Review of: “Hydrometeor classification algorithm through statistical clustering of polarimetric radar measurements: a semi-supervised approach.”

General comments:

This manuscript describes a proposed method for performing hydrometeor classification using a combination of statistical clustering of polarimetric radar observations and scattering calculations. The radar observations are first clustered purely statistically into subsets that represent the particle classes to be identified. A statistical test is then used to determine the hydrometeor type associated with each cluster or whether further clustering is needed. The algorithm is then demonstrated using a few different events and observations from different radars. The method described is innovative and appears to be an improvement over the fuzzy logic hydrometeor classification algorithms (HCAs) used in the study for comparison. As such, the readership of AMT should find this manuscript quite interesting. However, I have a few concerns that need to be addressed before this manuscript can be accepted for publication.

The first concern I have relates to the use of the scattering calculations in the algorithm. There needs to be more information provided to the reader about the microphysical assumptions that went into the scattering calculations (and thus the resulting membership functions), most importantly the aspect ratios and particle size distributions of the hydrometeors. These choices will influence the algorithm and therefore need to be justified.

Given that the authors' intention is to minimize the reliance of their HCA on the uncertain microphysical and electromagnetic properties of hydrometeors, the impact of these assumptions on the resulting algorithm needs to be shown. Therefore, I recommend the authors perform some sort of sensitivity analysis where the parameters used in the scattering calculations are modified and resulting changes in the cluster centroids are examined.

Another concern I have is the lack of discussion about the polarimetric variables used in the algorithm. It should be stated how these variables depend on the microphysical properties of the hydrometeors. It would also help to introduce in this section the hydrometeor types that will be used in the HCA. The degree of uncertainty in the polarimetric radar signatures associated with each hydrometeor type should then be discussed to motivate the addition of clustering to the HCA.

Finally, there are a number of instances of somewhat confusing wording, compounded in some cases by grammatical errors. Addressing these wording issues and grammatical errors will make the points highlighted by the authors clearer. Based on these concerns, I recommend that this manuscript undergo major revisions. My specific and technical comments follow.

Specific comments:

Page 1:

Line 11-12: It is unclear what is meant by “merging” here.

Line 19-20: Add references to support the uses of hydrometeor classification.

Page 2:

Line 16-21: Re-work this sentence or split it into two sentences. Also, define the polarimetric variables mentioned here and describe how they depend on the microphysical properties of the hydrometeors.

Line 26: Add references and discuss the uncertainty in the polarimetric signatures of different hydrometeor types.

Page 3:

Lines 10-12: The authors need to discuss further the differences between their proposed semi-supervised method and the method proposed by Bechini and Chandrasekar (2015) (i.e., why the method proposed in this manuscript is an improvement over the previous method).

Page 4:

Line 6: How can the standard deviation for each cluster be a scalar? Shouldn't this quantity be a vector since each polarimetric variable will have a different standard deviation for a given cluster?

Lines 11-12: Does the choice of algorithm affect the resulting clusters?

Page 5:

Line 16: How are the authors sure that no vertically aligned ice was present in the X band radar observations without assuming some signature for this class of particles *a priori*?

Line 23-24: Is this sentence implying that the four chosen cases from the Monte Lema radar were selected because more hail was reported during these events then the climatological frequency of hail events in the region?

Line 25: Are these plan position indicator (PPI) scans? Please clarify.

Page 6:

Line 15: Why is ΔH the relevant parameter for phase composition of the hydrometeors? Doesn't the degree of melting depend on both the height above or below 0 ºC and the lapse rate?

Line 24: I suggest adding the number of scans from each radar in addition to the number of pixels to table 1.

Page 7:

Line 3: Does this include correcting for non-uniform beam filling?

Line 5-6: I am confused by what is being done for the pixel assignment step to get the altitude of the 0 ºC isotherm. Why use a 6.5 ºC/km lapse rate when the temperature levels should be available in the model?

Line 7: How well do the 0 ºC isotherms altitudes derived from the polarimetric melting layer signature correspond to the 0 ºC isotherms altitudes from the model?

Line 12: Switch order of Figs. 2 and 3; as it is now, Fig. 3 is referenced in the text before Fig. 2.

Line 20: This discussion of the scattering calculations is vague. The details of the scattering calculations need to be presented if they differ from the calculations from the Dolan et al. (2013) and Dolan and Rutledge (2009) studies. There is some limited discussion of the scattering calculations in the Appendix that should be expanded upon and referenced in this section of the text.

Line 28: State the H_0 hypothesis being tested.

<u>Page 8:</u>

Line 5: In general, most radar sampling volumes will contain a mixture of different particle types. In many cases, one of these particle types will dominate the polarimetric variables. Therefore, in this section of the text, it would be better to say that the unlabeled clusters are mixtures of particle types that have more equal contributions to the polarimetric variables.

Line 6-7: What is the justification for limiting the observations to a range of 40 km?

Line 19: State what the number of samples S refers to specifically in the algorithm.

Line 30: Clarification: these 30 centroids come from the 30 iterations of the external loop of the algorithm?

<u>Page 11:</u>

Line 7-9: Rework this sentence.

Line 8: Determining the sensitivity of the classification algorithm to the assumptions in the scattering calculations/membership functions might be another way to illustrate the differences between semi-supervised and unsupervised methods. Doing this would also illustrate just how dependent the semi-supervised scheme is on the scattering calculation assumptions.

<u>Page 12:</u>

Line 6: It would be useful here to discuss the skill of the POH algorithm.

Line 10-11: Does the algorithm identify vertically-oriented ice during all cases where lightning is observed? One example where vertically-oriented ice with lightning is detected is not sufficient to suggest that vertically-oriented ice is actually present. Observations of depolarization streaks in Z_{DR} may be another way to support the presence of vertically-oriented ice during this period. See the paper by Hubbert et al. (2014).

Line 29: The microphysical assumptions, especially the assumed particle size distributions and aspect ratios used to generate the membership functions, need to be revealed either in section 3 or the Appendix. It is also important to summarize how these assumptions affect the centroids, given that the proposed method should not be overly dependent on the scattering calculations.

<u>Page 13:</u>

Line 15: How does considering only small sampling volumes eliminate hydrometeor mixtures?

Microphysical situations where mixtures of hydrometeors are likely (e.g., aggregates and pristine ice particles, rimed and unrimed ice particles, melting hail and rain) are still going to be encountered by the radar, regardless of the size of the radar sampling volume.

Technical corrections:

Page 1:

Line 5: Remove "eventually."

Line 13: Remove "finally."

Page 2:

Line 26: Remove "whereas" and split this sentence into two sentences.

Line 32: The word "distant" should be "distinct."

Page 3:

Line 1: Remove "whereas" and split into two sentences.

Line 4-9: Split this sentence into multiple sentences.

Line 13: Remove "whereas" and revise.

Page 5:

Line 26: Change "pointed" to "indicated."

Page 6:

Line 10-11: Remove parentheses and change "precipitations" to "precipitation."

Line 27: Remove "in particular."

Page 8:

Line 1: Revise this sentence.

Page 10:

Line 2: Change "pixels" to "pixel."

Line 28: Change "unsupervisely derived centroids" to "the centroids derived from the unsupervised method."

Page 11:

Line 7: Change “unsupervisely.”

Page 13:

Line 23: Change “was” to “were.”

Figures:

Figure 2: Add yes/no text to decision branches in the flow chart.

Figure 6: I recommend transforming K_{DP}' and ρ_{HV}' to K_{DP} and ρ_{HV} before plotting to make their interpretation easier. Change “CS” in the caption to “CR.”

Figure 8: Change the plotting color of the crystal class so that it can be better distinguished from the aggregate class. In the figure caption, describe the panels in the order of a, b, c, and d.

Figure 11: Change the color bars of panels b and c so that the negative Z_{DR} and K_{DP} values are distinguishable. Also, make the +, -, and x symbols in panel f larger.

Figure 12: Identify the locations of the radar and the two rain gauge stations on the map of panel c.

Refernences:

Hubbert, J. C., S. M. Ellis, W.-Y. Chang, S. Rutledge, and M. Dixon, 2014: Modeling and interpretation of S-band ice crystal depolarization signatures from data obtained by simultaneously transmitting horizontally and vertically polarized. *J. Appl. Meteor. Climatol.*, **53**, 1659–1677.