**Atmospheric
Measurement
Techniques**

Open Access

Discussions

Interactive comment on "Hydrometeor classification from polarimetric radar measurements: a clustering approach" *by* J. Grazioli et al.

Anonymous Referee #1

Received and published: 5 September 2014

First, I want to underscore that I think this is a very good paper with a good methodology and great contributions to how hydrometeor classification is done. I have reviewed other statistical papers using clustering analysis for HC purposes and this is by far the best one I've seen. However, there are many grammar and writing style corrections that should be made to this manuscript but that which I do not want to specify here because there are too many and it is to tedious to do over this format. Instead, I have listed more science related issues I had with the manuscript:

QUESTIONS AND CORRECTIONS:

8466, 13: 3000 hrs is 125 days - is this right? 8467, 3-8: I do not understand why you accentuate the use of Doppler radars? Doppler (velocity and spectrum width) data are not necessary or used at all in hydrometeor classification. However, they could be used to compare hydrometeor classification to storm motion and mesoscale flow patterns as in Dolan and Rutledge 2010.

This last sentence doesn't make much sense either. It needs to be rewritten and split into two sentences: "This allows a single instrument, the radar, to acquire multiple simultaneous measurements that are sensitive to... This facilitates ...."

8468, 5: I do not understand why you accentuate the use of Doppler radars? Doppler (velocity and spectrum width) data are not necessary or used at all in hydrometeor classification. However, they could be used to compare hydrometeor classification to storm motion and mesoscale flow patterns as in Dolan and Rutledge 2010.

This last sentence doesn't make much sense. It needs to be rewritten and split into two sentences: "This allows a single instrument, the radar, to acquire multiple simultaneous measurements that are sensitive to... This facilitates ...."

8468, 7: The dielectric properties are probably well known - it's either ice or water or some mixture. I could see this being an issue for graupel or rimed ice only.

The particle size distribution is the biggest unknown for scattering simulations. I would emphasize this in the text by writing "and largely unknown particle size distributions."

8470, 11: why isn't "Correlative Distance" included all in quotes like " Min... Distance"?

8472, 5-9: Yes, exactly! This could be emphasized more, this is the strength of your paper because you do not predefine or constrain your MBFs or expected value range clusters with a priori knowledge.

8472, 13-25: this is repetitive and many sentences could be combined.

8473, 18: attenuation is always in dBZ/km... not dBZ/km.

8474, 3: I think this should be radar range gates potentially contaminated... because you hopefully aren't removing the entire radar resolution volume if it contains some ground clutter. Radar resolution volume refers to all the scans. Usually you can just remove bad gates.

8474, 24: How many weather stations are within 40 km of the radar? In the US, that would be a touch criteria because our stations are much more spread out! Could you specify how many stations you used usually for a given algorithm run?

8475, 9-10: Great idea!!!

8475, 20: here, and in other instances of using this word, "components" is a little vague and your readers might understand it better if you clarified that X contains all the radar variables, like a structure. "The different radar variable fields contained in X..."

8476, 4 - you most likely mean that the bounds were allowed for each polarimetric variable, not assigned to?

8477, 13-17: rewrite for clarity and split these sentences up

8478, 18: "We define now a local... is poor word choice, awkward.

8480, 3: what do you mean by global? Throughout the radar resolution volume? How is "global" applicable to this radar algorithm?

8481, 21-25: rewrite, these sentences don't make very much sense.

8482-3, all lines: you are already using 1. 2. 3. to indicate the list, you should take out "A first one... " and "A second index" because it is redundant and the language is too casual to be suitable for a technical journal anyway.

On section 5.1 - Do you actually use all these methods that you describe? How? Clarify how/when/where they are used.

8484, 4-5: rewrite, these sentences don't make sense.

C2477

Section 5.2 - please specify which index you are talking about throughout this section instead of just saying "the index".

8485, 6-7: rewrite because this is confusing + I need a reminder what delta z means at this point too - redefine or remind the reader of it's meaning? Throughout the paper I was confused what delta z was supposed to represent

8485, 19: could clarify that this is the linear reflectivity

8486, 5-6: Instead of using the vague parenthetical phrase you currently have at the end of this sentence, you should add a clarifying sentence here, "We conclude that results from the proposed method agree well with BR2009."

8485, 10: You DEFINITELY need to add a summary statement to this paragraph - does this seem realistic? what does this mean for your algorithm? were you surprised? did this confirm your hypotheses? etc.

8486, 17-18: Could cite Brandes and Ikeda 2004 (Journal of Applied Meteorology, AMS) for this. Thompson et al. 2014 (Journal of Oceanic and Atmospheric Technology, AMS) also found that Kdp was not useful in melting layer detection algorithms. Your results suggest the same thing - Kdp is too noisy to be discriminatory or trustworthy in classification of melting snow.

8468, 21: why or how is it more difficult to interpret?

8475, 6: Kdp of 25 deg/km is ridiculous! This cannot be right, or it must be from a bogus simulation!

8476, 6: what does 3log mean? this doesn't mean anything in the way it is currently written

8478, 13: Should use real nouns to describe what you are actually talking about / referring to. "the dark green one" is vague and unprofessional for technical writing. - should be "dark green cluster" This happens often throughout this portion of the

manuscript.

8478, 14-15: rewrite, this doesn't make sense

8478, 22-25: Kdp doesn't really depend on concentration exactly. You definitely need to clarify in the last sentence that "the dark green cluster contains, on average, more OBLATE hydrometeors and/or OBLATE hydrometeors of larger size and density." Kdp ONLY increases beyond zero for oblate (nonspherical) particles, so it does not matter how many particles you have, only how many are oblate.

8488, 5-8: I think you mean "led us to hypothesize" and "These hypotheses" instead of "led us to assume" and "These assumptions"... which is a more accurate word for the context

8488, 19: This reference is for 2nd generation 2DVD. Schonhuber et al. (2008) talk about the third generation - might want to check which one you are using?

8488, 22-23: You need to specify here that 2DVDs capture SURFACE precipitation or images of surface precipitation particles.

8488, 23-24: hydrometeor type is probably not "recorded by the instrument." Can you please clarify whether you are manually looking through 2DVD photos? How are you classifying particle types with these images? Is it done manually or with an algorithm from the instrument?

8488, 25: what do you mean by "averaged in time"? It's every 60 sec... so do you average the 2DVD records about the time of each radar scan?

8488, last line to 8489, 1-2: do not capitolize Graupel or Rimed or any of these names - they are not proper nouns. Furthermore, why is it RIM if it is RI in the rest of the paper? What does the M even stand for?

8489, 3-4: need to clarify that "there is usually a mixture of different hydrometeors captured by the 2DVD so they do not necessarily exhibit a single pristine shape.

C2479

8490 8-13: A big problem I have with this comparison is that 9 deg elevation angle is VERY high, how is even possible for the 2DVD to remain within 400 m of the radar beam? You need to calculate how high off the ground the lowest beam is and then how far away this is from the surface. If the two data sources are indeed far away, then horizontal advection or blowing of particles in the wind will vastly affect what particles existed above the ground and what actually falls the ground. Crystals are very light and irregular and fall very slowly. Furthermore, during this gap between the radar and 2DVD, the snow could take on a new crystal habit with changing temperature and moisture conditions such that the hydrometeor type would change due to this as well.

8489, 13: I would clarify that the sampling times are different but the do overlap.

8489, 18: 150 m horizontal distance is usually the gate spacing along the beam! So are you just picking the closest gate to the 2DVD below?

8489, 23: should clarify that the clusters are "similar between methods". not just "similar".

8489, 23: Except for the graupel part of this result, this sounds fairly good. If the algorithm is putting dendrites and small particles into CR, aggregates into AG, that sounds like it's working? Maybe you should clarify that this is a positive result? You don't make any comment either way at this point.

8489, 26-27: rewrite: "Furthermore, small particles are least likely to be classified as RI by ... which algorithm?" Need to specify which algorithm you are referring to in all of these comparisons, it gets really confusing.

8490, 6-7: I would add that this is probably happening because aggregates tumble a lot as they fall, making their Zdr near zero but reflectivity quite high due to large diameters. Rimed ice would have similar signatures nearly because rimed ice becomes more spherical and therefore falls more uniformly, more dense.

8490, 18: You need to clarify how De is calculated - was it calculated for you by 2DVD

software? Do you know how it was calculated, i.e. under what assumptions of the PSD shape - usually exponential? This does affect De accuracy, as you mention. Since you mentioned the uncertainty, you should explain more about it.

8490, 18: Again, you need to use NOUNS. Cannot just say "those ones" or "these three" this is completely vague, undescriptive, I can't tell what you are talking about, and it's too casual of writing for a scientific journal.

8490, 22-25: I would add that as riming progresses, the original crystalline shape of ice particles becomes indistinguishable, therefore reducing the drag and irregularity of crystals as they fall in addition to making them more dense. Hence, their fall behavior is much more smooth and therefore faster, i.e. they are not fluttering or wobbling as much during descent because of the riming. See Pruppacher and Klett - Microphysics of Clouds and Precipitation textbook.

8491, 1-4: you need to clarify whether these statements are in agreement with observations or physical understanding of storms or if this is a comparison between algorithms - which ones? These are very vague sentences which don't really mean anything on their own.

8491, 9-11: descriptions of heights in parentheses is really confusing, the last one should just be "within 3.5 km of the radar"

8491, 19: how high? you can't just say "higher." You should specify altitude so we can compare directly with the altitudes you referenced for different hydrometeor types in the previous paragraph in order to compare... otherwise, really cool that you were able to use a model!

8492, 14: should be "with dual-polarization radar data"

8492, all remaining text: needs to mostly be in past tense, not present. The verbs need to agree and be in past tense because you are summarizing something that you have done.

8492, 3-4: what do you mean by "reduced" and what do you mean by "representative subset"???

8492, 14: if the abbreviation is RI, you shouldn't expand it as rimed particles... where's the i? Should say rimed ice particles.

8492, 19-20: should clarify that you probably mean that "this technique is immune to radar calibration"... "independent of systematic biases that may affect the polarimetric observations" is really vague and could mean anything... systematic bias like cool temperatures?

8492, 24: PLEASE emphasize/change text to read that the "limitations of the method are related to the MANUAL interpretation of the contents of the clusters. This may not be trivial, especially in the absence of surface precipitation type reports for comparison" It is important that the interpretation is manual and not automated, like other parts of your methodology.

8492, 27-29: do you expect the results not to differ very much for different systems besides your own because you were using a well-calibrated radar? was your radar well calibrated? I agree that the results hopefully won't change much from radar-to-radar if they are well calibrated. HOWEVER, one major point you should make is that Kdp is inversely proportional to wavelength, so you will DEFINITELY see differences in your clusters using Kdp at X, C, or S band.

8493, 1-3: (1) could you employ different spatial smoothness constraints on or between different categories? I could see this being very physical because some precipitation transitions are more gradual or abrupt in nature. (2) What does "when the full database of observation is hierarchically clustered" even mean? these words don't make sense; rewrite.

8493, 6-9: True! Good point, this is a strength of your results/algorithm.

8493, 24-25: (1) should be "...DR2009 is from Dolan and Rutledge (2009)". (2) What

adaptations did you make? How did you determine or find delta z? If you are going to cite their algorithm, you definitely need to specify how you "adapted" it!

Table 2 caption: should be "dual-polarization radar"

Table 3 caption: Should specify that Q5 = 5% quartile, your notation isn't readily obvious.

Table 4 caption: need to restate what a, b, and m are from the definition. Figures/tables and their captions should be self-sufficient from the text.

Table 5: Same problem as previous statement - need to define what r, I, T, and delta z are in this table.

Fig 2: should it be "n" underscore "opt" instead of Nopt? the text has it n_opt everywhere else - stay consistent

Fig. 3: when is this referenced or used in the text? Specify how the order 1-8 is important or used if you show it in the figure

Fig. 4: Make the dashed vertical line yellow to stand out from the rest of the graph, specify that "... the RKR, RS, and AS INDICES are a function.... " - again using nouns to describe what these acronyms are

Fig. 5: this figure is way too small and the figure quality is bad. You also really need to have a legend for what the different colors and symbols represent. Your figure caption should also better explain what the data are. - WHAT ARE THE 7 CLUSTERS AND WHAT COLOR ARE THEY?

Fig 6: Need to make legend with real names, not acronyms, doesn't make sense. Rewrite figure captions to say "Categories from Dolan and Rutledge (2009) are indicated on the x-axis while the proportion of cluster members from the proposed algorithm belonging to a given category are indicated by percentiles on the y-axis. .... I think? I really don't like this figure - which categories are from your cluster and which

are from DR2009? Need to indicate what's in the colors and legends better!

Fig 7: PLEASE add variable labels to each subplot just saying "Zdr", "Kdp", etc etc. (a) (b) (c) etc don't mean anything in this case and our eyes need to know immediately what field we are looking at without reading the caption. Also, you need to add UNITS to your colorbar [dB], [deg/km], etc.

Fig. 8: You could make the colorbars narrower (i.e. taking up less space vertically) in order to maximize the space you should be using to show the reader radar data. Also, PLEASE label each subplot as "Zdr", "Kdp", "Zh", etc instead of (a), (b), (c), etc because the letters mean nothing. Also, you need to add units to your colorbars to indicate dB, dBZ, etc.

Fig. 9: Need to label y-axis and make the legend out of real words, not acronyms. Do not use [-] for a y-axis label... tells us nothing. And the figure caption doesn't mean anything either, what does "distribution" mean? Do you mean a normalized histogram? so normalized frequency of occurrence?

Fig 10 and Fig. 11: See all comments for Fig. 6

Fig. 12: You should say that this is the probability density function and the cumulative density function (PDF and CDF)... much more descriptive and accurate.

Fig. 13: See all comments for 7

Fig. 14: I can't read this - everything is way to small and the lines are too light and the labels are not large enough to read on the contours. Why are legends for orange, green and TWO types of blue included? First of all - my eyes can't tell the difference between two colors of blue. Second of all, it doesn't look like there is any orange or green in the figure.

OTHER systematic things I noticed that could be improved:

reference to your figure panels in text: panel (a) looks and reads better than panel a -

see 8478, 4

Proper way to use i.e. is to have a comma before, not after the i.e.

Your way: a metric i.e., a measure... Proper way: a metric, i.e. a measure...

A list of three should have two commas: one, two, and three. You usually do this instead: one, two and three.

A list of two things between "and" should not have a comma but you often include one anyway. Example: One, and two (your way). One and two (proper way).

These should be revised throughout the manuscript.

———————————————————

Interactive comment on Atmos. Meas. Tech. Discuss., 7, 8465, 2014.