating the impact of AgI cloud seeding on hail sizes for cases categorized by CAPE (J/kg) and wind shear (m/s). Left panel (CTRL): Surface hail size. To focus on the mature, heavily precipitating stage, we use the 99th percentile over time of the spatial mean. This spatial average is calculated over the tracked convective storms. Plotted values are ensemble means. Right panel (SEED - CTRL): Change in hail size between seeded and control simulations. Marker sizes in both panels correspond to the standard deviation across the ensemble members.”

Ensemble mean is not a great matrix to look for relationship, given members exhibit a large spread. Please consider showing all members. Also an average over the entire domain would smear out the local signal. This also gives you more sample points with physically meaningful representations.

We generated these plots and added them to the supplementary material. We retained the ensemble mean and scatter plots in the main manuscript. While Figure S10 provides comprehensive information on the full ensemble spread, the ensemble mean is not driven by isolated outliers and accurately represents the physical response—with 6 to 10 out of 10 members consistently agreeing with its sign.

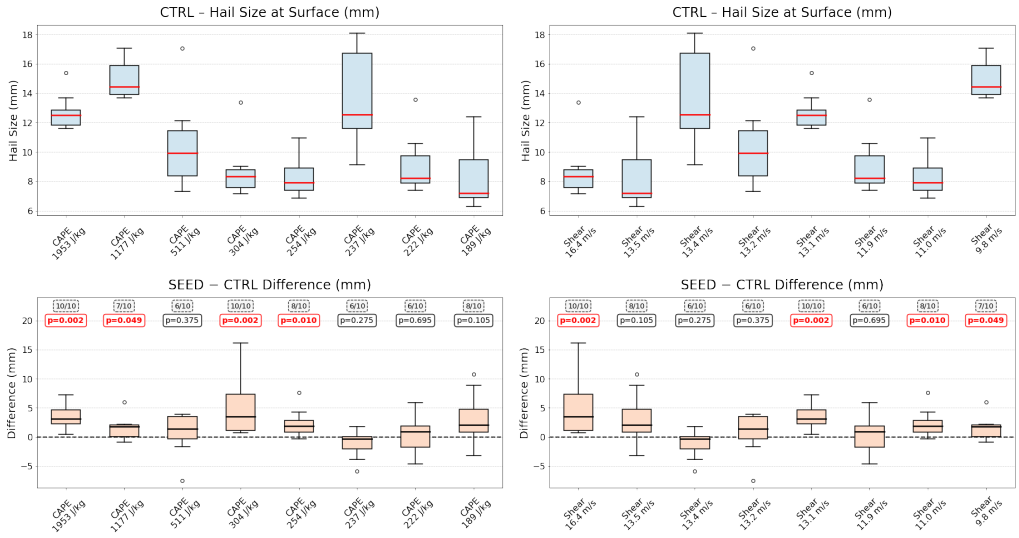

Figure S10) As in Figure 9 of the main manuscript, but displaying boxplots to show the full ensemble distribution (10 members per case) of surface hail size across all cases. Cases are sorted from left to right in descending order of environmental conditions: CAPE (left panels) and wind shear (right panels). Top panels (CTRL) show absolute surface hail sizes for unseeded simulations, denoted by blue boxes. Bottom panels (SEED - CTRL) illustrate the difference in surface hail size between seeded and unseeded simulations. The boxplots define the interquartile range (IQR), the horizontal line denotes the median, and the whiskers extend to the minimum and maximum values for each case's ensemble members. Additionally, p-values from the Wilcoxon signed-rank test are presented to compare each SEED case with its corresponding CTRL simulation. Finally, dashed boxes indicate the number of members (e.g., 8/10) that agree with the sign of the ensemble mean difference.

10. Figure 10: Please clarify how time-average values are computed? Did you first compute the time-average and then apply thresholds, and then average over these grid points, or first apply thresholds at each time, and then then take time and grid average? These can produce different results as vertical wind and AgI number vary over time and location. Correspondingly, Figure S3, please clarify how AgI thresholds are identified for CTRL since it does not have AgI? If locations are identified in SEED but values are sampled in CTRL, what is the physical meaning of these classification? (showing SEED results will have physical meaning but CTRL value based on the ghost AgI concentrations does not make physical meaning)

To clarify, we first applied the thresholds at each specific time step and grid point to create a spatio-temporal mask, and then we computed the average

over these selected points. By applying the thresholds first, we ensure that we only extract and average the values exactly where and when the updraft and seeding concentration criteria are simultaneously met.

Regarding the CTRL simulations and the physical meaning of this approach: Similar to Xue et al. (2013), we use the AgI plume from the SEED runs as a spatio-temporal mask to define the cloud volume affected by seeding. By applying this identical mask to the CTRL run, we examine the corresponding volume of air in its natural, unseeded state.

We acknowledge that due to the dynamical response of the convective clouds to the seeding process, the spatial locations of the storms may diverge slightly between these two simulations over time. However, by focusing only on the first 1.2-hour period of our study—after which two of our cases are no longer tracked by the tracking algorithm—we aim to minimize this spatial mismatch caused by dynamical responses to seeding.

11. Current analysis emphasize mostly on mean/medians/total. The manuscript would benefit from more from showing some spatial analysis, as a core strength of modeling over observations is resolved spatial structure. For example, for supporting key claims such as seeding reduces hail-impacted area, it would be helpful to show the hail impacted region (e.g., accumulated hail mass/number on ground) overlaid with terrain map before and after the seeding. This straight-forward comparison could also help understand how terrains play a role over the Alpine region.

    We have created the requested spatial map showing the hail-impacted area overlaid on the terrain. Since the main manuscript focuses on the statistical

evaluation across all cases, we have included this figure in the Supplementary Material (Fig. S11) to maintain the paper's flow.

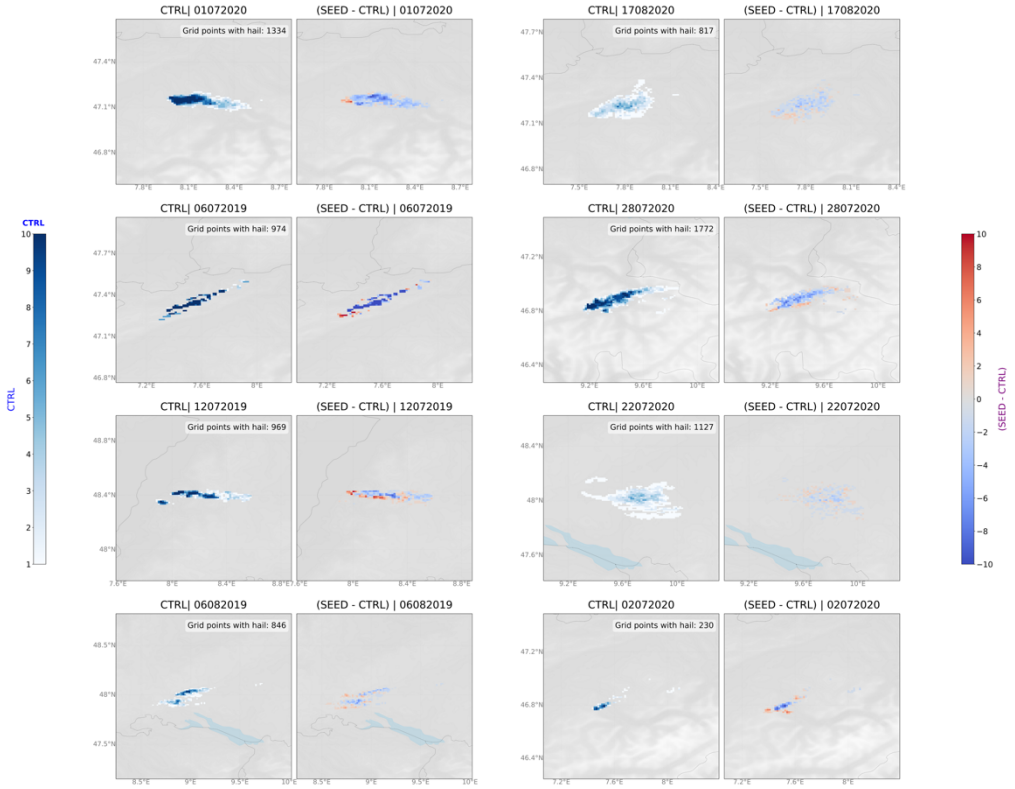

Figure S11) Spatial distribution of surface hail occurrence of the studied cases sorted by CAPE. The left column (blue) presents the CTRL, while the right column displays the difference (SEED - CTRL) for every case. The color intensity represents the number of ensemble members (out of 10) that simulated maximum surface hail exceeding the 5.0 mm threshold over the 1.2-hour analysis period. The total count of grid points satisfying this threshold for the CTRL is annotated in the upper right corner of the left panels. Background shading indicates the local topography. The lake depicted is Lake Constance.

References:

Auer, A., 1972: Distribution of Graupel and Hail With Size. Monthly Weather Review, 100, 325–328.

Brülisauer, M., N. Papaevangelou, and U. Lohmann, 2025: Simulations of selective seeding on hailstorms in northern Switzerland using the COSMO model: Effects on the Lightning

Potential Index. J. Appl. Meteor. Climatol., 1293–1305, https://doi.org/10.1175/JAMC-D-24-0129.1.

Knight, C. A., and N. C. Knight, 2001: Hailstorms. Severe Convective Storms, Meteor. Monogr., 28, 223–248.

LeMone, M. A., and E. J. Zipser, 1980: Cumulonimbus vertical velocity events in GATE. Part I: Diameter, intensity and mass flux. J. Atmos. Sci., 37, 2444–2457, https://doi.org/10.1175/1520-0469(1980)037<2444:CVVEIG>2.0.CO;2.

MeteoSwiss, 2017: Documentation of MeteoSwiss grid-data products: Hourly precipitation estimation through rain-gauge and radar: CombiPrecip. Federal Office of Meteorology and Climatology MeteoSwiss.

Miller, A. J., Fuchs, C., Ramelli, F., Zhang, H., Omanovic, N., Spirig, R., Marcolli, C., Kanji, Z. A., Lohmann, U., & Henneberger, J. (2025). Quantified ice-nucleating ability of AgI-containing seeding particles in natural clouds. Atmospheric Chemistry and Physics, 25, 5387–5407. https://doi.org/10.5194/acp-25-5387-2025

Nag, A., M. J. Murphy, W. Schulz, and K. L. Cummins, 2015: Lightning locating systems: Insights on characteristics and validation techniques. Earth Space Sci., 2, 65–93, https://doi.org/10.1002/2014EA000051.

Papaevangelou, N., D. Villanueva, G. K. Eirund, J. Chen, Z. Dedekind, and U. Lohmann, 2025: Simulations of selective seeding of hailstorms—A summertime case study over Switzerland. J. Appl. Meteor. Climatol., 1509–1524, https://doi.org/10.1175/JAMC-D-24-0121.1.

Rasmussen, R. M., and A. J. Heymsfield, 1987: Melting and shedding of graupel and hail. Part I: Model physics. J. Atmos. Sci., 44, 2754–2763, https://doi.org/10.1175/1520-0469(1987)044<2754:MASOGA>2.0.CO;2.

Shrestha, P., S. Trömel, R. Evaristo, and C. Simmer, 2022: Evaluation of modelled summertime convective storms using polarimetric radar observations. Atmos. Chem. Phys., 22, 7593–7618, https://doi.org/10.5194/acp-22-7593-2022.

Xue, L., A. Hashimoto, M. Murakami, R. Rasmussen, S. A. Tessendorf, D. Breed, S. Parkinson, P. Holbrook, and D. Blestrud, 2013: Implementation of a silver iodide cloud-seeding parameterization in WRF. Part I: Model description and idealized 2D sensitivity tests. J. Appl. Meteor. Climatol., 52, 1433–1457, https://doi.org/10.1175/JAMC-D-12-0148.1.