Review of "Hydrometeor classification from polarimetric radar measurements: A clustering approach" by J. Grazioli, D. Tuia and A. Berne

This is an interesting paper that is deserving of publication on the basis of the novelty of the HCA approach alone. This "microphysics-ignorant" approach to HCA needs to be contrasted with the more traditional ones (one of which is the Dolan-Rutledge HCA which is addressed). The early portion of the paper is fairly mathematical, and while some important integration of the mathematical analysis and the radar analysis has occurred, I am still left with the view that the mathematicians in the authorship could have interacted more closely with the radar expertise. The details of the clustering need to be more fully exposed in all the dual pol variables for the cases shown, particularly the cases shown in Figures 8 and 13. A number of substantive issues occurred to this reviewer in the reading of the manuscript and they are elaborated on below. All these points should be considered in preparing a final manuscript for publication. This discussion of substantive issues is followed by detailed comments and edits on the manuscript text.

Summary: Publish after appropriate revision

Substantive Issues:

(1) Radar data drawn from two locations/campaigns

The dual pol radar data for the analysis in this paper comes from two campaigns, and I found myself confused in several places about which data were used for what. When the two experiments are discussed in Section 3.1, it should be made clear that no information on liquid phase precipitation was available in the Davos case, but that abundant info of that kind was obtained in the Ardeche region. And then this should be followed up in each Figure caption to specify the region (Davos or HyMex) that provided the data that were used in each Figure. This is done in some cases, but not in the majority of them. Also, the information in the Figure 1 caption should match that in the text for the two regions. (For example, 'Ardeche' is used in the text and 'Montbrun' in the caption.) It would also be valuable to be more specific about the nature of the weather in the two places. If snowstorms were prevalent in Davos, what was their synoptic origin and were they ever sufficiently vigorous to make lightning. And if convective weather was present in HyMex, then was it sufficiently convective to make lightning, and to make deep mixed phase cores with graupel and hail. This info is quite important for revealing the full range of hydrometeor types that might be encountered in each case. The information about lightning is also relevant to the points made by the authors about "vertically aligned ice", also discussed in another substantive issue below. The cases illustrated in the RHIs in Figures 8 and 13 for HyMex and Davos, respectively, both suggest rather modest vertical development, but are these two cases generally representative? The description "30 snowfall events" for Davos could be elaborated on.

(2) Separation of radar data by meteorological regime

Very little is said about the nature of the convection from which radar data are drawn for cluster analysis. Since the hydrometeors that one has access to are obviously strongly dependent on that aspect, it would be useful to include. I realize that the authors are wanting a kind of microphysical-blindess to the analysis, but is it also really necessary to be meteorologically blind as well? If the conditions are entirely stratiform, with radar bright band, then the authors should say so. If the conditions are convective enough for graupel to form in isolated columns (no prominent examples shown) then the authors should say so. If there is any sign of electrification in the storms studied, then the authors should say yes or no. If rimed/unrimed snow, or graupel/hail, or rain were observed on the ground under the storms, that information would also be useful to report.

(3) The 2DVD instrument

This instrument is clearly important for the analysis in this paper (notable in Section 6) , but the instrument itself is not included in the "Data and Processing" section, where important details about the resolution of the hydrometeors should be included. In the Figure caption, reference is made to "Parsivel-type disdrometers". If they are the same as the 2DVD, that should be clarified. Furthermore, and most importantly in this topic, all the valuable rules for hydrometeor classification using the 2DVD instrument seem to be contained in another paper, but those rules should also be summarized briefly in Section 6 so that the reader is better qualified to judge the findings.

(4) Vertically aligned ice

"Vertically aligned ice" is insufficiently defined in this study. Ice particles can be aligned by both gravity and by vertical electric fields, and independent studies with an X-band dual pol radar in Brazil (Mattos et al., paper in review) have shown evidence for both behaviors, in incipient thunderstorms. This includes the evidence for a transition from positive ZDR values in the mixed phase to negative ones in advance of the first lightning discharge, and for the presence of negative Kdp values in highly electrified storm conditions. Too little info is given in this paper to make judgments about the importance of electric fields in this context. The gravit- aligned ice particles are most commonly conical graupel, and in this context it is worth noting that Dolan and Rutledge do not identify conical graupel, even though it is perhaps the most common hydrometeor in convective mixed phase precipitation. Furthermore, it is well known from Roland List's early work that conical graupel are present in storms in Davos, though it is still not clear whether the storms investigated here are in that category. Evaristo et al. (2013, AMS Annual Meeting, PDF available online) has shown evidence that conical graupel can exhibit negative values of ZDR, raising questions about what is really the target in Figure 9b for the rimed ice particles.

(5) Figure 5 (and Figure 8) pertaining to KdP

In light of evidence cited above for negative KdP values, together with the authors own evidence for vertically aligned ice, it certainly seems appropriate to include a negative axis for KdP in Figure 5. The big pileup of values near zero is surely suggestive that negative values also appear in the data set. The text on page 8486 reports that "Kdp does not seem to play a particular role in the classification of MS",

but that needs to be shown. The values of KdP should also be shown below the 0C isotherm in Figure 5. And in Figure 8 it might be helpful to display negative KdP values also. This storm may not be sufficiently electrified however for vertical alignment of ice particles by electric fields.

(6) The general problem (and quantification) of overlapping clusters

Despite the mathematical treatment, and the identification of a multitude of "cluster quality metrics" in section 5.1, way too little is said in this paper about the problem of overlapping clusters. Good examples are found in Figure 5, in the line plots of Figure 9, and in the analysis of EF in Figure 12. The great problem with use of dual pol data to distinguish various hydrometeor classes (i.e., supercooled drizzle from dry snow, dry snow without supercooled water from dry snow with supercooled water) is directly attributable to problems with cluster overlap. In Figure 5, the cluster overlap problem is illustrated and at the same time obscured by the overprinting of different colors, and the authors have little to say about this important issue. To be more specific, the Aggregates and the Rimed Ice are largely overlapped. The Drizzle and the Melting Snow are also strongly overlapped. The ability to distinguish among CR, RI and AG in Figure 12 on the basis of EF alone is largely discouraged by strongly overlapping behavior, particularly CR and AG. What best metric are the authors proposing to quantify this problem? Is it RS which is the index of dissimilarity? Why aren't numbers provided for real cluster examples? What values of RS are needed to clearly distinguish two clusters? My intuition is that some measure of cluster size must be sufficiently small relative to cluster "distance" (and there are a lot of those measures provided in Table 1, but I don't see them applied to the real clusters in the real radar data).

(7) Distinguishing dry snow with and without supercooled water (riming)

One of the most challenging problems faced in a recent FAA-supported field campaign (BAIRS- Buffalo Area Icing and Radar Study) aimed at validating NEXRAD dual pol data in winter storms, was the dual pol distinction between dry snow alone and dry snow with supercooled water (and riming). In that context, the contrasted results found in this paper in Figure 8 (essentially no riming in the cold part of the system) and in Figure 13 (abundant riming in the cold part of the system) really gets our attention. (Presumably in the first case the Bergeron process is eating up the supercooled water as fast as it appears, but in the more vigorous second case, excess supercooled water is available for riming.) But the frustration here is that the details are missing as to how the authors distinguished the two cases on the basis of the clustering of the dual pol data., and what is decidedly missing in this paper is the kind of analysis shown in Figure 5 (with dual pol variables) for the case from Davos shown in Figure 13 (in RHI form but without dual pol variables). If the authors have identified a reliable means to make this distinction between microphysical conditions, it would be highly valuable in the context of the aircraft icing hazard and the effective use of dual pol radar networks in both the United States and in Europe.

(8) Drizzle category

When I first compared the HCA categories identified with this new approach with the more conventional HCA that we have used in the FAA work with NEXRAD radars, I was delighted to see the authors' drizzle category. But I had also hoped that this was supercooled drizzle. On looking more closely at Figure 5, it

appears that not only is it not supercooled, but that the RHI in Figure 8 shows that it coincides with rain with reflectivity of order 30 dBZ (or maybe a little less). This is not likely real drizzle as this kind of stratiform rain will contain drops as large as 1 mm in size, and the definition of drizzle involves much smaller drop sizes (see for example the AMS Glossary on Meteorology). The authors need to have a careful look at this aspect, as it did not get much discussion in the text. In the US, some investigators are inclined to use precipitation with small reflectivity in stratiform conditions to calibrate the differential reflectivity, but the presence of drops of the order of 1 mm prevent the kind of high quality ZDR calibration that one would have if genuine drizzle were present.

(9) Role of differential attenuation

Since this study involves an X-band radar, it is surprising that no discussion is included about possible effects of differential attenuation on ZDR values. Could some mention be included about that?

(10)Comparison with other HCA algorithms

The authors do compare their hydrometeor classification results from their own examples with the Dolan/Rutledge results, but quantitative judgment is lacking. Does their method result in qualitative/quantitative improvement? (It seems not since the results between theirs and Dolan/Rutledge generally agree.) It would seem that their method needs to be implemented as many times in the future as necessary until all expected classifications are suitably encountered (sampled). (So as to apply the classifications carte blanche in the future on any data set in real time without requiring new clustering) Is that a true statement? If the authors have evidence after their comparison exercise that their method is superior, then those results should be highlighted.

(11)Robustness of final clustering

Along the way, the authors had to make a number of choices pertaining to distance measures, similarity measures, and ultimately, a choice in the number of clusters (7) that they chose. When using these techniques, it is important to demonstrate the *robustness* of their results in light of these choices. For example, in Figure 4, I see little difference in the choice between 6 & 7 clusters relative to the quality metrics the authors use. Are their conclusions sensitive to the choices made along the way in their algorithm?

Detailed edits on the text:

Page 8466

Abstract

Shouldn't you indicate somewhere that nc = 7?

Page 8467

You might tell why algorithms are different at different radar frequencies.

Page 8468

Line 5 Why do you say that the choice of classes is “mostly subjective”, when well defined microphysical entities are usually chosen? Please give a specific example of subjectivity in this context?

Line 11 Definitely an interesting approach.

Lines 15-17 Nothing is said here about distinct separability of the groups by the clustering technique. (See also substantive issue (6) above)

Page 8470

Line 17 It seems to me that this distance should always be discussed in relation to the cluster size, in some sense.

Page 8472

Line 7 I have tried to encourage above the provision of more prior knowledge about the radar data.

Line 25 OK, this tells that you are not going to do any analysis on liquid phase precipitation with this data set. But then you should say that you will be doing analysis on this aspect with the other data set, when it is described on the next page.

Page 8473

Lines 4-5 ZDR is not really calibrated with vertically pointing, but rather only zeroed.

Lines 8-9 The authors could say much more about the meteorological characteristics of these storms that they do here.

Line 12 Again, it is not really a calibration of ZDR, but a zeroing of this quantity. (Other values like +2.7 dB are not checked by this vertically pointing method.

Line 17 OK, but work on rain will require the HyMeX data set. You should say so.

Page 8474

Ground clutter sample volumes are not included. How was it determined that a sample volume was ground clutter (automated or human-in-the-loop)?

Page 8475

Lines 16-18 Negative values can be meaningful and should be examined.

Page 8476

Equation 10: units should be included. It is not clear to me why you want X to be zero if you are in warm part.

Page 8477

Line 3 I don't understand what "full distance matrices" are.

Lines 7-9 But always with radar making observations above the melting level?

Line 12 Useful to remind reader exactly what an "observation" is. If that is equation (7), then give it.

Line 27 What is meaning of "correlative distance metrics"?

Page 8478

Line 5 Why was the number 1000 chosen here?

Line 16 This jumps ahead in the figures list.

Page 8481

Line 25 How is goodness quantified? There is nothing about "goodness" in Figure 2b. And nothing about cluster quality indices.

Pages 8482 and 8483

Why don't these cluster quality metrics get applied to the real data. If they do not get applied, they are orphaned here. Did I miss something?

Page 8485

Section 6.2 should begin by saying that HyMeX radar data are being examined here.

Line 10 That would be a bad assumption if large graupel or hail were involved, but the authors have not said anything about that aspect.

Page 8486

Section 6.3 should begin by saying that the same HyMeX data are being examined here.

Line 13 It would be useful to cite the observed thickness as additional evidence for this assertion.

Line 17 Many dual pol experts claim that rhohv is the best variable to show the melting layer. Do the present authors agree?

Page 8487

First line: what quantitative info is provided? This could be beefed up.

Page 8488

Line 3 Change "colder" to "lower"

Line 13 What is mechanism for “vertically aligned crystals”

Lines 20-21 We need to know about the rules in this paper.

Page 8489

Line 23 The color looks turquoise rather than light blue on my print.

Line 24 How are graupel ever associated with ice crystals? This is just one reason that the spatial resolution of the 2DVD instrument needs to be given, as well as the rules for identifying hydrometeor types.

Page 8490

Line 7 Additional need for knowing spatial resolution.

Line 15 Why not say “1 min”?

Line 21 Yes the snowfall intensity differs, but there is substantial overlap. (The differences do not allow for a unique specification.)

Page 8491

Line 5 I would add at the end of this first sentence “in Davos”

Line 13 Quantify “very low values of Kdp”

Line 14 OK, but are clusters cleanly distinguishable? What about the overlap problem?

Line 25 “to a dataset of radar data” is redundant

Page 8492

The authors state: “The main limitations of the method are related to the interpretation of the content of the clusters, that might not be trivial especially if no ground reference is available for comparison.” Never mind the ground reference. What is really needed here is in situ aircraft validation.

Lines 24-25 You can compare with still other HCA schemes. Maybe there is some reluctance because other schemes were not developed for X-band (other than Dolan and Rutledge).

Lines 28-29 This would be easy to test with NEXRAD and European radar networks, and the authors should say something about this possibility.

Appendix A

It would be helpful to see the actual step-by-step calculations for their 5x5 matrix that leads to the SS scores?

Table 1

Where do we find example numbers for these parameters? This is one example of need for tightening of linkage between mathematical development and the actual dual pol radar data in this paper.

What about discussion about relationship between 'distance' and cluster size in the context of real data (Figures 5 and 9 for example.)

Page 8501

Do you mean "orthogonal"?

Table 3 This table should include the actual final categories on the left side (Drizzle, Light Rain, Heavy Rain)

The estimate of 0.12 mm/hr for drizzle is too high, and adds to my point above that this is not really drizzle. (Authors should look at the disdrometer data in the literature in stratiform rain, and almost invariably one has drop diameters exceeding 1mm which is definitely out of the drizzle category)

Table 4 The caption should include the meanings of a,m,b

I would like to know the man ZDR value for high density graupel, as one expects gravity-aligned conical graupel in many instances.

Table 5 The caption should give information about the four parameters given on the right.

Figure 1 The info in this caption about radar sites should agree with that given in the text. Why thes choices for "the directions of the RHI scans"? Why is this important for the paper? Reference to "polarimetric power laws"? I don't recall much about that in the text? Are these power laws on rain?

Figure 2 Shouldn't the flow chart include something about which data are being used? It seems decoupled from the real observations.

Figure 4 Which of two data sets is used here? The caption here should give some idea why the final choice of nc=7 is what it is. (This is not obvious on the basis of the figure alone.)

Figure 5 This is a key figure and there should be others like it in the paper, and particularly for the Davos analysis in Figure 13 showing abundance of riming in contrast . The caption here should say which data set. The color coding should be included, in the same form as in the RHI analysis. The caption should also address the issue of overprinting of data, and the implication for cluster overlap and difficulty in separability.

Figure 6 Which data set is used? Presumably HyMeX but this should be stated.

Figure 7 "collected on the 24 September"

Figure 8 It is remarkable that there is no riming at all, and the discussion on the contrast with Figure 13 should be expanded in the revised manuscript. I wonder if you had a +ZDR "bright band" in this case?

Figure 9 “Distribution of dual pol parameters”. What data set was used? Your rimed ice showed no evidence for negative ZDR values? This is odd given other results with X-band dual pol radar data. There is a lot of overlap here, in nearly every case. But the potential problems of the overlapping are not discussed.

Figure 10 Suggest change from “freezing” to “sub-zero” temperatures. Additionally, just a reminder that Dolan and Rutledge did not recognize conical graupel, despite having two categories for that hydrometeor.

Figure 11 Ditto on change of “freezing”

Figure 13 This case stands in marked contrast with Figure 8, but the details about the dual-pol clustering (that is, in the same form of Figure 5) are not included and should be. The caption should also address all five parts of the figure, and should repeat the info from Figure 8. I am puzzled about the radar being at 2133 m when you have complete RHIs to lower levels. If the zero of altitude is the radar altitude then that should be so stated.

Figure 14 More should be said about the synoptic situation to put this rimed case in context.

End review

Earle Williams

September 25, 2014