egusphere-2023-290
Recommendation: Accept pending major revisions

This article provides some of the first observations of insoluble particle concentrations within hailstones collected at the ground, as well as the opportunity for embryo analysis. My primary concern is that while this observational dataset is unique and worthwhile, the authors do not use it to draw any conclusions about physical processes related to hailgrowth. It isn't clear what the purpose of these observations is. The description of the classification and clustering technique is also confusing. Once these issues are addressed I am comfortable with publication.

Major Comments:

1. The authors have clearly spent a lot of effort to gather these observations, and they are some of the first studies of insoluble particles in hailstones. However, I would like to see some conclusion about the physical processes that these observations can now help us to understand. What, physically, can we now conclude about hailstone growth, or insoluble particles in hailstones, that we couldn't before? As it stands, the article is simply a reporting of what the authors found. What do these observations *mean*?

What information does the knowledge about how the particle distributions change among embryos/shells provide about physical processes involved in hailstone growth? For example:

- Which hailstones had size distributions that changed significantly between one shell to the next? Why might this occur?
- Do shell B and embryo distributions usually look the same or different? Shell C and B? Do size distribution/particle type vary more among shells, or among hailstones? Why?
- Do you have a sense if these hailstones might have taken similar paths through the storms that generated them?
- Were the storms that generated these hailstones all of similar convective mode, or different? Would you expect storms with stronger updrafts to transport more particles, changing the insoluble particle size distribution? Why or why not?

2. The description of the classification and clustering techniques in Section 2.2 is very difficult to understand, making the subsequent results in the rest of the paper unnecessarily hard to follow.

- Fig 2: This figure is hard to follow. I'd label each step with a single action (e.g., polishing, slicing, shell extraction, etc.), and then list and describe each action by name in the caption (or in the text itself) Numbering the steps and boxes would also help. That way it is clear which stage of the process corresponds to which box in the figure. Names of equipment can go below each image, in smaller font perhaps.
- Lines 126- 127: This step should be included as a step in Fig. 3, showing what M1 - M100 are and how they are generated.
- Lines 128- 137: The SOM description could be clearer. What is being produced by this process? Identification of particle species? How is that determined from a given centroid matrix? Are the number of outcome clusters predefined (a perhaps that is k?) If so, how are the numbers k-2 through 10 selected?
- Lines 133-134: Does the "same neuronal network setting" mean k is the same, or k, i, and j are all the same? For that matter, what does "neuronal network setting" mean?
- What are "particle sample replicates"?
- Lines 134-136: What information does each of these indices provide? Most importantly, how do you determine accuracy with this method?

- Lines 144- 149, Fig. 5: Unfortunately, I can't follow this description at all. What information does a centroid matrix have in it? Are carbon and oxygen included in the classification matrix, and as SOM inputs, or not? Lines 118- 121 seem to indicate no.
- Lines 144 -159: Please explicitly state what is being used as the "truth" dataset for the random forest. I had thought it would be the centroid matrix (line 144), but instead I think it is the classification outcome of the centroid matrix (Fig. 6)?

3. The number of studies of observed hailstone embryos in the literature less than 30 years old is almost zero, so it seems a real missed opportunity not to offer some details about the embryos collected here. These were all graupel embryos, not frozen drops? How do the characteristics of these embryos and hailstorms correspond to the hailstone embryo research of Knight (1981, J. Appl. Meteorology and Climatology)? How big were each of the embryos? Are you able to estimate their density? Differently sized embryos would also have to have impacts on their insoluble particle makeup, I would think. Do you find that to be the case?

Minor Comments:

- Lines 170- 173: I think I get your meaning here, but it should be clearer. Do you mean that because you assume that the random subsample of the filter is representative of the entire filter, Ncount is determined by multiplying the observed Nfilter by the area ratio between the whole filter and the obsered image (Simages/Sfilter)? If so, I would explain it like that.
- Lines 174-177: Move these sentences to the start of the subsection immediately following (1). Also, adding a sentence after each equation explaining the physical meaning of it would be helpful In the reader. E.g., The number of insolvable particles in the melted shell (Nliquid) can be found by multiplying their number concentration (nliquid) by the volume of the melted shell (Vliquid); this total particle number does not change when the solution is diluted (Ndilute).
- Line 180: How is Ncount determined? I thought only Nfilter could be observed.
- Lines 185- 188: How are these equations determined?
- Line 196: What data? Particle number concentrations (nliquid)? Binned by particle diameter size?
- Lines 200 -203: Why was a log-normal distribution chosen? What are rg and Tg? Why this form of the distribution? What does line 202-203 mean physically?
- Lines 216- 225, Fig. 7: Given the log scale and the small y-axes of Fig. 7, it is difficult to see the differences between any hailstones, let alone among specific storms. I recommend shifting the standard deviation results to a new figure. I would also be curious to see what the standard deviation values are across all storms but excluding the GY1 and GY2 hailstones. I am curious how much of the increased standard deviation for all 7 hailstorms in sum is due to those 2 storms. Once those two storms are removed, that should have an impact on the conclusions in Lines 216- 225. Possibly, using just one Beijing hailstone is not representative.
- Lines 237 - 238: That's a lot of “possiblies”. One could just as easily argue the insoluble particles were contributed by the riming supercooled water acting to form the embryo.
- Lines 248-251: I’m having trouble following these sentences. Why would industrial coal burning result in an increased number of organic aerosols specifically 10 microns in diameter?

- Lines 252- 253: Rephrase to make it clear "at the same diameter" refers to particle diameter, not hailstone embryo diameter.
- Line 256: Would change "since" to "and", as the following phrase agrees with your previous phrase, but does not offer a possible causal factor.
- Lines 2.60-261: Do these uncertainties mean it is possible potential biological aerosol particles were misclassified in your study? If not, what is the reasoning for including this statement here?
- Line 288: Shouldn't the geometric mean diameter be D_g or d_g, not r_g? r could be too easily confused with radius.
- Fig. 9 is much too small to make out necessary detail. While I appreciate the authors' conscientiousness in ensuring the reader is aware of the geographical locations where the hailstone samples were sourced, I think the responsibility rests with the reader at this point and the maps are no longer necessary. I would split this figure into 2-3 figures to allow points to be made about each individually, as there is a lot of information here. Plus, more detail can be gleaned.
- Section 3.4: Is this section about particle concentrations from the embryos, the shells, or both? Are there concentration size distributions of these particle types for the air at large, or in emissions from specific cities, that these distributions could be compared to? What does having these equations accomplish?
- Lines 325-326: I'm not sure why this statement couldn't be gleaned from Fig. 8 alone, without needing to fit to Eq. 12.

<u>Grammatical/Typographical corrections:</u>
There are quite a few minor grammatical errors throughout, things like “a” or “the” missing before words, misplaced commas, or subject/verb tense agreement. These don’t obscure the science being presented, but I recommend the authors ask for proofreading help from a source with professional proficiency in English. I’ve included some examples from the first couple pages below.

- Line 13: "to little <u>regard paid to</u>"
- Line 16: "A total of 289,461..."
- line 17: comma after bioprotein
- Line 17: vary → varies, in → among
- Line 18: "were performed as" → "were found to follow"
- Throughout: need a space between last letter of a word and the first parenthesis of a citation
- Line 27: "that leads" → "leading"
- Line 27: Add "the" before "number concentration"
- Line 235: "graupels" → "graupel particles"