Authors' comments to Anonymous Referee #1

We would like to thank the reviewer for the thorough review of our manuscript and insightful feedback. These comments have significantly improved the quality of our work. In the following sections, we present the reviewer's comments (in black), our responses (in red), and the changes made in the revised manuscript (in blue). Please note that all line numbers in our responses correspond to those in the revised manuscript.

Overall comments:

I have thoroughly read the article titled "Inferring the controlling factors of ice aggregation from targeted cloud seeding experiments" by Zhang et al. Overall, I find the article very interesting and of high scientific quality although slightly limited due to experimental range. The article reports on the quantification of ice particle aggregation rate in cloud with several controlling factors in mind. They use machine learning to identify the main controlling factors. Although the study is very novel and the results are of high interest to the scientific community, there is an insurmountable modest dissatisfaction because the experimental space is so limited (the temperature range is limited and liquid droplets are necessarily present in high concentrations). These limitations only moderately take away from the novelty of the experimental design and the quality of the analysis. Overall, I feel that the article is definitely suitable for publication in ACP with some comments below. My main concern is about the utility of the results. While it is spectacular that the results have been so well studied, the authors should consider more closely where the results will be used and in what form it is best for those who will use them. Aggregation rates are used in high resolution models. The final results presented, while they may be the most accurate based on the data, produce a moderately rough line. I feel the article would be far more useful if the authors were to present suggestions as to how the results could be used in a model.

Overall criteria:

Scientific significance (4 Excellent): The publication tackles a very significant unknown through experiments. Aggregation rate of ice crystals in a cloud is a very difficult, yet important, number to know. This publication presents the results of the aggregation rate estimate after an extensive field experiment which attempts to accurately measure the range of possible values in the real world. The experimental design has been well documented. The analysis of the results (the heart of this publication) stand up in quality to the experimental design. The reason I don't give it a 5 is that the significance would be higher if they had been able to calculate aggregation rate in a wider range of temperatures which was likely limited by experimental constraints.

Scientific quality (4 Excellent): Overall, the scientific quality of the research is excellent. The authors use the latest in Machine Learning techniques to get at the analysis. See specific comments regarding areas where the scientific quality could be improved.

Presentation quality (4 Excellent): The manuscript is, for the most part, concise and the figures are well presented and easy to interpret. English grammar is perfect. The authors have done a good job moving some details into Appendices which makes the article read well.

Specific comments:

1. Line 10-11: In many microphysics parameterizations, an uncertainty of 0.08 might be considered "close enough" to 1 to be within the margin of error. I would suggest presenting a level of confidence here to help the reader understand that you are very confident that it is not 1.0 and random experimental factors drove your mean to be below 1.

   Thanks for this suggestion. We now report the uncertainty of the ICNC exponent: across the 21 experiments, the mean exponent is 0.92 (standard error 0.024), yielding a 95% confidence interval of 0.876–0.968, which does not include 1.0. This indicates that the sub-quadratic scaling is statistically significant rather than an artefact of experimental variability.

   L10-11:

   We report, however, a subquadratic dependence of the aggregation rate on $ICNC_{t0}$ (mean exponent ~0.92; 95% CI: 0.88–0.97), in contrast to theoretical expectations (quadratic dependence).

2. Line 19: Don't forget that ice crystals do grow from vapor in regions that do not include supercooled liquid drops such as in cirrus at -50C where WBF can't exist.

We agree that vapor depositional growth also occurs in fully glaciated clouds where the WBF mechanism is not active, such as cirrus at very cold temperatures and have the revised the manuscript as follows:

L18-20:

During the early stage of ice growth, diffusional processes such as the Wegener–Bergeron–Findeisen mechanism dominate (Korolev, 2007) and vapor deposition in fully glaciated clouds such as cirrus (Gierens et al., 2003).

3. Line 100: Since EDR wasn't an important factor, perhaps it might be easier to reduce the text here and just say that EDR wasn't an important factor rather than giving the details on how EDR was measured.

   Thank you for the suggestion. We have simplified the description of the EDR retrieval by retaining only the methodological essentials and removing detailed instrument-specific parameters.

   L103-105:

   Primary estimates were obtained from the Mira35 MBR7 Ka-band radar. In addition, the RPG94 W-band radar was used to supplement data gaps. The retrieval was temporally averaged over 30 s, and the resulting EDR fields have a spatial resolution of approximately 30 m.

4. Line 124: smaller crystals rather than smaller ones.

   Fixed, thanks!

   L122-123:

   The uncertainty in cloud droplet number concentration is approximately ±5 %, while that for ice crystal number concentration ranges from 5–10 % for crystals larger than 100 µm and about 15 % for smaller ice crystals.

5. Line 146: Please use terms such as "variability" and "confidence" when possible rather than "uncertainty" (which is interpreted by the general public as "we don't really know").

   We appreciate the reviewer's suggestion regarding the use of "uncertainty" in contexts where the term may be misinterpreted. In the specific paragraph noted, however, we are referring to potential detection errors of IceDetectNet rather than statistical variability or confidence levels. In this technical context, "uncertainties" denotes possible misidentification of monomers and is the standard terminology for model- or algorithm-related error sources. To avoid confusion, we revised the text to use more precise terms.

L149-150:

Other potential detection errors and classification errors are discussed in Appendix A and are considered negligible.

6. Line 172: While you eliminate EDR as being too important, you might comment on measured EDR versus what would be expected in deep convection. (The same for other parameters, knowing how your data fit into the zoo of cloud types will help the reader to understand how representative your results are of situations of interest to the reader)

   Thank you for the suggestion. We have revised the manuscript to place the observed EDR values primarily in the context of other boundary-layer cloud environments, which are dynamically more relevant to the clouds sampled in this study. For reference, we also note that EDR values reported for deep convective environments are typically one to two orders of magnitude higher (typically 10^{-2}–10^{-1} $m^2 s^{-3}$; Barber et al. 2019), highlighting that the turbulence levels encountered here are comparatively weak.

   L188-191:

   EDR values measured during the seeding flights were modest (on the order of 10^{-4}–10^{-3} $m^2 s^{-3}$, Fig. B1), comparable to turbulence levels reported for other boundary-layer cloud environments (Chu et al., 2025). This indicates that aggregation developed under a weak-turbulence conditions characteristic of stratiform mixed-phase clouds.

7. Figure 2: Are there symbols missing from the top? There is a blue line, then it gives the Cold temperature range twice, then a red line and warm twice. I suspect there should be a symbol there but it could be my computer.

   Fixed, thank you!

8. Line 240 and paragraph below: Since there is a significant habit transition between warm and cold, it might be interesting to see if there are enough data points in each of the groups to identify a trend in each. There might be a natural functional change due to habit that could be hidden by linear analysis.

   Thanks. In response, we quantified the number of observations in each temperature–habit group. The colder regime ($T \leq -7^{\circ}\mathrm{C}$) contains 797 data points from 7 experiments, and the warmer regime ($T > -7^{\circ}\mathrm{C}$) contains 2508 data points from 14 experiments. Both groups therefore have sufficient sampling to examine trends within each habit regime. We now report these sample sizes explicitly in the revised manuscript.

L169-170:

Based on this habit difference, the experiments were classified into “warmer” (T > −7 ◦C; 14 experiments, 2508 data points) and “colder” (T ≤ −7 ◦C; 7 experiments, 797 data points) regimes (Fig. 2a).

9. Line 256: RR can influence geometry and stickiness if there is still some quasi liquid present?

Thanks. Indeed, riming can potentially modify not only ice crystal geometry but also the stickiness of colliding. We have added this clarification to the revised text.

L268-270:

Because riming influences ice crystal geometry — particularly major size and aspect ratio — and it may also influence surface stickiness, especially shortly after the riming event, it can, in principle, affect aggregation indirectly.

10. Line 275: I am not sure that I see how RR decreases ICNC_t0.

Thanks for pointing this out. We agree that riming cannot physically reduce $ICNC_{t0}$, and the RR → $ICNC_{t0}$ link in the preliminary DAG does not represent a causal mechanism. The weak negative coefficient arises from a statistical dependence: cases with higher ICNC exhibit lower riming efficiency per crystal because the available supercooled liquid water is shared among more particles. This resource-competition effect induces a negative correlation even though the causal direction is the opposite, and the link is not relevant to the aggregation-rate analysis. We therefore removed the RR → $ICNC_{t0}$ arrow from the final figure and clarified this point in the text.

L290-295:

RR has a negligible direct effect on aggregation (−0.01). RR indirectly increases the major axis length (+0.31), which tends to promote aggregation; however, the positive contribution of major size to the aggregation rate is itself very small (+0.08).

RR has a negligible direct effect on aggregation (−0.01). RR indirectly increases the major axis length (+0.31), which tends to promote aggregation; however, the positive contribution of major size to the aggregation rate (+0.08) is itself very small. Moreover, most rimed ice crystals in our measurements exhibit only light riming (Fig. E1), limiting the potential for riming-induced enhancements in fall speed or sticking efficiency. Under these specific experimental conditions, the overall influence of RR on aggregation appears minimal, suggesting that any coupling between riming and aggregation is weak or not detectable within these experiments.

11. Line 300: It would be nice to have a reference for the SHAP calculations.
Thanks. Reference added.
L316-317:
Model interpretability was assessed using SHapley Additive exPlanations (SHAP; Lundberg and Lee 2017).

12. Figure 10: The obvious question is what happened at -5.3 degrees? The CatB values all dip there especially the 10^1L^-1 line, the other CatB lines are impacted there as well. As stated in the initial comments, how is a modeler supposed to incorporate these data into a weather forecasting model?
Thanks. The feature near −5.3 °C reflects the higher intrinsic variability of aggregation rates at low ICNC rather than data sparsity. This has now been quantified and clarified in the revised manuscript using the coefficient of variation.
L428-439:
This behavior reflects the high variability of observations in the low-ICNC regime. We quantify this variability using the coefficient of variation (CV), defined as CV = σ/μ where σ and μ are are the standard deviation and mean of Ragg within a given bin. Across the observed temperature range, CV values at low ICNC are typically 1.2--2.1, compared to 0.7--1.0 at intermediate ICNC and $\lesssim$ 0.6 at high ICNC. Such large CV values indicate that aggregation rates in this regime are dominated by strong fluctuations rather than a clear temperature signal. In combination with the imposed lower bound on the target variable during training, this lack of a resolvable signal causes the tree-based model to revert to the lower-bound baseline, resulting in an approximately constant prediction at low temperatures. The flat behavior therefore reflects the absence of a learnable temperature dependence in this regime, rather than a physically meaningful relationship. The fact that a localized dip near -5.3°C appears consistently across

all ICNC regimes, while being most pronounced at low ICNC and progressively weaker at higher ICNC, further supports this interpretation. This feature is more likely to reflect experiment-to-experiment variability and condition mixing near this temperature, together with the non-smooth, threshold-based nature of tree models, rather than a robust, generalizable physical transition.

L443-446:

From a modeling perspective, this implies that physically based parameterizations provide a more reliable baseline representation of aggregation under such conditions, with observational or machine-learning-based approaches best suited for diagnosing variability or informing uncertainty estimates rather than defining deterministic temperature dependence.

13. Lines 491-526: I think it would be reasonable to add the Geron, 2022 reference on line 491 and not have it on every (but one) of the model algorithms.

    Thank you for this suggestion. We agree that citing Géron (2022) repeatedly for individual algorithms is unnecessary. We now only cite Géron (2022) once at the beginning of the section.

    L522-523:

    We used a total of 10 supervised regression algorithms. A general introduction to these machine learning methods can be found in Géron (2022). Below we summarize the basic principles of each model:

Reference:

Barber K A, Deierling W, Mullendore G, et al. Properties of convectively induced turbulence over developing oceanic convection[J]. Monthly Weather Review, 2019, 147(9): 3429-3444.

Lundberg, S. M. and Lee, S.-I.: A unified approach to interpreting model predictions, Advances in neural information processing systems, 30, 2017.