The authors explore the spatial variability of riming and its contribution to the clustering of ice within clouds associated with wintertime precipitation in the midlatitudes (IMPACTS) and marine cold air outbreaks over the high latitudes (HALO-$(AC)^3$). They use a synergistic radar and in situ product to produce estimates of ice water content (IWC) with and without the influence of riming. By applying pairwise correlation functions to bulk microphysical parameters for a number of long flight segments (26 segments), the authors aim to capture length scales associated with IWC clustering and can further compare the functions separately between those applied to IWC including riming and those excluding riming. I particularly like the section where long swaths of in situ observations are broken up into smaller segments to essentially maximize the sample size of environments, thereby producing a more robust statistical analysis. This analysis reveals clear modes in the spatial clustering of IWC.

The paper follows logically towards its conclusions, and the figures are easily discernable and readable, and for that I thank the authors. However, I do have concerns with the robustness of the analysis. Additionally, I had confusion understanding the derived rimed and unrimed IWC, which hopefully could be better articulated/reorganized to improve upon the paper (mentioned in major comments).

I recommend this paper be reconsidered with major revisions.

Major comments:

Concerning the robustness of the results, I have two major points. First, I worry "artificial" positive pairwise correlation values are being produced by applying the moving average. This moving average is on scales of ~2 km, which is on the order of the largest observed positive correlations values (less than this value). I would propose sensitivity tests whereby varying the window size of the moving average. While I understand more robust measurements are obtained by averaging the in situ observations, it is very common to examine ice microphysical properties at 1 Hz scales (~ 100m). It would be especially prudent to use smaller windows for moving averages especially when looking at lags below a few km.

Second, there is no testing for the statistical significance of the pairwise correlation functions. This is especially a concern as standard deviations of the functions mostly overlap values equal to 0 (values expected of a homogeneously distributed system; Figure 7a,b). Further, some of the results of the rimed and "assumed-unrimed" IWC spatial inhomogeneity are nearly identical. If it's possible, applying some sort of bounds for rejection testing using white noise at some XX percentile could be helpful.

I also experienced confusion in the methodology in deriving the rimed and unrimed IWC. Concerning the organizational comment, for example, separating section 3.3 and 4.2 confused me. The derivation of IWC influenced by riming and IWC not influenced by riming seems to be separated into multiple sections (3.3 & 4.3), when section 3.1 is titled "Quantifying riming". I'm also still not sure how IWC can be separately obtained assuming riming and no riming. Are you simply using different coefficients in the mass-diameter relationship for the two variables (which I assume would be an issue since riming would in theory impact the diameter of "unrimed ice")? I'm sure it's explained in the text, however, it's difficult to determine.

Additional comments:

Line 6: delete comma after “understood”

Line 10: delete “closely” or rephrase

Line 39-40: Citation for this statement?

Line 43-44: What are the actual length scales of these smaller bands (also a citation speculating these processes would be nice).

Line 53-54: Why the long dashes?

Line 55-56: Citation showing the P3 scheme still struggles with ice processes (I get there are still broad concerns but a citation would be good when specifying a specific microphysics scheme)?

Line 63: “space-borne radar” is more commonly accepted nomenclature.

Line 65-66: should specify why measurements of IWC remain challenging (since the ensuing text implies a synergistic remote sensing/in situ method reduces uncertainty in IWC, which is misleading).

Line 94: You never define normalized rime mass. Please do.

Line 95: I’m not sure what you mean “by closure” (this was also said in the abstract). Please specify.

Line 117: rephrase “…and sampled at different frequency rates producing different spatial resolutions" or something similar.

Line 118: change “fly” to “flew”

Line 121: What does “good collocation” mean?

Line 142-143: Why is a CDP and a Fast-CDP used? Was one in error for different flights? Also, what are these probes used for exactly? Is it for PSD measurements the radar uses for calibration? Results from these probes aren’t shown anywhere in the paper.

Line 149: Although understood to be somewhat common to assume >50um is all ice, it is possible droplets can get much larger than this. While the potential of icing is often the rationale for this assumption, kinetic heating of the aircraft can avoid icing at temperatures a few degrees less than 0C. In fact, I wonder if the large ice particle concentrations in IMPACTS might actually be large drops in the -5 to 0C range.

To test this by doing a temperature dependent sensitivity test, I’m curious whether results overall might be sensitive to temperature ranges (possibly not, since you do the height analysis in Appendix B, but might be worth checking).

Line 179: the collocation of radar and in situ measurements can be as far as 5 km off? That seems pretty significant based on the spatial scales you’re using the pcf analysis.

Line 191-192: Perhaps I’m confused of what M really is, but isn’t it possible to obtain a sum of total M over the particle population? Unsure why an average M is being obtained.

Line 202: What are synthetic rimed aggregates?

Equation 2: Are results being binned at some specified length(s) (i.e., should r be r+dr)?

Line 225: Define gaps. I'm actually unsure of what your in-cloud threshold is.

Line 257: "…and second, on spatial…"

Line 277: "…particles larger than 50 um…"

Line 280: Isn't this the definition of effective diameter (area weighted mean diameter)? Are these properties equivalent?

Line 314: "…values associated with large particles…"

Line 321: "parameters"

Line 326: Did you mean by three orders of magnitude?

Line 336-337: What are the observed ranges of M? Reflectivity seems to vary by up to 30 dBZ.

Line 337: "dBZ"

Line 343-347: What is meant by riming being "minimal" while also increasing IWC by about 2/3s?

Line 350-353: "Isn't this only true where positive values of pcf(IWCr) overlap this pcf difference?

Line 354: "…larger than zero."

Line 364: "lags"

Line 403-404: Can portions of sub-segments be resampled?

Line 411: King probe LWC results aren't shown correct? If so state it.

Line 411: "suggests" not "indicates"

Line 412: "…supersaturation with respect to ice…" although this is somewhat dependent on whether you are within that -5 to 0C range or at colder temperatures.

Line 414-416: It's difficult to tell how significant the differences are between IWCr and IWCu. I get doing the difference to highlight this but the subpanels for the respective quantities' pcfs are nearly identical (also seen in Figure 7). I think doing some statistical robustness testing would sell this point.

Line 419-420: Again, LWC is not shown correct? I get the pcf(LWC)=0 contour is shown but the modes aren't. Worth at least specifying the modes are similar.

Line 427-428: Citation?

Line 435-436: Please refer to the M panels in Figure 10.

Line 447: "deposition" not condensation.

Line 444-446 & Line 449-450: where did you show these IWC clustering results?

Line 465: “…from cloud top.”

Line 480: “…larger than 6 km...”

Appendices: would be nice to keep the figures within the respective appendices.

Figure B1&C1: would be nice to show altitude or some sort of normalized cloud height rather than distance below the higher aircraft (unless this higher aircraft is essentially flying at constant altitudes).