Hydrometeor classification from polarimetric radar measurements: a clustering approach
amt-2014-238

Jacopo Grazioli, Devis Tuia and Alexis Berne

November 12, 2014

Responses to the reviewers

With the present supplementary document we provide our responses to the comments of the three anonymous reviewers of the manuscript *amt-2014-238*, entitled "Hydrometeor classification from polarimetric radar measurements: a clustering approach ".

The comments of the reviewers are reported in *italic* font. Quotations of the manuscript in its revised or in its original form are reported in blue . The figures used to support our answers to the reviewers can be found at the end of this document. Before to proceed with our answers, we would like to list some of the main changes with respect to the AMTD manuscript:

1. Heavy rephrasing and improvement of the sections describing the interpretation of the content of the seven hydrometeor classes.
2. More details and statistics are given for the seven clusters.
3. Negative K_{dp} values are now included in the clustering algorithm.
4. We identified a problem with the Z_{DR} zeroing (calibration) during some events of the Davos campaign, causing too high Z_{DR} values to appear. These values were only slightly affecting the clusters, but they degraded the comparison with HC2DVD, presented in Sec 6.4.3. We recomputed all the comparison once this issue was fixed.
5. The previous appendix A has been removed.
6. The sections describing the comparisons between HC2DVD, DR2009 and our method have been largely rephrased and a quantitative evaluation has been performed as well.
7. The most difficult figures to interpret have been substituted by tables.
8. The manuscript has been edited by a native English speaker.

Anonymous reviewer 1

We would like to start by thanking the reviewer for his/her valuable contribution, for the many useful suggestions and corrections and for his/her constructive criticism that lead, in our opinion, to a significant improvement of the quality of the manuscript.

General comments

1. *Substantial contribution to scientific progress? i.e. concepts, ideas, methods, data*

   This paper presents a very well-described, well-executed hydrometeor classification method. They compare their methods to Dolan and Rutledge (2009) and found good agreement. They also used a comparison to 2DVD data that was informative, although the presentation of this comparison was quite confusing and needs to be revised.

   We thank the reviewer for his/her positive feedback. We agree that the comparison with the 2DVD-based hydrometeor classification method was not presented clearly enough. The section has been largely rephrased. We would like to underline that the algorithm describing the hydrometeor classification from 2DVD data is now published and available online as Grazioli et al. (2014). We are confident that after the major rephrasing and with the availability of the final version of Grazioli et al. (2014), this part of the manuscript is now clear and understandable.

2. *Are the results within the scope of AMT? i.e. remote sensing, laboratory and in-situ measurement techniques for constituents and properties of Earth's atmosphere, i.e. development, intercomparison, and validation of measurement instruments and techniques of data processing and information retrieval for gases, aerosols, and clouds.*

   Yes.

3. *Are scientific approach and methods valid?*

   Yes, except for the 2DVD and BD2009 comparison, which was not presented or discussed well. The graphs and text were very unclear and definitely need revising.

   We agree, and we modified the text and the figures according to the suggestions of the reviewers.

4. *Is the experiment reproducible?*

   It would be much more reproducible and comparable to future/past studies if the author's new clusters for various hydrometeor types were outlined in such a way that shows each category's polarimetric variable ranges: max, min, mean, median, etc. You say your algorithm works, but how can I compare it to mine if I don't know your ranges of expected values that result from the cluster analysis? Furthermore, a primary contribution of the paper is to establish the expected value ranges or membership beta functions of various hydrometeor types from a large European X-band radar dataset. You should provide this to readers! They could use the data and implement it in whatever algorithm they have. Your algorithm goes on to use some other methods of final classification, but as you say in the conclusions, your expected value ranges could be of use in any algorithm design.

   We thank the reviewer for the good suggestion. The clusters found in our database needs to be better outlined, for future comparisons with data collected by similar instruments and/or in similar climatologies. We added a table in the appendix with all these statistics (Table A1 in the revised manuscript).

5. *Are results discussed in an appropriate and balanced way?*

   Yes.

6. *Consideration of related work and appropriate references?*

   Yes.

7. *Are scientific results and conclusions presented in a clear, concise, well-structured way? i.e. number and quality of figures/tables, appropriate use of English language?*

   The grammar and writing style of the paper started out fantastic, but greatly degraded as the paper went on. I would gauge the writing quality as poor for all sections but the 1st and 2nd. To complicate things further, the paper was VERY LONG. The major problems with writing and use of the English language were: the very frequent misuse of commas; placement of prepositional phrases in the sentence structure; misuse of dataset/ data set/dataset; using dual polarized radar (incorrect) instead of dual-polarization radar (correct); overusing the words "employ," "following," "present study," "here"; unnecessarily using [-] to denote unitless variables – this is confusing; beginning too many sentences with extraneous prepositional phrases that make the sentences very awkward and hard to read like "In this way, At first, In the following, Additionally, Eventually, etc. etc – these can be great ways to start sentences but they are very commonly misused and misplaced in this paper; the misuse/mixture of present and past verb tenses when it should be past tense in most cases.

   We carefully revised the manuscript with additional help from a native English speaker. About the length of the manuscript, in the present revision we managed to shorten the description of the algorithm but we increased a bit the discussion part. Given the many constructive suggestions received, the best that we could achieve was to keep the paper of approximately the same length, and reduce the number of figures. We do not agree only about the notation "[–]", that we do not consider confusing to denote unitless variables.

   In addition, some figures need to be improved and some need much more clarification so the reader can easily understand them. Many figures rely on the reader to know acronyms and methods from the text, but these details should be reiterated in clear legends with real words (not acronyms!) and the same pertinent information should be clearly stated in the figure caption. The figures and their captions are incomplete and not understandable in the current state.

   We improved the figures according to the suggestions of the reviewer(s). Some of the figures previously used to compare among different methods are now replaced by contingency tables (or confusion matrices when one method is considered as "reference"), that are easier to interpret.

Technical corrections

1. *TECHNICAL CORRECTIONS: TOO MANY corrections - will take more time to denote these, trying to figure out a proper format to do so. I will probably annotate a PDF and attach it.*

   We thank the reviewer for the many corrections provided in his/her second PDF. We address these comments in the following sections.

Specific comments

Section addressing individual scientific questions/issues

1. *I think the paper could have a greater impact and be much more reproducible and be more comparable to future/past studies if the author's new clusters for various hydrometeor types were outlined in such a way that shows each category's polarimetric variable ranges: max, min, mean, median, etc. You say your algorithm works, but how can I compare it to mine if I don't know your ranges of expected values that result from the cluster analysis?*

   We agree with the reviewer about this point and we added this information in a new appendix, were we provide a table containing the statistics of each polarimetric variable within each hydrometeor class (clusters shown in Table A1 of the revised manuscript).

2. *Furthermore, a primary contribution of the paper is to establish the expected value ranges or membership beta functions of various hydrometeor types from a large European X-band radar dataset. You should provide this to readers! They could use the data and implement it in whatever algorithm they have. Your algorithm goes on to use some other methods of final classification, but as you say in the conclusions, your expected value ranges could be of use in any algorithm design.*

We believe that by addressing the previous point of the reviewer (Specific comment 1) we answer to this request as well. Now we provide in the manuscript much more information about the polarimetric variables ranges for each cluster. This will allow any interested reader to build its own set of MBF with the information provided.

We want to underline that in the manuscript we never used MBFs to classify images with the novel algorithm, but only when we employ the reference algorithm DR2009, based on Dolan and Rutledge (2009) and described in appendix B.

3. *The other scientific problem I noticed in this paper was the presentation and discussion of HC-2DVD and BD2009 comparisons with your proposed method. The text, figures, and figure captions need to be revised.*

   We followed most of the suggestions of the reviewer in order to improve the comparisons with HC2DVD and DR2009 and these sections have been revised. Additionally, following a suggestion of reviewer 2 and reviewer 3 we added some statistical scores to quantitatively compare the algorithms. This is an additional contribution of the present manuscript. We are also aware of the difficulties of such quantitative evaluations, and we clearly state our hypothesis, methods and limitations in the manuscript. In fact, the goal of the present manuscript is not to compare with a set of other algorithms, but to illustrate how the application of unsupervised techniques is promising for the field.

4. *There was a confusing mix of the words "index" and "score" in section 5.1*

   The term "score" in section 5.1 has been substituted by "index" or "smoothness". As an example we substituted expressions like "the lower this score is" with "the lower this index is".

5. *I think "spatial smoothness" makes sense to me, but you mention "data-wise" similarity or "data similarity" many times. It is less clear what this means, perhaps using more physical words to describe what you mean in the beginning could help.*

   Data similarity refers to the similarity of polarimetric observations, i.e. the radar "data". Two observations, or range gates, are data-wise similar if the polarimetric variables collected are similar , independently of the location (and collection time) of these gates. In the revised manuscript we simplified the first sections (describing the clustering algorithms) and we added some reminders throughout the text of the definition of the terms.

6. *The introduction section (1) was very well-written. A few changes (listed below) need to be made for the purposes of correctness and to improve the writing style, grammar, and readability*

   The changes proposed by the reviewer are addressed in the next section.

7. *The background and clustering techniques section (2) was well-written with minor errors in writing style and word choice. It was very well-explained. I do not have a statistics background but I learned a lot and could understand most of this, thank you!*

   We thank the reviewer for the positive comment. The goal of this section was indeed to provide this information to researchers who are working with weather radars or other instruments to measure precipitation. We believe that this short introduction can generate at least some curiosity for techniques that might be helpful in other aspects of precipitation measurements.

8. *The data and processing (3), clustering of polarimetric radar data (4), and selection of the optimal cluster partition (5) sections had many more technical writing errors related to word choice, grammar, and sentence structure that should be rewritten for clarity.*

   The changes proposed by the reviewer are addressed in the next section. Additionally, these sections have been further revised according to the suggestions of a native English speaker.

9. *Section 6 "from unlabelled clusters to hydrometeor classes" had great scientific content but needs major revisions in order to better understand the comparisons made between your proposed method, BD2009, and the 2DVD.*

   Please, refer to our answers above: answer (1) of the "General comments", and answer (3) of the "Specific comments".

10. *The conclusions section (7) contained technical writing errors.*

    The section has been further revised according to the suggestions of the reviewers and according to the revision of a native English speaker.

Additional questions and corrections

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 too tedious to do over this format. Instead, I have listed more science related issues I had with the manuscript:

We really appreciate such a positive feedback and we are confident that the revised manuscript will now meet a higher standard of quality for the grammar, for writing style, and for the scientific contents.

1. *8466, 13: 3000 hrs is 125 days - is this right?*

   3000 hrs does not strictly mean 125 days. During the field campaigns described in Sec 3.1 the X-band weather radar (MXPol) was operating continuously. The totals provided in the section referred only to hours when precipitation was visible in the domain covered by MXPol. In order to clarify this concept, we rephrased the sentence as:

   The approach is applied to data collected by an X-band polarimetric weather radar during two field campaigns (from which about 50 precipitation events are used in the present study).

   In this way we refer only to precipitation events, and not anymore to hours of precipitation.

2. *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).*

   We removed the word "Doppler" from this sentence, as it results to be confusing for the reviewer. We rephrased as follows:

   The most recent HC techniques require polarimetric capabilities. This allows a single instrument, the radar, to acquire…

   However, when we mentioned "Doppler" radars we did not mean only velocity and spectrum width measurements, but we referred more in general to the ability to measure the phase of the waves. Strictly speaking, a non-Doppler polarimetric radar would be able to measure Z_{DR} but it could not measure Ψ_{dp}, hence K_{dp}. We use now the expression "polarimetric", as suggested by the reviewer, and the context makes clear that the radar is Doppler too.

3. *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…"*

   We agree with the reviewer. The sentence now reads:

The most recent HC techniques require polarimetric capabilities. This allows a single instrument, the radar, to acquire multiple simultaneous measurements that are sensitive to distinct characteristics of precipitation. This facilitates the understanding of many microphysical processes. . .

4. *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."*

   We agree with the rephrasing proposed by the reviewer. The sentence now reads:

   . . . uncertain for ice-phase hydrometeors, because of the complex geometries, dielectric properties and largely unknown size distributions of ice particles

   However we do not agree with the reviewer when he/she says that the " dielectric properties are probably well known". Ice-phase hydrometeors are mixtures of air, ice and liquid water in different proportions. Their dielectric properties would be "well known" only if these proportions would be well known (Westbrook et al., 2006), and their shapes as well (Tyynela et al., 2011).

5. *8470, 11: why isn't "Correlative Distance" included all in quotes like " Min... Distance"?*

   We rephrased the sentence according to the suggestion of the reviewer:

   . . . Conversely, the "Correlative distance" (see Table 1) is less affectes. . .

6. *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.*

   We emphasized this aspect by rephrasing a sentence in the introduction:

   In the present paper we propose a different approach to HC, in which the classifier is built on actual measured radar data and is not constrained by the output of numerical simulations.

7. *8472, 13-25: this is repetitive and many sentences could be combined.*

   This section has been largely rephrased, taking into account the suggestions of the three reviewers.

8. *8473, 18: attenuation is always in dBZ/km. . . not dBZ/km.*

   We believe that the reviewers is suggesting here to express the attenuation in $dB\text{km}^{-1}$. We modified the sentence as follows:

   . . . specific horizontal attenuation α_{H} $[dB\,km^{-1}]$, and differential attenuation α_{DR} $[dB\,km^{-1}]$. . .

9. *8474, 3: I think this should be radar range gates potentially contaminate. . . 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.*

   For the sake of clarity, we substituted the term "resolution volumes", with the term "range gates" in this specific sentence and in many other cases in the revised manuscript. However, when in the manuscript we use the term "radar resolution volume", we have in mind the convention used in Bringi and Chandrasekar (2001). The radar resolution volume is the sampling volume associated with each range gate, defined by the range resolution and the angular aperture of the antenna. The terms "radar range gate" or "radar resolution volume" are synonyms in this context.

   Our guess is that the reviewer thought that we were talking about "radar volume scans" in this sentence.

10. *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?*

    This is an interesting question. Before to clarify this aspect we would like to underline that the temperature is used in the method only to get a first order separation between observations taken at positive or negative temperatures. Temperature is not used in absolute terms but in a semi-binary way (value 0 for $T > 0°$, value 1 for $T < 0°$, see Sec. 4.1) precisely because we are aware that the temperature information is precious but "external", i.e. not measured by the radar itself.

    We do not consider this so much as a "touch criteria". This simplified information could come from different sources like soundings, models, local climatology or, in ideal cases, from radar data. We mention these possibilities in Sec 4.1:

    It ($z_{0°}$) could also be estimated from other sources like soundings, numerical models or radar data directly, when a melting layer is sampled.

    We come back to the question of the reviewer. For the data of Davos there were 13 temperature measurements available (precise locations given in Schneebeli et al., 2013) within 20km from the radar. During HyMeX 30 temperature measurements within 40 km from the radar were available. The HyMeX data are accessible (after registration) on the HyMeX database (`http://mistrals.sedoo.fr/?editDatsId=627&datsId=627&project_name=HyMeX`). We rephrased the sentences describing the use of weather stations (Sec 4.1) as follows:

    The altitude of the 0°C isotherm $z_{0°}$ is approximated by means of linear interpolation of ground-based temperature measurements collected at a distance ≤ 40 km from the radar location and by assuming a constant lapse rate with altitude. It could also be estimated from other sources like soundings, numerical models or radar data directly, when a melting layer is sampled.

11. *8475, 9-10: Great idea!!!*

    Thanks!

12. *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..."*

    We agree with the suggestion of the reviewer, and we rephrased the sentence accordingly.

13. *8476, 4 - you most likely mean that the bounds were allowed for each polarimetric variable, not assigned to?*

    We rephrased this part of the sentence according to the suggestion of the reviewer:

    ... is a minimum (maximum) bound allowed for each polarimetric variable.

14. *8477, 13-17: rewrite for clarity and split these sentences up*

    The sentence has been rephrased as:

    This amount, consisting of 20 000 observations $\mathbf{x}^*$ (defined in Eqs. 6, 8, and 9), is used as input of the subsequent cluster analysis. Different seeds of the initial random selection led to the same results, suggesting that the random sampling does not affect the outcome of the clustering technique presented in the next section.

15. *8478, 18: "We define now a local... is poor word choice, awkward.*

    We rephrased the sentence as follows:

    We start by defining a spatial smoothness index (SSI) associated with...

16. *8480, 3: what do you mean by global? Throughout the radar resolution volume? How is "global" applicable to this radar algorithm?*

    The term "global smoothness" has been substituted in this sentence by "spatial smoothness" . We originally used the term "global smoothness" to indicate the spatial behaviour of n clusters in the whole dataset, while the term "local smoothness" to indicate the spatial behaviour of each cluster over individual range gates. In other words the "global smoothness" is a way of summarizing many individual "local smoothnesses".

17. *8481, 21-25: rewrite, these sentences don't make very much sense.*

    We rephrased fully these sentences as follows:

    An important step of hierarchical clustering is the selection of the optimal partition (n_{opt}) of the dataset. In the present section we introduce some indices that evaluate the quality of data partitions and that are used to guide the final selection of n_{opt}.

18. *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.*

    We agree with the suggestion of the reviewer and the manuscript has been revised accordingly.

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

    Section 5.1 was very dense, and too many indices were defined. In the revised manuscript we define only 3 indices in this section. Two of them, Kappa and accuracy spread AS, quantify the spatial smoothness of the partitions of the dataset. The third one, the SD index (introduced in the revised version of the manuscript), quantifies the compactness and separability of the content of the clusters. We select these indices because they are conceptually easier to understand and because they can be applied in relative terms (the minimum of SD and AS corresponds to optimal partitions).

    The indices defined in section 5.1 are used immediately in the following section (Sec 5.2), that starts with this sentence:

    Figure 3 illustrates the behaviour of the quality indices defined in Sec 5.1 as a function of the number of clusters in the dataset...

20. *8484, 4-5: rewrite, these sentences don't make sense.*

    The sentence has been removed from the revised version. The first sentence of Sec 5.2 in the revised manuscript has been reported in the previous point.

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

    Section 5.2 has been completely rephrased (see also the two previous points).

22. *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*

    This section was expanded and rephrased a lot. We also included in the manuscript a reminder of what Δz represents:

    ...where the relative altitude with respect to the local 0°C (i.e. Δz...

23. *8485, 19: could clarify that this is the linear reflectivity*

A clarification has been added:

...where $\zeta_{\mathrm{H}} = 10^{0.1 Z_{\mathrm{H}}}$ $[mm^6\, m^{-3}]$ (i.e. the horizontal reflectivity expressed in linear units).

24. *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".*

We agree with the suggestion and we introduced the sentence:

We conclude that results from the proposed method agree well with DR2009 for liquid phase hydrometeor classes.

25. *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.*

We believe that the reviewer means: 8486, 10. We added the following sentence:

The transition from LR to HR within few kilometres appears qualitatively to be a satisfactory illustration of the incoming front.

26. *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 K_{dp} was not useful in melting layer detection algorithms. Your results suggest the same thing - K_{dp} is too noisy to be discriminatory or trustworthy in classification of melting snow.*

We thank the reviewer for the interesting (and relevant) literature suggestions. We rephrased the sentence in the following way:

K_{dp} does not exhibit any obvious signature in the regions classified as MS (in agreement with observations documented by Thompson et al., 2014).

27. *8486, 21: why or how is it more difficult to interpret?*

This sentence has been removed from the revised version of the manuscript.

28. *8475, 6: K_{dp} of 25 deg/km is ridiculous! This cannot be right, or it must be from a bogus simulation!*

The reviewer is right about this sentence. In fact, most of the available literature presents K_{dp} values ranging between -1 and 15° km^{-1}. Our previous choice (upper boundary of 25° km^{-1}) was driven by two considerations:

(a) Figure 1c of Dolan and Rutledge (2009) was showing K_{dp} values up to 25° km^{-1}.

(b) We have observed in radar measurements collected with the same X-band radar of the manuscript K_{dp} values larger than 20° km^{-1} in rare cases. We provide here (for the reviewers and for any interested reader) one example of such measurements, collected during HyMeX, on the 24*th* September 2012. Figure 1, at the end of this document, shows a profile of K_{dp} reaching values larger than 20°km^{-1} in the most intense part of a storm.

However, to be consistent with the large majority of the peer-reviewed literature available, we modified the sentence of Sec 4.1 as follows:

...at X-band, K_{dp} ranges approximately from -1 to 15° km^{-1} (e.g. Otto and Russchenberg, 2011; Schneebeli and Berne, 2012)

29. *8476, 6: what does 3log mean? this doesn't mean anything in the way it is currently written*

    For the sake of clarity, we rephrased the sentence as follows:

    . . . -5.3 to 3 for the logarithmically transformed K_{dp},

30. *8487, 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.*

    We revised this part of the manuscript while paying attention to use real nouns.

31. *8487, 14-15: rewrite, this doesn't make sense*

    The full section has been largely rephrased, and the sentence now reads:

    Z_{DR}, shown in panel (b), exhibits a different pattern. The cyan cluster and the pink cluster show some variability in Z_{DR}. Z_{DR} ranges from -0.3 to 2.5 dB (mode 0.8 dB) in the cyan cluster and from -0.3 to 1.6 dB (mode 0.5 dB) in the pink cluster. We interpret this behaviour as the signature of particle shape and orientation variability in the cyan and pink clusters, with the pink cluster containing on average hydrometeors that are more geometrically isotropic.

32. *8487, 22-25: K_{dp} 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." K_{dp} ONLY increases beyond zero for oblate (nonspherical) particles, so it does not matter how many particles you have, only how many are oblate.*

    We are thankful to the reviewer for this suggestion. The sentence is now rephrased accordingly.

33. *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*

    We rephrased the sentences according to the suggestion of the reviewer.

34. *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?*

    We are using the second generation 2DVD (not the compact and lighter "third generation" one). We believe that Kruger and Krajewski (2002) is in any case the best available reference to describe this instrument. We clarify our 2DVD version in a new section of the manuscript (Sect. 3.2):

    One 2DVD (second generation, "low" profile version) was deployed, during the Davos field campaign. . .

35. *8488, 22-23: You need to specify here that 2DVDs capture SURFACE precipitation or images of surface precipitation particles.*

    We added a short section (Sec 3.2 in the revised manuscript) that introduces the 2DVD. Here we specify that:

    The 2DVD is a disdrometer that measure surface precipitation with a sampling area of about 125 cm^2. It provides the fall velocity as well as. . .

36. *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?*

    We clarified the main features of the algorithm performing the classification (Grazioli et al., 2014). We rephrased the beginning of Sec 6.4.3, as:

An additional comparison is conducted with the output of a HC scheme developed for two-dimensional video disdrometers. This method, hereafter called HC2DVD is described detail in Grazioli et al. (2014). HC2DVD takes as input the set of two-dimensional particles images, collected by a 2DVD, and it provides as output an estimate of the dominant hydrometoer type within time intervals of 60s. The method does not classify individual particles but populations of hydrometeors. HC2DVD discriminates between eight hydrometeor classes: small particle-like (SP), dendrite-like (D), column-like (C), graupel-like (G), rimed particle-like (RIM), aggregate-like (AG), melting snow-like (MS), and rain (R)[1]. The "-like" is added to underline that this algorithm assigns a dominant type of hydrometeor to each time step but there is usually a mixture of different hydrometeors captured by the 2DVD so they do not necessarily exhibit a single pristine shape.

37. *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?*

    Refer please to the previous point.

38. *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?*

    We rephrased the section avoiding capital letters for hydrometeor types. "RIM" is the acronym used in Grazioli et al. (2014) for the 2DVD-based classifier called HC2DVD in the paper. In the revised manuscript we clarify with a footnote that the notation is taken verbatim from that manuscript.

39. *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.*

    The sentences have been rephrased as follows:

    this algorithm assigns a dominant type of hydrometeor to each time step but there is usually a mixture of different hydrometeors captured by the 2DVD so they do not necessarily exhibit a single pristine shape.

40. *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.*

    The radar, during this field campaign, was located at an altitude of ≈2100 m, while the 2DVD was deployed on the other side of the valley of Davos, at a higher altitude of ≈2500m. The horizontal (projected) distance between the 2DVD and the radar was about 5.2 km. Therefore, the altitude difference between the radar beam and the 2DVD is:

    $$\Delta \approx 2100 + 5200 \times \tan(9 \times \pi/180) - 2500 = 423\text{m}$$

    Additionally, if we consider that the beam of the radar is broadening with distance (with a 3dB beamwidth of 1.45°, as given in Table 2 of the manuscript), at a horizontal distance of 5.2 km the beam is about 200 m wide, and therefore the bottom of the beam is only ≈ 320 m higher than the location of the 2DVD. In order to clarify the manuscript, we added the altitude of the 2DVD in Figure 1.

    We are aware that even if the actual difference of altitude is ≈400 m, the crystal habits can change along this space. We clearly state that this is a potential limitation of the comparison (Sec 6.4.3):

[1] We use here the abbreviations of Grazioli et al. (2014)

...the closest radar resolution volume center was about 400 m above the 2DVD and crystal habits can change over this altitude range...

41. *8489, 13: I would clarify that the sampling times are different but the do overlap.*

    We clarified the sentence as follows:

    ...the sampling times and volumes of the two instruments are different, even though the sampling times overlap.

42. *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?*

    The range gate spacing of MXPol is actually 75m. This operation is conducted in Cartesian-projected data and this results most of the times in 9 radar pixels (consider ±150 m, centred at the location of the 2DVD).

43. *8489, 23: should clarify that the clusters are "similar between methods". not just "similar".*

    This part of the manuscript has been largely rephrased, and the sentence does not appear any more.

44. *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.*

    This part of the manuscript has changed. We now group the classes of HC2DVD into three macro-classes that should be directly comparable with our algorithm. Additionally, we solved a Z_{DR} zeroing issue that was influencing these statistics.

45. *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.*

    This part of the manuscript has been largely modified and the sentence does not appear any more.

46. *8490, 6-7: I would add that this is probably happening because aggregates tumble a lot as they fall, making their Z_{DR} 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.*

    Thanks! This is a very well phrased explanation. We removed the sentence from this part of the manuscript but we used this interpretation in a previous section. In section 6.4.1 we mention:

    Aggregates generate larger Z_H returns due to their larger sizes, and they tumble as they fall thus lowering Z_{DR}.

47. *8490, 18: You need to clarify how D_e 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 D_e accuracy, as you mention. Since you mentioned the uncertainty, you should explain more about it.*

    We believe that the reviewer has in mind the median volume diameter D_0 because he/she refers to particle size distributions. D_e is instead the equivalent "spherical" diameter, i.e. the diameter of a sphere having the same volume of the sampled particle. This parameter is classically used to describe raindrops (rotationally symmetric hydrometeors with well known size-shape relationships), and it was extended to describe snowflakes by Huang et al. (2010). For snowflakes, D_e is estimated from the apparent bulk dimensions of the particles as they appear in the two cameras of the 2DVD. Anyway, snowflakes are not in general rotationally symmetric and therefore D_e has to be considered only as an index describing the characteristic size of particles of irregular shape.

We included in the text the appropriate reference:

De_i is the equivalent diameter of each snowflake or particle [mm], as defined in Huang et al. (2010).

48. *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*

    We rephrased this part of the manuscript.

49. *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 (1997) Microphysics of Clouds and Precipitation textbook.*

    We thank the reviewer for the suggestion. These concepts (with the Pruppacher and Klett (1997) reference) have been included in the text. The sentence is now rephrased as:

    This is the expected behaviour of rimed particles. As riming progresses, the original shape of ice particles becomes imperceptible, the drag decreases and the particles become more dense (Pruppacher and Klett, 1997). Hence, their fall behaviour is smoother and therefore faster, leading to larger EF.

50. *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.*

    These sentences have been removed. We preferred to move this discussion in the previous sections and paragraphs, that have been expanded.

51. *8491, 9-11: descriptions of heights in parentheses is really confusing, the last one should just be "within 3.5 km of the radar"*

    We agree with the reviewer. The description of heights was confusing. We rephrased the sentence by using mathematical notation to describe the heights (y dimension in the RHI figure). Additionally, we use now always absolute heights and not relative heights with respect to the radar. We rephrases as follows:

    At higher altitudes ($y > 5\,\mathrm{km}$) there is a thin layer classified mostly as crystals (CR). The crystals turn into aggregates (AG), that dominates the precipitation in a second layer ($4 \leq y \leq 5\,\mathrm{km}$), and finally ($y \leq 4\,\mathrm{km}$) RI dominates the precipitation.

52. *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!*

    The sentence has been rephrased as follows:

    COSMO predicts the presence of supercooled liquid water (QC) at altitudes between 2.5 and 3 km. Additionally, at altitudes between 2 and 6 km, we observe large quantities of graupel (QG) mixed with snow (QS).

    We appreciate the positive comment of the reviewer about the use of a NWPM for direct comparison. Even though the comparison is rather qualitative at this stage, we believe that it illustrates nicely the complementarity of radar measurements and model reanalysis.

53. *8492, 14: should be "with dual-polarization radar data"*

    We rephrased the sentence as follows:

Finally, three clusters were found at negative temperatures and their polarimetric signatures were interpreted as the ones of crystals/small aggregates (CR), aggregates (AG) and rimed ice particles (RI).

54. *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.*

    We rephrased the conclusion. In the revised manuscript they are mostly in past tenses.

55. *8492, 3-4: what do you mean by "reduced" and what do you mean by "representative subset"???*

    This sentence was summarizing the first step of the proposed approach, described in Sec 4.2. We propose now the following rephrasing:

    A subset of 20000 polarimetric observations was randomly extracted from the available dataset.

56. *8492, 14: if the abbreviation is RI, you shouldn't expand it as rimed particles... where's the i? Should say rimed ice particles*

    We agree, here and in the rest of the manuscript RI is now defined as "rimed ice particles".

57. *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?*

    With the term "systematic bias" we wanted to define any offset that might affect one of the polarimetric variables systematically (always and constantly). We agree that the sentence is confusing and therefore we rephrased it according to the suggestion of the reviewer:

    Some of the advantages of this approach are that it is immune to possible radar miscalibration…

58. *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.*

    We agree. This important aspect is emphasized in the revised text with the sentence suggested by the reviewer.

59. *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 K_{dp} is inversely proportional to wavelength, so you will DEFINITELY see differences in your clusters using K_{dp} at X, C, or S band.*

    As the reviewer points out, our statement was too strong especially with respect to different radar frequencies. We rephrased the sentence as follows:

    We nevertheless expect the number and type of clusters to be very similar for other X-band dual-polarization radars of similar sensitivity.

    About the calibration: the calibration constant of our radar was validated in Davos by means of a corner reflector, and during the HyMeX campaign by comparing radar and disdrometer measurements (in terms of Z_H).

60. *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.*

The spatial constraint employed in the manuscript is very simple. It is based on the assumption that on average hydrometeors of a certain type tend to have as neighbours hydrometeors of the same type. We agree with the reviewer about the added-value of prior physical knowledge for the clustering, but this is beyond the scope of this manuscript, which is focussed on presenting an innovative unsupervised clustering approach. The development of a supervised (or semi-supervised) clustering approach taking advantage of prior knowledge will be conducted in future work. We rephrased the sentences at the end of Sec 7, according to the suggestion of the reviewer:

It is interesting to note that the method exploits a simple hypothesis of spatial smoothness of the hydrometeor types and that this rule is applied only in the initial steps (when the n_{opt} clusters are identified). Future work will be devoted to extend the constraints involving spatial smoothness also to newly classified images, or to include physically-justified contiguity rules for specific hydrometeor types.

61. *8493, 6-9: True! Good point, this is a strength of your results/algorithm.*

    Thanks!

62. *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!*

    We clarified this aspect with the following sentences in the section (Appendix B):

    - The algorithm denoted as DR2009 in this paper is based on the work of Dolan and Rutledge (2009), with some adaptations that we will highlight as they appear in this section.
    - Δz is the relative altitude with respect to the 0°C isotherm, as defined in Sec 4.1, and this input is not used in Dolan and Rutledge (2009).
    - ... vertical ice (VI) and wet snow (WS, not present in Dolan and Rutledge, 2009).

63. *Table 2 caption: should be "dual-polarization radar"*

    We changed the caption according to the suggestion of the reviewer.

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

    We changed the caption by adding the percentage sign after each quantile, as follows:

    Some relevant quantiles (Q5%, Q25%, Q50%, Q75%, Q95%) of the full distribution are given here.

    A small clarification for the reviewer. We talk in the manuscript about general quaNTiles, while the term quaRTiles should be used only for Q25%, Q50%, Q75%, that divide a distribution (or a sampled population) in quarters.

65. *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.*

    The caption now includes these definitions. It is rephrased as:

    Parameters of the membership beta functions β employed in the DR2009 algorithm: midpoint m, width a and slope b for the available hydrometeor classes.

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

    The caption is rephrased as:

    Parameters of the trapezoidal membership function T applied to the relative altitude with respect to the 0°C isotherm (Δz [m]). l_1, l_2, l_3, l_4 are the four vertices of the trapezoid T.

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

    Fig 2 has been revised according to the suggestion of the reviewer.

68. *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*

    This figure was referenced in the end of Sec 4.3.2, and it was used to illustrate the fact that raw radar data are collected in polar coordinates, and that the constraint of spatial smoothness is applied in polar coordinates too. Because the length of the manuscript is increasing after taking into account the suggestions of the reviewers, we decided to remove the figure from the revised version. We rephrased instead the end of Sec 4.3.2 as follows:

    The identification of the nearest neighbours is performed in polar coordinates and therefore the distance between objects is the distance between their respective radar resolution volumes.

69. *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*

    The yellow line has been added and this graph has been revised and simplified. The caption is rephrased according to the suggestion of the reviewer as:

    Evolution of Kappa, accuracy spread (AS) index, and SD index as a function of the number of clusters in the dataset. The SD index is stretched between 0 and 1 for illustration purposes. The yellow vertical line at $n_c = 7$ shows the selected final number of clusters, corresponding to a minimum AS and SD. Each curve shows the mean behaviour over 100 runs of the clustering algorithm.

70. *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?*

    We added a legend and produced better resolution figures. Additionally, we produced this image now using points from 100 realizations (only one realization was shown in the previous manuscript) of the clustering algorithm. This serves two purposes: (i) the density of points per cluster shown is higher (the figure looks therefore better) and (ii) we show that the method is robust. We also rephrased the caption to clarify the content of the graph.

71. *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!*

    We substituted this figure with a contingency table, keeping in mind the suggestion of the reviewer for the caption.

72. *Fig 7: PLEASE add variable labels to each subplot just saying "Z_{DR}", "K_{dp}", 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.*

    We agree with the reviewer and we applied his/her suggestions.

73. *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 "Z_{DR}",*

"K_{dp}", "Z_H", 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.

We applied the suggestions of the reviewer.

74. *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?*

    We substituted this figure with a contingency table.

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

    Also this figure has been revised according to the suggestions of the reviewer.

76. *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.*

    The caption is rephrased as follows:

    Sampled probability density function of snowfall intensity, as quantified by the equivalent flux EF $[mm\,h^{-1}]$, measured by the 2DVD and associated with the radar hydrometeor classification output above the location of the 2DVD. The data used in this figure were collected during the field deployment of Davos (CH).

    The CDF is not shown any more, to avoid too many lines on the same graph.

77. *Fig. 13: See all comments for 7*

    The figure has been revised according to the suggestions of the reviewers.

78. *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.*

    This figure has been revised.

    (a) The domain has been "zoomed" to a maximum distance of 20 km from the radar.
    (b) Lines are thicker.
    (c) The color of graupel has been changed to green.

Other systematic errors

OTHER systematic things I noticed that could be improved:

1. *reference to your figure panels in text: panel (a) looks and reads better than panel a - see 8478, 4*

   We agree with the reviewer and we implemented the following changes through the manuscript (e.g):

   - Fig 13a → Fig 13 (a)
   - panel b → panel (b)
   - (Fig 13a)→ (Fig 13a) No changes

2. *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...*

   We corrected it in all the manuscript, and we removed also many unnecessary "i.e.".

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

   We thank the reviewer for all the grammatical clarifications and suggestions. We revised the manuscript keeping in mind this and the following remarks. As a last step, a native English speaker proof-reads the manuscript.

4. *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).*

   We thank the reviewer and we paid attention to this aspect during the revision process.

These should be revised throughout the manuscript.

Anonymous reviewer 2

This paper presents an exceptional work. It describes a novel approach of a hydrometeor classification using polarimetric data. It is a "a-prioriless" method where the hydrometeor types are not defined in the beginning. 7 common behaviors or clusters have been detected where each one has been relied to a hydrometeor type at the end. The results are compared to a classical hydrometeor classification algorithm and to a 2DVD classification output.

We kindly thank the reviewer for his/her very positive feedback and for the recognition of the originality and novelty of our approach. In the following we will address one by one his/her valuable suggestions and comments.

My comments starts with some general comments or ideas and then by some minor remarks:

General comments

1. *The authors mentioned at the beginning that 3000h have been used. Then filtered with a SNR threshold and ground clutter contamination. What is the number of hours used at the end.*

   We realized that mentioning "3000 hours" was generating confusion. 3000 hours was in fact the total amount of relevant (i.e. not "dry") observations collected during the two field campaigns described in Sec 3.1. Each run of the unsupervised part of the algorithm takes as input 20000 measurement points, i.e. 20000 "range gate" observations. We tried to clarify the concept by rephrasing some parts of the manuscript. At first in the abstract:

   The approach is applied to data collected by an X-band polarimetric weather radar during two field campaigns (from which about 50 precipitation events are used in the present study).

   Then, in Sect. 4.2:

   About 50 precipitation events belonging to the data set of Sect. 3.1 were manually selected.

   Finally, in Sect. 4.2:

   A subset of data is taken randomly from these 50 precipitation events, from PPI scans conducted at elevation angles between 3.5° and 10° (free of ground clutter contamination). This amount, consisting of 20 000 observations...

   its ok to not correct in solid precipitation cases but what about attenuation with liquid precipitation.

   We do correct for attenuation both Z_H and Z_{DR} in liquid precipitation, using Φ_{dp}. This is mentioned in Sect. 3.2, with the sentence:

   Z_{H} and Z_{DR} collected in rain are corrected for attenuation using the relations linking K_{dp}, Z_{H}, specific horizontal attenuation α_{H} $[dB\, km^{-1}]$, and differential attenuation α_{DR} $[dB\, km^{-1}]$ according to the method of Testud et al. (2000)

   Does the authors planned some sensitivity study?

   This is an important question. We realized that in the old manuscript the robustness of the method was not mentioned. In order to show that the method is robust, now we perform the selection of $n_{opt} = 7$, in section 5.2, by using 100 realizations of the clustering output instead of a single one. Also, we now use 100 runs to produce Fig 4, 5, 9 (numbering of the old manuscript), to show that not only $n_{opt} = 7$ is stable, but also the content of the clusters.

   We are aware that there are many choices in the manuscript. Some examples are the standardization of the data presented in section 4.1, or the random sampling of section 4.2. We verified the robustness of the algorithm with respect to these choices but we preferred not to present them in the manuscript that (already extremely dense of contents). Instead, we introduced in the text sentences like (Sec 4.1):

   Variations of the order of ±20 % around the proposed boundaries have a negligible impact on the results presented in the following sections, and the most sensitive boundaries are associated with Z_H .

   ... Gaussian, sigmoid and logistic functions have been tested and appeared to be equally adequate...

   Then, in Sec 4.2:

   Different seeds of the initial random selection were leading to the same results, suggesting that the random sampling does not affect the outcome of the clustering technique presented in the next section...

Finally, in Sect. 4.3:

Among the possible combinations of distance metric and linkage criteria presented in Sect. 2 we obtained similar results also using the correlative distance and/or WC linkage.

We evaluated these statements even though we did not show the results. As an example for the reviewer, figure 3 (end of the present document) illustrates the robustness of the classification output as a function of a bias on the standardization extremes of Sect 4.1 (left panel), and as a function of a bias on the polarimetric variables ($+\Delta z$) themselves. We observe that variations on the order of $\pm 20\%$ result in less that 15% of changes in the classification output (the error is mitigated and not amplified).

2. *What about calculation time? can this method used in operational environment?*

   We thank the reviewer for the question, that allow us to clarify this aspect. Our approach to hydrometeor classification is composed of two basic steps:

   (a) Clustering, i.e. identification and interpretation of the n_{opt} hydrometeor classes. This step is computationally expensive and slow (the complexity of agglomerative clustering alone is $\approx O(n^3)$, and we include additionally the spatial consistency constraint). However, this step is conducted only once to obtain the n_{opt} clusters.

   (b) Classification of new radar images. New radar images are classified into one of the n_{opt} already available classes. This is the step that would need to be run in an operational context. This step is fast (i.e. a few seconds for a PPI image with precipitation) and can be even made faster by means of lookup tables.

   We clarify this aspect in the conclusions of the revised manuscript. We mention that:

   The initial identification of the clusters is computationally expensive, but this operation is performed only once and the classification of newly collected radar images can be conducted in real time.

3. *At the beginning, the paper is really interesting. After when the equations start, it becomes really hard to follow.*

   We rephrased the manuscript according to the suggestion of the reviewers and we believe that now it is easier to read and understand. We reduced the length of the initial sections in favour of the final discussion. About the equations in particular: this manuscript is presenting an approach that is new to the weather radar research community. We are convinced that the mathematical details, even though they make the paper "heavier", are fundamental to ensure that future studies can take this work as a starting point and use similar techniques for hydrometeor classification or other related topics.

   In my opinion, the Appendix B is not really important. I suggest, instead, to add a numerical example which better explains better the concept

   We realized that, after the revision, the manuscript size was increasing and we thought about removing an appendix. However, we decided to remove Appendix A, which was generating confusion to the reviewers, and not Appendix B. Appendix B describes our implementation of the DR2009 algorithm (Dolan and Rutledge, 2009). We think that this information is important because of two reasons:

   - It underlines some modifications that we implemented to Dolan and Rutledge (2009). It is therefore fair to the authors of the method to clearly state these differences.
   - With some exceptions (e.g. Marzano et al., 2010; Al-Sakka et al., 2013), it is really difficult to find examples of full parametrizations of hydrometeor classification schemes in the literature. Therefore we find this information relevant for the scientific community.

4. *this paper missed some statistical or scores to show better the performance of the approach vs others*

   Following the suggestion of the reviewer we added in the revised manuscript a quantitative evaluation that, with the 2DVD as a reference, shows that our algorithm achieves better scores than DR2009 in the classification of winter events collected in Davos (see the confusion matrices of the revised manuscript, and related scores). This statistical evaluation is an innovative aspect of the revised manuscript, with respect to the available literature on the subject. However, a thorough quantitative evaluation among classification techniques is by itself a very innovative and difficult topic and we stress all the possible limitations of such exercises in the manuscript.

This analysis does not aim at demonstrating that our technique is better, but rather that this innovative approach is promising.

5. *Some sentence are long and commas are missed.*

   We deeply rephrased the manuscript according to the suggestions of the reviewers and with the help of a native English speaker.

Different remarks

1. *8473, 4 and 11 : can the authors add the total scanning strategy including the time?*

   We rephrased both sentences in order to include this information in the paragraph. This information is in our opinion of particular interest for the HyMex data, that are available in an online database (`http://mistrals.sedoo.fr/HyMeX/`). This clarification can be useful for any scientist potentially interested in the data set.

   The description of the scanning sequence for the "Davos" campaign has been rephrased as:

   The scanning sequence of the radar, repeated approximately every 5 minutes, included Plan Position Indicator (PPI) sector scans over the valley of Davos (at elevation angles of 0°, 2°, 5°, 9°, 14°, 18°, 20°, and 27°), a Range Height Indicator (RHI), and a vertically-pointing PPI used for the zeroing of Z_{DR}.

   The description of the scanning sequence for the "HyMeX" campaign has been rephrased as:

   The scanning sequence of the radar included in this case wider (200° in azimuth) sector scans at elevation angles of 3.5°, 4°, 6°, 9°, and 10°. Additionally, two or three RHIs towards different directions and a vertically-pointing PPI were collected during each cycle of 5 minutes.

2. *8473, 18 : the unit is dB km-1 and not dBZ km-1*

   The units have been corrected.

3. *8473, 25 : can the authors add some references?*

   A reference is added, citing the work of Matrosov (1992).

4. *8474, 16 : what is the degree of confidence about the Z0 and what is the error margin?*

   We thank the reviewer for this interesting question. We calculated the standard deviation of z_{0° estimates coming from the different temperature measurement stations (... at a distance $\leq$ 40 km from the radar location..., Sect. 4.1) and we obtained a value of $\approx$ 0.4 km, which is the reason why we use $\pm$400 as a boundary in Eq. (10). The reviewer make us realize that this was not written in the manuscript. Therefore we rephrased the explanation of Eq. (10) by including the following sentence:

   The threshold of $\pm$400 m is the (rounded) standard deviation of z_{0° estimates.

5. *8475, 6 : 25 deg/km !!! I think it is obvious. The intervals should be more realistic.*

   The reviewer is right about this sentence. In fact, most of the available literature shows K_{dp} values ranging between -1 and 15° km^{-1}. We changed the boundaries in the revised manuscript. For additional information on why we originally selected a value of 15° km^{-1}, please refer to point 28 of "Additional questions and corrections" of our answers to reviewer 1.

6. *Table 2 : I prefer to see more details about the scanning strategy.*

   We respectfully disagree with the reviewer in this case. This table gives only the general and fixed characteristics of the radar system. The scanning sequence instead is modified and adapted to cover the specific needs of different field campaign, and therefore we do not consider it as a general characteristic of the radar. In any case we agree with the reviewer about the need of additional information about the scanning sequence and we added this information in Sec 3.1 (see also point 1 of "Different remarks", above).

7. *Figure 1 : couldn't be better if the background is the topography instead of the google view?*

   We tried this option (with available open-source topographies), but we still believe that the map view gives a better visual impression.

8. *Figure 5 : the light blue color is not really shown.*

   The figure has been largely revised. Additionally, the reader is warned of the fact that this is a 2-D projection of a 5-D space and some clusters result slightly hidden. The caption includes the following warning:

   Note that some of the clusters are not fully visible in this 2-D projections, because they are defined in a 5-D space

Reviewer 3: Dr Earle Williams

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.

We would like to thank Dr Williams for his great contribution to the manuscript. In the following section we will answer to the substantive issues of the reviewer and to his technical corrections.

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.

   We agree with the reviewer and we have clarified the source of the data in the different parts of the manuscript. About Section 3.1, our message is slightly different with respect to what understood by the reviewer. The "Davos database" contains in fact many cases of liquid precipitation (the radar was deployed at ≈2100 m of altitude so during summer months rainfall was often measured), even though its main peculiarity (uniqueness) is the large amount of snowfall cases sampled in fall, winter, and spring. We clarified this aspect by rephrasing as follows the first part of Sec 3.1, describing the "Davos" campaign:

   The altitude of the deployment site made it possible, during cold seasons, to collect many observations of ice-phase precipitation when the radar itself was located above the melting layer, and therefore did not suffer from liquid-water signal attenuation. Such radar observations represent the main peculiarity of this field campaign (e.g. Schneebeli et al., 2013; Scipion et al., 2013). However, during warm seasons, the melting layer was often higher than the radar site and relevant observations of liquid phase precipitation, both in stratiform and convective cases, were collected as well.

   Later in the section, when describing the "HyMeX" observations, we make clear with the following sentence that abundant liquid phase precipitation was sampled:

   Stratiform and convective Mediterranean precipitation events were sampled during this campaign, with the radar always located below the melting layer.

   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.

   We modified the captions of tables and figures in order to clarify the source of the data. We used sentences as:

   The data used in this figure were collected during the deployment of Davos (CH).

   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.)

   The term " Ardèche" is now used in the caption instead of "Montbrun".

   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?*

Where possible we tried to add some information about the weather in Davos and in the south of France as well as some information about lightning. In Sect. 3.1 we introduce the climate of Davos campaign as:

The climate of the Davos region is characterized by approximately 130 days of precipitation per year, and total yearly accumulations of about 1100 mm. The most intense winter snowfall events are associated in this context to air masses coming from the North and North-East (Mott et al., 2014).

About the HyMeX campaign, conducted in Ardèche (FR), we clarify that strong thunderstorms were observed, together with more mild stratiform rain. Also, we add some synoptic scale information about the weather. We quote from the revised manuscript:

This deployment was part of the HyMeX experiment (`www.hymex.org`, ?Bousquet et al., 2014). Stratiform and convective Mediterranean precipitation events were sampled during this campaign, with the radar always located below the melting layer. Convective precipitation included vigorous thunderstorms with intense electric activity. In Ardèche, precipitation (in the fall season) is mainly associated with eastward-moving troughs from the Atlantic region, that are at first slowed by the anticyclonic system over Russia and interact with the complex topography of the coastal region in the south of France (Miniscloux et al., 2001; Boudevillain et al., 2011).

However it is not possible (and relevant) to provide a thorough description of the synoptic condition for all events, given the length of the paper.

The description "30 snowfall events" for Davos could be elaborated on.

The 30 snowfall events that the reviewer quotes refer to the comparison between the 2DVD and the radar. We clarified the sentence in the following way:

The comparison is conducted on a subset of about 30 manually selected snowfall events. We excluded any precipitation event with a visible melting layer, or positive temperatures at the radar location as well as any event characterized by evident spatial and temporal variabilities at the small scale.

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.

   In the manuscript we tried to provide as input for the algorithm a database that should cover a wide range of weather events and conditions. This includes stratiform rain, convective storms, summer thunderstorms and snow storms. The algorithm then is used to find the best way to group together similar radar observations in the database and to select how many groups should be considered (seven, in our case). These groups correspond to the partition that better adapts to the radar system and to the database. We are confident that all the weather types mentioned by the reviewer are found in our data. However, as mentioned earlier, we believe that it is not relevant for the paper (given also its length) to provide a thorough description of all the cases and we decided to provide (global) additional information about the weather in Davos and in Ardèche in Sect. 3.1. Please refer to the previous point for the details.

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.

We followed the suggestion of the reviewer and we added a short section (3.2 in the revised manuscript), in order to introduce the 2DVD.

In the Figure caption, reference is made to "Parsivel-type disdrometers". If they are the same as the 2DVD, that should be clarified.

Parsivel type disdrometers are not the same as the 2DVD, the main difference being that they are not able to provide actual images of the hydrometeors crossing their measurement area. These disdrometers were used to parametrize the attenuation correction scheme for reflectivity Z_H and differential reflectivity Z_{DR} in the data collected by MXPOl during the HyMeX campaign. We rephrased a part of the caption as follows:

Red circles are used to mark the locations of instruments directly employed in the study (MXPol and a 2DVD two-dimensional video disdrometer) while blue squares are used for laser disdrometers (Parsivel) employed only to parametrize the attenuation correction of Z_H and Z_{DR}.

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.

An extensive description of the algorithm would unnecessarily lengthen the manuscript, and the interested reader should refer to Grazioli et al. (2014). However, we largely rephrased this part of the manuscript. We tried to provide to the reader the main characteristics of the HC2DVD algorithm with the following sentences (Sect. 6.4.3):

An additional comparison is conducted with the output of a HC scheme developed for two-dimensional video disdrometers. This method, hereafter called HC2DVD is described detail in Grazioli et al. (2014). HC2DVD takes as input the set of two-dimensional particles images, collected by a 2DVD, and it provides as output an estimate of the dominant hydrometoer type within time intervals of 60s. The method does not classify individual particles but populations of hydrometeors. HC2DVD discriminates between eight hydrometeor classes: small particle-like (SP), dendrite-like (D), column-like (C), graupel-like (G), rimed particle-like (RIM), aggregate-like (AG), melting snow-like (MS), and rain (R)[2]. The "-like" is added to underline that this algorithm assigns a dominant type of hydrometeor to each time step but there is usually a mixture of different hydrometeors captured by the 2DVD so they do not necessarily exhibit a single pristine shape.

4. *Vertically aligned ice*

"Vertically aligned ice" is insufficiently defined in this study.

We added a clarification, in Sect 6.4.2. The clarification reads:

. . . vertically aligned ice (VI, that denotes oblate ice crystals aligned vertically because of an electric field).

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.

We prefer not to comment the results of a manuscript that we cannot access. However, according to classical literature, we can address the reviewer to Caylor and Chandrasekar (1996); Ryzhkov and Zrnic (2007). These papers showed that K_{dp} is more illustrative than Z_{DR} in detecting strong electric fields. We would like to stress the fact that highly electrified storm conditions are very special cases, and the fact that our clustering algorithm do not detect such situations with a specific cluster is a consequence of the fact that their occurrence is limited (in our dataset).

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

[2] We use here the abbreviations of Grazioli et al. (2014)

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.

To be fair, Dolan and Rutledge (2009) do not name a single category as "conical graupel", but in Sec. 2 of their manuscript they take into account this hydrometer type in their scattering simulations and then they associate it with their graupel categories (high density graupel HDG and low density graupel LDG).

About negative Z_{DR} signature, we solved an issue that was affecting Z_{DR} zeroing during some periods of the Davos campaign. This improvement, even though did not have a major impact on the global statistics of the clusters, is leading to a better outline of the ice-phase clusters (notably, we see now negative Z_{DR} in the rimed ice particles, RI) and it improved the comparison with the 2DVD observations as well.

5. *Figure 5 (and Figure 8) pertaining to K_{dp}*

   In light of evidence cited above for negative K_{dp} values, together with the authors own evidence for vertically aligned ice, it certainly seems appropriate to include a negative axis for K_{dp} in Figure 5. The big pileup of values near zero is surely suggestive that negative values also appear in the data set.

   We agree with the reviewer about negative K_{dp}. Now the method includes negative K_{dp} values as an input. The differences are not major. Negative K_{dp} values did not cluster together in a new hydrometeor class but they distributed among the other classes. However, in this way we do not artificially bound K_{dp} and we obtain more reliable statistics on the content of each cluster.

   The text on page 8486 reports that "K_{dp} does not seem to play a particular role in the classification of MS", but that needs to be shown.

   This comment was referring to a single image, qualitatively inspected. We rephrased the sentence (and added an appropriate reference) as follows:

   K_{dp} does not exhibit any obvious signature in the regions classified as MS (in agreement with observations documented by Thompson et al., 2014).

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

   We now show negative K_{dp} values in the RHI images and we show the distribution of K_{dp} values for all the clusters. (i) We added a summary table in the appendix with the main statistics of each cluster (Table A1 of the revised manuscript). (ii) We added a figure (as Figure 9 of the previous manuscript but for clusters appearing at positive temperatures). This figure is the number 5 in the revised manuscript.

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 reviewer has asked some good questions. We add here two small premises. At first, in our opinion, the questions are valid for Figure 5 and Figure 9 (numbering of the old manuscript), that show polarimetric data only, while Figure 12 is based on the comparison with the 2DVD and does not include polarimetric data (EF is calculated only from 2DVD data, and then stratified according to the classification output of the radar in the proximity of the instrument.) Secondly, we believe that the reviewer is here tackling two aspects/issues at the same time and we will answer separately in the following paragraphs.

   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.

   Issue 1, i.e. the classification problem. Mapping continuous variations into a discrete set of classes.

We agree with the fact that different hydrometeor classes can exhibit similar polarimetric signatures, this is a well known fact and this is the reason why classical hydrometeor classification schemes are performed by means of fuzzy logic (well described in Straka et al., 2000; Liu and Chandrasekar, 2000)[3]. Hydrometeor classification is mapping a set of continuously varying polarimetric observations into a discrete set of most likely hydrometeor classes, and this justifies the possibility to have overlapping classes.

Fuzzy logic, by means of a set of mono- or multi-dimensional membership functions, assigns each radar measurement to the hydrometeor class that is receiving the highest score from all the membership functions. The "overlapping issue" highlighted by the reviewer is evident when two or more classes have very similar scores, and the discrete assignment is therefore questionable. With our approach, this would mean that a set of polarimetric observations (object to be classified) is situated at a similar distance with respect to two or more clusters.

The goodness of classification can be evaluated in both cases (fuzzy logic or our approach) in a similar way, i.e. by showing not only the final output but also the scores (for the fuzzy logic) or the distances (for our approach) of each classified observation to all the hydrometeor classes. In this way the output of the hydrometeor classification would be described by many images:

(a) An image showing the classical hydrometeor map (like in the PPIs and RHIs presented in the manuscript)

(b) An image showing the reliability of the classification.

(c) $n-1$ images showing the scores (distances) of each classified point with respect to the other $n-1$ hydrometeor classes available.

This obviously complicates the visual interpretation of the results. The reviewer made us realize that such scores are rarely shown in hydrometeor classification papers and we decided to add this information in the revised manuscript. We added to each PPI and RHI image a map that shows the "classification accuracy". We use a rather restrictive index of accuracy, defined in the text as (Sect. 6.2, end):

This parameter is defined for each observation (valid range gate) as the difference between the distance of the observation with respect to the two closest clusters, and it is normalized by the smaller distance. The classification accuracy is therefore lower in the areas of transition between different hydrometeor types, where the polarimetric signatures change as the dominant hydrometeor type changes.

This index illustrates nicely that in the areas of transition between different hydrometeor types the reliability of classification decreases (as well as on the edges of the domains, where we reach the lower boundary in sensitivity of the radar) and therefore reminds to the reader that hydrometeor classification is a discrete mapping of otherwise continuously varying quantities.

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.

Issue 2. Are the overlapping clusters of Fig.5 and Fig.9 distinguishable?

Before to proceed, a very important clarification is necessary. We quote, from Sect. 6.1:

These clusters exist in the 5-dimensional space given by the dimensions of $\mathbf{x}^*$ and it is therefore not trivial to illustrate their content. A way to reproduce a partial visualization of the clusters is to display pairs of 2-dimensional projections of the objects $\mathbf{x}$, keeping anyway in mind that their original nature is 5-dimensional.

Also, in the caption of Fig.5 (numbering of the old manuscript), rephrased as:

Three examples of 2-D projection of the 7 clusters found in the dataset. The clusters include observations collected in Davos (CH) and Ardèche (FR) . . . Note that some of the clusters are not fully visible in this 2-D projections, because they are defined in a 5-D space

[3]To overcome this aspect, many methods use temperature as input and they treat it as an additional radar variable, despite the fact that absolute temperature is not measured by the radar.

The clusters exist in a 5-D space, while Figure 5 showed 2-D projections and Figure 9 showed 1-D projections. Therefore what the reviewer observes is an overlapping range of variation of polarimetric variables within each cluster. Each object shown in Fig. 5 is unambiguously assigned to one and only one cluster, according to a distance metric, a merging rule and a spatial constraint (Sect. 4.3).

However, it is instructive to follow the suggestion of the reviewer and demonstrate that the clusters are distinguishable. We can look at the objects $\mathbf{x}$ belonging to a cluster i and see what is their distance $D(\mathbf{x}, j)$ with respect to a cluster j. Then we compute the ratio $D(\mathbf{x}, j)/D(\mathbf{x}, i)$. If this ratio is systematically greater than one (for many $\mathbf{x} \in i$) then we can be confident that the clusters are actually separable.

We show this comparison in Figure 2 (at the end of the present manuscript). Panel (a) and panel (b) show the two cases highlighted by the reviewer. We can observe (panel a) that DZ (now renamed light rain, LR, in the revised manuscript) and MS are actually well separable and the distance ratio is almost never lower than 1. AG and RI (panel b) are more similar among each other. However, only a negligible proportion of distance ratios is lower than one, indicating that the two clusters are separable. As a reference, panel (c) shows the behaviour of AG and HR, two highly separable clusters.

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.

Figure 12 is based on 2DVD data, and we use it to validate our assumptions about the content of the ice-phase clusters. In this case it does not matter that the histograms overlap as long as we can see a clear ranking for these three classes, that supports our hypothesis. The comparison with the 2DVD is anyway very delicate, and we warn the reader in the end of the section:

The results presented in this section are not a rigorous validation but they are in agreement with our initial assumptions.

Additionally, the section describing the use of the 2DVD in this manuscript has been largely revised and we believe that the comparison is now described better.

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).

We agree with the reviewer on this point. Section 5.1 was very dense, and too many indices were defined. In the revised manuscript we define only 3 indices in this section. Two of them, Kappa and accuracy spread AS, quantify the spatial smoothness of the partitions of the dataset. The third one, the SD index (introduced in the revised version of the manuscript), quantifies the compactness and separability of the content of the clusters. We select these indices because they are conceptually easier to understand and because they can be applied in relative terms (the minimum of SD and AS corresponds to optimal partitions).

The indices defined in section 5.1 are used in the following section in order to find the optimal partition corresponding to $n_{opt} = 7$ clusters. Please note that in the revised version we reduced the indices to 3, but we evaluate their mean behaviour over 100 runs of the clustering algorithm, to show the robustness of this choice.

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.*

We would like to clarify some aspects, related to the question of the reviewer:

- The clusters are obtained over all the database (HyMeX+Davos), and therefore the two cases mentioned by the reviewer refer to the same available categories.
- We interpret the content of the clusters in Sec. 6.1, 6.2, and 6.3. These sections have been largely rephrased and improved, especially with respect to the interpretation of polarimetric measurements. Again, the content of the clusters comes both from Davos and from Ardèche (HyMeX) measurements.
- We do not claim to be able to identify supercooled liquid water, but only its major effects on the hydrometeors. It may happen that in the areas classified as "rimed ice particles" (RI) from our algorithm there is actually no supercooled liquid water, but the supercooled water was present where these particles formed. An example of our interpretation of RI can be found in Sec. 6.4.1, at the end:

  The dark green cluster corresponds to heavily rimed ice particles (RI). The larger density of rimed particles lead to significant Z_H signatures, and the dielectric properties of dense ice, very different with respect to dry crystals and aggregates (Vivekanandan et al., 1994), lead to generate a response also in K_{dp}. Z_{DR} is low because riming tends to smooth particle shapes, and it shows negative values when conically-shaped rimed particles are formed (Evaristo et al., 2013).

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).

We thank the reviewer for the clarification about the "drizzle" category and for addressing us to the AMS glossary. We decided to change the names of the liquid-phase categories to: light rain (LR), rain (R), and heavy rain (HR). Drizzle is supposed to be a subset of LR. We rephrased the text of Sect. 6.2 as:

The green cluster is characterized by extremely low rainfall intensity, and therefore it is associated hereafter with a hydrometeor class named light rain (LR). This cluster contains mainly precipitation characterized by small spherical drops. It is worth noting (Fig 5, panels b and c) that LR contains Z_{DR} values lower than 1 dB, with a mode around 0.25 dB, and K_{dp} values always close to $0\,°\mathrm{km}^{-1}$. LR therefore contains drizzle and the lightest rainfall intensities.

Also, we want to underline that we (and also Dolan and Rutledge (2009)), consider only drizzle at temperatures $> 0°$C, and we do not claim identification of supercooled liquid water with an X-band radar. This would be too ambitious given the radar frequency.

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.

We agree on the fact that Z_{DR} zeroing should not be performed in presence of oblate drops. However, also drops larger than $1mm$ must give $Z_{DR} = 0$ at vertical incidence. Only pathological cases, with large and non-random canting of the raindrops at all the altitudes, will create significant uncertainties in the zeroing of Z_{DR}. In our case, the standard deviation of Z_{DR} offset estimates was about 0.12 dB during the HyMeX campaign.

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?

   We mentioned the differential attenuation in Sect. 3.2, and we additionally provide the reference for the attenuation correction algorithm that we employed. We rephrased this part of the manuscript as follows:

   Z_{H} and Z_{DR} are corrected for attenuation (in rain only) using the relations linking K_{dp}, Z_{H}, specific horizontal attenuation α_{H} [$dB\,km^{-1}$], and differential attenuation α_{DR} [$dB\,km^{-1}$] according to the method of Testud et al. (2000). The power laws between these variables are parametrized using disdrometer measurements for the data collected in France (locations shown in Fig.1b) and using simulated realistic drop size distribution fields (Schleiss et al., 2012) for the data collected in Switzerland.

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 judgement is lacking. Does their method result in qualitative/quantitative improvement? (It seems not since the results between theirs and Dolan/Rutledge generally agree.)

    Following the suggestion of reviewer 3 and reviewer 2 we added in the revised manuscript a quantitative evaluation that, with the 2DVD as a reference, shows that our algorithm achieves better scores than DR2009 in the classification of winter events collected in Davos (see the confusion matrices of the revised manuscript, and related scores). This statistical evaluation is an innovative aspect of the revised manuscript, with respect to the available literature on the subject. However, a thorough quantitative evaluation among classification techniques is by itself a very innovative and difficult topic and we stress all the possible limitations of such exercises in the manuscript.

    This analysis does not aim at demonstrating that our technique is better, but rather that this innovative approach is promising. As an example we mention that (Conclusions):

    The proposed approach is the first and successful attempt, using unsupervised classification, to move the starting point of a classification algorithm away from scattering simulations conducted over an arbitrarily defined number of hydrometeor classes to the identification of relevant clusters in the data themselves.

    and also:

    Among the advantages of the approach, we remind that it is immune to radar calibration and the data-driven approach ensures that the identified clusters take into account the accuracy of the instrument

    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.

    The method is designed to be optimal for a given instrument and a given set of measurements. In our case, we applied it on a large database and we are confident that the clusters cover the dominant situations encountered with MXPol, or by similar radars in similar climatologies. We believe it also because our clusters are overall similar to the ones of Dolan and Rutledge (2009). Of course anyone can implement the method on smaller databases (for examples one focus only on thunderstorms), or with different spatial constraints. The clusters obtained will be representative of the dataset and of the given conditions.

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?

We thank the reviewer for this important remark. About the robustness of the selction of 7 clusters and their content, now we base the analysis of Fig 4, 5, 9 (numbering of the old manuscript) on 100 realizations of the clustering algorithm and not on a single one, as it was done previously. We agree that the choice of 6,7, or 8 clusters is not so different, but n_{opt} is nevertheless a local optimum,

We do agree that there are many choices in the manuscript. Some examples are the standardization of the data presented in section 4.1, or the random sampling of section 4.2. We verified the robustness of the algorithm with respect to these choices[4] but we preferred not to present them in the manuscript that is already long. Instead, we introduced in the text sentences like (Sec 4.1):

ans Variations on the order of $\pm 20\,\%$ around the proposed boundaries have a negligible impact on the results presented in the following sections, and the most sensitive boundary has found to be $\mathbf{x}_{\min}[1]$ ($\mathbf{x}_{\max}[1]$, associated to Z_H...

... Gaussian, sigmoid and logistic functions have been tested and appeared to be equally adequate...

Then, in Sec 4.2:

Different seeds of the initial random selection were leading to the same results, suggesting that the random sampling does not affect the outcome of the clustering technique presented in the next section...

And in Sect. 4.3:

Among the possible combinations of distance metric and linkage criteria presented in Sect. 2 we obtained similar results also using the correlative distance and/or WC linkage.

We evaluated these statements even though we did not show the results. As an example for the reviewer, Figure 3 (end of the present document) illustrates the robustness of the classification output as a function of a bias on the standardization boundaries of Sect 4.1 (left panel), and as a function of a bias on the polarimetric variables (and Δz) themselves. We observe that variations on the order of $\pm 20\%$ result in less that 15% of changes in the classification output (the error is mitigated and not amplified).

Detailed edits on the text

1. *Page 8466, Abstract. Shouldn't you indicate somewhere that nc = 7?*

   We modified a sentence of the abstract according to the suggestion of the reviewer:

   Seven hydrometeor classes ($n_{\mathrm{opt}=7}$) have been found in the dataset and they have been identified as ...

2. *Page 8467. You might tell why algorithms are different at different radar frequencies.*

   We rephrased a sentence in the introduction as follows:

   Different HC algorithms are used at different frequencies, such as Straka et al. (2000); Liu and Chandrasekar (2000) for S-band, Marzano et al. (2007); Dolan et al. (2013) for C-band, and Dolan and Rutledge (2009); Snyder et al. (2010); Marzano et al. (2010) for X-band. This is necessary because the scattering properties of hydrometeors vary with respect to the incident wavelength.

3. *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?*

   In this case we wanted to underline the fact that there is a certain degree of subjectivity in the selection of the available categories of an algorithm. To give some practical examples: (Dolan and Rutledge, 2009) classifies radar data into one of the following classes : rain, drizzle, high density graupel, low density graupel, vertically aligned ice, aggregates, crystals. Al-Sakka et al. (2013) has the following classes: rain, dry snow, wet snow, hail, ice. Keenan (2003) classifies drizzle, rain, dry snow (low and high density), wet snow, dry graupel, wet graupel, small hail, large hail, rain mixed with hail. Note that some algorithms divide graupel by density, and other by "wetness", some algorithms identify hail, other do not, and so on.

[4] Even though for the metrics and the merging rules we tested only the ones presented in the manuscript and not the hundreds of possible combinations that can be found in the literature.

The subjectivity is given by the fact that each "developer" is free to select the hydrometeor classes that he/she prefers (or needs), then simulate their scattering properties and finally classify actual measurements according to the simulations. Little has been done up to now to investigate if the number and type of hydrometeor classes that one selects is reasonable. With our approach we start from the data instead that from numerical simulations also because of this reason, i.e. we want to see which groups are "resonable" in a given dataset, and then think about their content in a second step only.

We rephrased the sentence highlighted by the reviewer as follows:

First, the choice of the hydrometeor classes, meaning their content and their number, is mostly subjective.

4. *Page 8468, Line 11. Definitely an interesting approach.*

   Thanks!

5. *Page 8468, Lines 15-17. Nothing is said here about distinct separability of the groups by the clustering technique. (See also substantive issue (6) above)*

   Please note that $n_{opt} = 7$ actually is the best partition in terms of separability (trade-off between scattering within clusters and distance between clusters). Also, please refer to our answer to the substantive issue (6).

6. *Page 8470, Line 17. It seems to me that this distance should always be discussed in relation to the cluster size, in some sense.*

   This is indeed what a merging rule does (Sec 2.3). It adapts the concept of distance between object and translates it to distance between clusters, taking into account their sizes. See for example the definition of the WC merging rule in the section:

   WC defines the distance between clusters as the distance between the (weighted) centroids of each cluster. The centroid is the centre of mass of a cluster C_I. It is computed as the average position of all the sub-clusters $C_K \subset C_I$, weighted by the number of objects in each C_K.

7. *Page 8472, Line 7. I have tried to encourage above the provision of more prior knowledge about the radar data.*

   We agree with the reviewer (and with reviewer 1) about the added-value of prior physical knowledge for the clustering, but this is beyond the scope of this manuscript, which is focussed on presenting an innovative unsupervised clustering approach. The development of a supervised (or semi-supervised) clustering approach taking advantage of prior knowledge will be conducted in future work.

8. *Page 8472, 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.*

   We clarified this sentence in the following way:

   The altitude of the deployment site made it possible, during cold seasons, to collect many observations of ice-phase precipitation when the radar itself was located above the melting layer, and therefore did not suffer from liquid-water signal attenuation. Such radar observations represent the main peculiarity of this field campaign (e.g. Schneebeli et al., 2013; Scipion et al., 2013). However, during warm seasons, the melting layer was often higher than the radar site and relevant observations of liquid phase precipitation, both in stratiform and convective cases, were collected as well.

9. *Page 8473, Lines 4-5. ZDR is not really calibrated with vertically pointing, but rather only zeroed.*

   This is true. In the radar meteorology research domain this approach is commonly called "Z_{DR} calibration" and we used this expression. We acknowledge anyway the fact that the vertically pointing PPI is only zeroing Z_{DR} and we substituted this expression in Sec 3.1.

10. *Page 8473, Lines 8-9. The authors could say much more about the meteorological characteristics of these storms that they do here.*

We understand the request of the reviewer. However, we would like to avoid to give an excessive amount of information that we will then not comment further. Also, many storms were collected in different years and it is rather difficult to provide meteorological information about all of them. We tried to add in the manuscript some sentences that should help the reader to understand the basic characteristics of the climatology of the places where the data were collected (see Substantive Issue 1, above).

11. *Page 8473, 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.*

The sentence has been rephrased according to the suggestion of reviewer 2, and the term "calibration" does not appear any more.

12. *Page 8473, Line 17. OK, but work on rain will require the HyMeX data set. You should say so.*

We clarified throughout the manuscript (see comments before and after) that the clustering algorithm uses together data from both campaigns.

13. *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)?*

This is a very good question. An automated clutter filter (based on Doppler) was integrated in the radar software. However, when the clutter signal is very strong, it "survives" the clutter filter. The measurements contaminated by clutter are separated and censored by means of a threshold on ρ_{hv} (0.7), which is a polarimetric variable that has very low values in clutter (e.g. Doviak and Zrnić, 2006; Gourley et al., 2007).

14. *Page 8475, Lines 16-18. Negative values can be meaningful and should be examined.*

Negative values are examined in the revised manuscript. We followed the suggestion of the reviewer and we believe that it contributed to a better definition of the clusters, even though we did not obtain a "8th" cluster containing only negative K_{dp} values. The sentence has been rephrased as:

This issue is tackled by log-transforming K_{dp} values. Before log-transforming we add $1^\circ\,km^{-1}$ to K_{dp}, in order to consider K_{dp} values in the range $[-1,15]$ [5].

15. *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.*

Units have been included in the equation. It is not fundamental for the approach that $\mathbf{x}^*[5] = 0$ in the warm part. The only important things for $\mathbf{x}^*[5]$ are: (i) to span between 0 and 1 like the other normalized components $\mathbf{x}^*[1, \ldots 4]$ and (ii): to take constant values (in this case 0) below the local 0°C isotherm, and constant values (κ, $0 < \kappa \leq 1$.) above the local 0°C isotherm as well. We explain the meaning of this component of $\mathbf{x}$ some lines after the equation:

...this parameter is intended only to flag positive and negative temperatures in a quasi-binary way and not to substitute the information provided by the polarimetric variables (therefore κ is kept strictly ≤ 1).

16. *Page 8477, Line 3. I don't understand what "full distance matrices" are.*

We rephrased the sentence as follows:

Agglomerative clustering algorithms are generally computationally expensive, because the distances among all samples (and then groups) to be clustered are computed at each step of the hierarchical aggregation chain.

[5] $K_{\mathrm{dp}} < -1^\circ\,km^{-1}$ occurs less than 0.01% of the cases in our database.

What we called a "full distance matrix", generating some confusion, was simply a "distance matrix". A very simple example can be found at: `http://en.wikipedia.org/wiki/Distance_matrix`.

17. *Page 8477, Lines 7-9. But always with radar making observations above the melting level?*

    In this case we referred to both datasets, aggregated. The main goal of this section is to find the clusters (and then hydrometeor classes) that globally emerge from our datasets. For sake of clarity, we rephrased the sentence as follows:

    These events cover the range of precipitation types observed by MXPol during the field campaigns of Davos (CH) and Ardèche (FR) and they are assumed to be a representative sample of mid-latitude temperate precipitation.

18. *Page 8477, Line 12. Useful to remind reader exactly what an "observation" is. If that is equation (7), then give it.*

    We added a reference to the appropriate equations:

    ... 20 000 observations $\mathbf{x}^*$ (defined in Eqs. 7, 9, and 10)...

19. *Page 8477, Line 27. What is meaning of "correlative distance metrics"?*

    "Correlative distance metric" refers to the metric presented in Table 1. However, in order to ease the readability of the paragraph, we rephrased this section in the revised manuscript.

20. *Page 8478, Line 5. Why was the number 1000 chosen here?*

    1 000 is a number large enough to:

    - Be confident that any optimal partition of the dataset will be found at a lower (i.e. we do not expect to obtain a classification algorithm with 1200 hydrometeor classes).
    - Be sure that the spatial smoothness constraint is taken into account since the early stages of the aggregation of the dataset.

    The choice of 1 000 was first introduced in Sect. 4.3.1, with these sentences:

    A first hierarchical aggregation is conducted on the data, until reaching a number of 1 000 clusters in the dataset. This step aims at merging the most similar objects before proceeding with more computationally expensive calculations.

    Now, in order to clarify the choice, we added a footnote after the word "dataset". The footnote reads:

    By doing this we assume that the optimal partitions of the dataset are found when $n_\mathrm{c} \leq 1000$.

21. *Page 8478, Line 16. This jumps ahead in the figures list.*

    We thank the reviewer for noticing that. Because the figure has more sense if it appears later than here, we decided to remove the sentence from the revised manuscript. This figure is anyway described in the following sections.

22. *Page 8481 Line 25. How is goodness quantified? There is nothing about "goodness" in Figure 2b. And nothing about cluster quality indices.*

    The full paragraph has been rephrased as follows:

    An important step of hierarchical clustering is the selection of the optimal partition (n_opt) of the dataset. In the present section we introduce some indices that evaluate the quality of data partitions and that are used to guide the final selection of n_opt.

23. *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?*

    We realized that there were too many quality indices defined here (Sec 5.2). We revised the manuscript by defining only three indices, that are applied in the very next section (Sec. 5.3). Section 5.3 begins in fact with this sentence:

    Figure 3 illustrates the behaviour of the quality indices defined in Sec 5.2 as a function of the number of clusters in the dataset…

24. *Page 8485. Section 6.2 should begin by saying that HyMeX radar data are being examined here.*

    We respectfully disagree. In fact the data of the clusters come from both datasets. Here the HyMeX dataset (rich of pure rain events) is used to evaluate the content of the three "liquid phase" clusters, and this is explained in the first paragraph, with the sentence:

    At first, all data classified into one of these three categories are extracted from the whole field campaign of HyMex…

25. *Page 8485. 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.*

    We agree with the reviewer. We assume that this is a good assumption given the large size of our database (i.e. hail has very limited occurrence with respect to rain, in any climatology). However, we believe that hail is a "sub-category" of our heavy rain (HR) class. In the revised manuscript, in Sect. 6.2 we mention that:

    Finally, the red cluster contains by far the highest rainfall intensities, and it is named hereafter heavy rain (HR). We also hypothesize that when hail occurs it is classified as HR. We base this assumption on the fact that HR includes observations with low correlation coefficient ρ_{hv} (Fig 5 d) as well as near-zero or negative Z_{DR} (Fig 5 b). These signatures have been documented in cases where hail was measured by polarimetric weather radars (Al-Sakka et al., 2013).

26. *Page 8486. Section 6.3 should begin by saying that the same HyMeX data are being examined here.*

    As mentioned in point 24. (above), the data of the clusters come from both campaigns and only the specific example presented in this section is taken from Ardèche (HyMeX). We followed another suggestion of the reviewer and the figure referenced in this section now clearly mentions in the caption the HyMeX campaign.

27. *Page 8486, Line 13 It would be useful to cite the observed thickness as additional evidence for this assertion.*

    We do not agree with the reviewer on this point. The melting layer in the figure is in our opinion clearly visible according to its standard polarimetric signatures (low ρ_{hv}, enhanced Z_H and Z_{DR}, Doviak and Zrnić, 1993). There is absolutely no doubt that we have a melting layer in this case.

28. *Page 8486, Line 17. Many dual pol experts claim that rhohv is the best variable to show the melting layer. Do the present authors agree?*

    We do agree. ρ_{hv} has the big advantage to be very close to 1 both in pure rain and in pure snow, and it drops only in the melting layer. Z_H and Z_{DR} are instead only enhanced in the melting layer, but they do not have a priori fixed (or similar) values below and above it. About this topic there are two interesting works by Brandes et al. (2004); Giangrande and Ryzhkov (2008) that outline the different behaviour of the polarimetric variables in the melting layer.

29. *Page 8487. First line: what quantitative info is provided? This could be beefed up.*

    We rephrased the sentence as:

    Subsequently we compare the classification with qualitative (hydrometeor classification) and quantitative (snowfall intensity) observations provided by a two-dimensional video disdrometer (2DVD) and with the output of a numerical weather prediction model (COSMO).

30. *Page 8488. Line 3 Change "colder" to "lower".*

    The change has been made.

31. *Page 8488. Line 13. What is mechanism for "vertically aligned crystals".*

    We added a clarification in the section:

    . . . vertically aligned ice (VI, that denotes oblate ice crystals aligned vertically because of an electric field).

32. *Page 8488. Lines 20-21 We need to know about the rules in this paper.*

    An extensive description of the algorithm is beyond the scope of the present manuscript, and the interested reader should refer to Grazioli et al. (2014). We tried to provide the main characteristics of the algorithm with the following sentences:

    An additional comparison is conducted with the output of a HC scheme developed for two-dimensional video disdrometers. This method, hereafter called HC2DVD is described detail in Grazioli et al. (2014). HC2DVD takes as input the set of two-dimensional particles images, collected by a 2DVD, and it provides as output an estimate of the dominant hydrometoer type within time intervals of 60s. The method does not classify individual particles but populations of hydrometeors. HC2DVD discriminates between eight hydrometeor classes: small particle-like (SP), dendrite-like (D), column-like (C), graupel-like (G), rimed particle-like (RIM), aggregate-like (AG), melting snow-like (MS), and rain (R)[6]. The "-like" is added to underline that this algorithm assigns a dominant type of hydrometeor to each time step but there is usually a mixture of different hydrometeors captured by the 2DVD so they do not necessarily exhibit a single pristine shape.

33. *Page 8489, Line 23. The color looks turquoise rather than light blue on my print.*

    We call the color "cyan" in the revised manuscript.

34. *Page 8489, 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.*

    We revised deeply this section, we added a short description of the basis of HC2DVD and we simplified the comparison among different methods by aggregating some of the hydrometeor classes. Additionally, as mentioned previously, we found some minor issues in some Z_{DR} values collected in Davos. These issues did not have a major impact on the content of the clusters, but they degraded the comparison with HC2DVD. The comparison should now be easier to interpret (quantitative scores are also given).

35. *Page 8490, Line 7. Additional need for knowing spatial resolution.*

    The sentence has been removed from the revised manuscript.

36. *Page 8490, Line 15. Why not say "1 min"?*

    We prefer to express the quantities with the units that are used in the equation (EF has units of $\mathrm{mm\,h^{-1}}$).

37. *Page 8490, Line 21. Yes the snowfall intensity differs, but there is substantial overlap. (The differences do not allow for a unique specification.)*

    There is overlap of snowfall intensities. However, we use this comparison with a lot of care and we clearly state in the end of the section that it is not a definite proof, but just a qualitative confirmation.

[6] We use here the abbreviations of Grazioli et al. (2014)

The results presented in this section are not a rigorous validation but they are in agreement with our initial assumptions.

38. *Page 8491, Line 5. I would add at the end of this first sentence "in Davos"*

    We added the location in the text, according to the suggestion of the reviewer.

39. *Page 8491, Line 13. Quantify "very low values of K_{dp}"*

    We quantify the statement as:

    between -0.1 and 0.1 °km^{-1}

40. *Page 8491, Line 14. OK, but are clusters cleanly distinguishable? What about the overlap problem?*

    As mentioned before, the clusters themselves are clearly distinguishable. Then, new data associated to the clusters can be "close" to more than one cluster. In order to illustrate this point we added to each PPI and RHI plot a map that shows the "classification accuracy". We use a rather restrictive index, defined in the text as:

    This parameter is defined for each observation (valid range gate) as the difference between the distance of the observation with respect to the two closest clusters, and it is normalized on the smaller distance. The classification accuracy is therefore lower in the areas of transition between different hydrometeor types, where the polarimetric signatures change as the dominant hydrometeor type changes.

    This index illustrates well that in the transitions between different hydrometeor types the accuracy decreases (as well as on the edges of the domains, where the received signal is close to the radar sensitivity) and therefore reminds to the reader that hydrometeor classification is a discrete mapping of otherwise continuously varying quantities.

41. *Page 8491, Line 25. "to a dataset of radar data" is redundant*

    We rephrased the sentence as follows:

    The method is applied to polarimetric data collected by an X-band radar in the Swiss Alps and in the French prealps.

42. *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.*

    We agree with the reviewer, it would be very convenient and relevant to have in-situ (in-cloud) validation. Nevertheless the reviewer should consider that the comparison between aircraft data and radar measurements is not as trivial as it seems. In particular, to our knowledge, there are no study that quantitatively compare statistically significant sets of radar hydrometeor classification outputs and aircraft measurements.

43. *Page 8492, 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).*

    We compare with DR2009 indeed because it was specifically developed for X-band. In the revised version of the manuscript we provide also a tentative quantitative evaluation of the classification output between the two methods. However, the goal of the present manuscript is not to compare against a set of other algorithms, but to illustrate how the application of unsupervised techniques is promising for the field. We believe that a complete inter-comparison among methods is beyond the scope of the manuscript and it would distract the reader from the main messages of the paper.

44. *Page 8492, Lines 28-29 This would be easy to test with NEXRAD and European radar networks, and the authors should say something about this possibility.*

We clarify now that we expect similar results if the radars are similar as well as the databases (our previous statement was too strong.):

We nevertheless expect the number and type of clusters to be very similar for other X-band dual-polarization radars of similar sensitivity.

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

We removed Appendix A from the revised version of the manuscript to shorten the paper.

46. *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.*

Table 1 is used in the "background of clustering methods" section. The goal of the section is to give to the reader a basic understanding of the main theoretical basis of clustering. The table illustrates how can we define the concept of distance between two vectors (objects) in a $d-$dimensional space. In order to clarify the section we simplified the table by removing one metric, and we discuss explicitly the concept of "Euclidean distance" in a d-dimensional space, that should be the most intuitive metric. In Sec. 2.1:

For instance, the Euclidean distance [7] is defined in a $d-$dimensional space as:

$$D(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^{d} |\mathbf{x}[i] - \mathbf{y}[i]|^2}$$

and it is a good metric to evaluate the similarity...

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

We rephrased and simplified Sects. 5.1 and 5.2 and the concepts should be exposed more clearly now. Please note that the SD index, introduced in Sect. 5.2 of the revised manuscript, is actually used to find the optimal trade-off between cluster size and cluster compactness and separability.

48. *Page 8501. Do you mean "orthogonal"?*

Yes, thanks for spotting the typo.

49. *Page 8501. Table 3 This table should include the actual final categories on the left side (Drizzle, Light Rain, Heavy Rain)*

The table now includes the final categories.

50. *Page 8501. 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.*

We changed the name of the hydrometeor class, from "Drizzle" to "Light rain" (LR). We still assume anyway that actual drizzle is a subset of LR. We thank the reviewer for pointing out the exact definition of drizzle.

[7] A particular case of "Minkowski distance" when $p = 2$, according to the notation of Table 1.

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

The caption includes now these definitions. It is rephrased as:

Parameters of the membership beta functions β employed in the DR2009 algorithm: midpoint m, width a and slope b for the available hydrometeor classes.

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

We believe that the reviewer refers here to the parametrization of Dolan and Rutledge (2009). This method is based on fuzzy-logic (see Straka et al., 2000, for a complete overview), and in particular on mono-dimensional membership beta functions (MBF). Therefore, strictly speaking, we do not have access to the mean Z_{DR} but instead we have access to the parametrization of the MBF $\beta(Z_{DR})$, defined in Appendix B. By looking at the values in table 4 (numbering of the old manuscript), we observe that the mid point m for Z_{DR} and for high density graupel (HDG) is 1.2 dB, and the width a is 2.5 dB. This means that, from scattering simulations, HDG showed Z_{DR} values in the range $[-1.3, +3.7]$ and the value that gives the highest score to the HDG class in the fuzzy-logic scheme is $Z_{DR} = 1.2$ [dB].

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

The caption is now rephrased as:

Parameters of the trapezoidal membership function T applied to the relative altitude with respect to the 0°C isotherm (Δz [m]). l_1, l_2, l_3, l_4 are the four vertices of the trapezoid T.

54. *Figure 1 The info in this caption about radar sites should agree with that given in the text.*

We rephrased the caption to be consistent between text and caption.

Why these choices for "the directions of the RHI scans"?

The RHI scan directions were selected in a way that each scanning sequence contains the largest amount of "quasi 3-D" information, given the physical constraints of the topography. Additionally, during HyMeX, the RHIs were designed to scan in the directions of other instruments, for purposes that are beyond the scope of the present manuscript.

Why is this important for the paper?

We believe that it is important to provide the information about the scanning strategy to put the data into context. Additionally, this manuscript is a contribution to the HyMeX programme and other scientist participating in HyMeX can find this information useful.

Reference to "polarimetric power laws"? I don't recall much about that in the text? Are these power laws on rain?

We refer here to the parametrization of the attenuation correction scheme. In Sect. 3.2 (now 3.3 in the revised manuscript) we mention:

The power laws between these variables are parametrized using disdrometer measurements for the data collected in France locations shown in Fig. 1b) and

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

We prefer to keep the flow chart decoupled from our specific observations, because the idea of this method can apply to different datasets and not only to ours.

56. *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 4 uses data from both field campaigns and it is employed to find the number of clusters that better adapts to our dataset. The figure now shows only 3 curves because we simplified Sect.5.2 and 5.3 that were not very clear to the reviewers. We rephrased the caption as:

Evolution of Kappa, accuracy spread (AS) index, and SD index as a function of the number of clusters in the dataset. The SD index is stretched between 0 and 1 for illustration purposes. The yellow vertical line at $n_c = 7$ shows the selected final number of clusters, corresponding to a minimum AS and SD. Each curve shows the mean behaviour over 100 runs of the clustering algorithm.

57. *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 5 is showing the content of all the seven clusters, that are obtained using data from Davos <u>and</u> HyMeX. The goal of the clustering technique was to find the clusters (hydrometeor classes) that better adapt to the range of observations of our X-band radar system (i.e. the clustering is not performed separately for different locations). Also, the purpose of the figure is to give an idea of 2-D projections for 5-D clusters, and to show that some clusters appear always at positive (negative) temperatures. We rephrased the caption, according to the suggestion of the reviewers as:

    Three examples of 2-D projection of the 7 clusters found in the dataset.The clusters include observations collected in Davos (CH) and Ardèche (FR). **(a)**: Z_{DR} vs. ρ_{hv}, **(b)**: Z_{H} vs. K_{dp}, **(c)**: Δz vs. ρ_{hv}. Note that some of the clusters are not fully visible in this 2-D projections, because they are defined in a 5-D space

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

    Figure 6 was using data collected during both campaigns. It used the output of the clustering algorithm, that applies to the complete database. We substituted now the figure with a contingency table and we added in the caption the sentence:

    The data are obtained with 100 runs of the clustering algorithm.

59. *Figure 7 "collected on the 24 September"*

    The change has been made.

60. *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 8 (numbering of the old manuscript) now shows a case with some riming. However, the purpose of this figure is to illustrate a case with a clear melting layer and justify the reason why we associate the yellow cluster with melting snow (MS), and not to compare with Figure 13. We selected now for Fig 8 a case with a clearer Z_{DR} enhancement around the melting layer.

61. *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.*

    This figure can be considered as a 1-D interpretation of Fig 5 (numbering of the old manuscript), so it considers data from both field campaigns. We clarified now this fact by rephrasing the caption as follows:

    Distribution within the three clusters found at negative temperatures ($\Delta z > 0$) of: (a) Z_H [dBZ], (b) Z_{DR} [dB], (c) K_{dp} [$^\circ$ km^{-1}], **(d)** ρ_{hv} [–], (e) Δz [km]. The curves are obtained considering the content of 100 runs of the clustering algorithm.

    The clusters are slightly revised with respect to the previous manuscript, mainly because of the suggestion to include negative K_{dp} and because we solved an issue affecting Z_{DR} zeroing during some periods of the

Davos campaign. Negative Z_{DR} is now observed in the RI class, that should include graupel. About the overlap, we believe that we answered to this question in the previous points.

62. *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.*

    We rephrased, by naming these clusters as "ice-phase" clusters. Additionally, this figure has been replaced by a confusion matrix and some classes of Dolan and Rutledge (2009) have been grouped together.

    It is true that DR2009 does not name explicitly a category as "conical graupel". However, the simulations (Sect.2 of Dolan and Rutledge, 2009) are designed to include this type of hydrometeor within the two graupel classes (HDG, LDG).

63. *Figure 11 Ditto on change of "freezing"*

    Please refer to the previous point.

64. *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.*

    We would like to underline that Fig.5 of the previous manuscript was showing all the clusters and not only the ones of Fig. 8. The 7 clusters shown in Fig 5 contained data coming from the database of Davos and HyMeX together. They correspond to the partition that better applies to our radar system and our full database. We decided to give additional information about the content of the 7 clusters and we included a new figure (same as figure 9 of the old manuscript, but for the liquid-phase clusters) and also a table in the appendix with statistics on the content of all the clusters (Table A1 in the revised manuscript).

    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

    The zero of altitude in the figures was indeed the radar location. We agree that it was confusing and in the revised manuscript we display the actual altitude on the y-axis of the RHI maps (i.e., starting from $\approx$ 2133 m in Davos and 605 m for HyMeX.)

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

    We added in the manuscript a sentence describing the main characteristics of the snowfall event:

    This event was associated with the passage of a cold front over Europe that was leading in Davos to a significant temperature drop (of about 15°C in few hours)...

References

Al-Sakka, H., A.-A. Boumahmoud, B. Fradon, S. J. Frasier, and P. Tabary, 2013: A new fuzzy logic hydrometeor classification scheme applied to the french X-, C-, and S-band polarimetric radars. *J. Appl. Meteor. Clim.*, **52 (10)**, 2328–2344, doi:10.1175/JAMC-D-12-0236.1.

Boudevillain, B., G. Delrieu, B. Galabertier, L. Bonnifait, L. Bouilloud, P.-E. Kirstetter, and M.-L. Mosini, 2011: The Cévennes-Vivarais Mediterranean Hydrometeorological Observatory database. *Water Resour. Res.*, **47**, W07701, doi:10.1029/2010WR010353.

Bousquet, O., et al., 2014: Multiple-frequency radar observations collected in southern france during the field phase of the hydrometeorological cycle in the mediterranean experiment (hymex). *Bull. Amer. Meteor. Soc.*, in review.

Brandes, E. and K. Ikeda, 2004: Freezing-level estimation with polarimetric radar. *J. Appl. Meteor.*, **43 (11)**, 1541–1553, doi:10.1175/JAM2155.1.

Brandes, E., G. Zhang, and J. Vivekanandan, 2004: Drop size distribution retrieval with polarimetric radar: Model and application. *J. Appl. Meteor.*, **43 (3)**, 461–475.

Bringi, V. N. and V. Chandrasekar, 2001: *Polarimetric doppler weather radar.* Cambridge University Press, 662 pp.

Caylor, I. and V. Chandrasekar, 1996: Time-varying ice crystal orientation in thunderstorms observed with multiparameter radar. *IEEE T. Geosci. Remote Sens.*, **34 (4)**, 847–858.

Dolan, B. and S. A. Rutledge, 2009: A theory-based hydrometeor identification algorithm for X-band polarimetric radars. *J. Atmos. Oceanic Technol.*, **26 (10)**, 2071–2088, doi:10.1175/2009JTECHA1208.1.

Dolan, B. and S. A. Rutledge, 2010: Using casa ip1 to diagnose kinematic and microphysical interactions in a convective storm. *Mon. Weather Rev.*, **138 (5)**, 1613–1634, doi:10.1175/2009MWR3016.1.

Dolan, B., S. A. Rutledge, S. Lim, V. Chandrasekar, and M. Thurai, 2013: A robust C-band hydrometeor identification algorithm and application to a long-term polarimetric radar dataset. *J. Appl. Meteor. Clim.*, **52 (9)**, 2162–2186, doi:10.1175/JAMC-D-12-0275.1.

Doviak, R. and D. Zrnić, 1993: *Doppler radar and weather observations, second edition.* Academic Press, San Diego, 562 pp.

Doviak, R. and D. Zrnić, 2006: *Doppler radar and weather observations, second edition.* Dover Publications, 592 pp.

Evaristo, R., T. Bals-Elsholz, E. Williams, A. J. Fenn, M. Donovan, and D. Smalley, 2013: Relationship of graupel shape to differential reflectivity: theory and observations. *29th Conference on Environmental Information Processing Technologies*, Austin, USA.

Giangrande, S. E. and A. V. Ryzhkov, 2008: Estimation of rainfall based on the results of polarimetric echo classification. *J. Appl. Meteor. Clim.*, **47 (9)**, 2445–2462, doi:10.1175/2008JAMC1753.1.

Gourley, J. J., P. Tabary, and J. Parent-du Chatelet, 2007: A fuzzy logic algorithm for the separation of precipitating from nonprecipitating echoes using polarimetric radar observations. *J. Atmos. Oceanic Technol.*, **24 (8)**, 1439–1451, doi:10.1175/JTECH2035.1.

Grazioli, J., D. Tuia, S. Monhart, M. Schneebeli, T. Raupach, and A. Berne, 2014: Hydrometeor classification from two-dimensional video disdrometer data. *Atmos. Meas. Tech.*, **7 (9)**, 2869–2882, doi: 10.5194/amt-7-2869-2014.

Huang, G., V. N. Bringi, R. Cifelli, D. Hudak, and W. A. Petersen, 2010: A methodology to derive radar reflectivity liquid equivalent snow rate relations using c-band radar and a 2d video disdrometer. *J. Atmos. Oceanic Technol.*, **27**, 637–651.

Keenan, T., 2003: Hydrometeor classification with a C-band polarimetric radar. *Aust. Meteorol. Mag.*, **52 (1)**, 23–31.

Kruger, A. and W. F. Krajewski, 2002: Two-dimensional video disdrometer: a description. *J. Atmos. Oceanic Technol.*, **19 (5)**, 602–617.

Liu, H. P. and V. Chandrasekar, 2000: Classification of hydrometeors based on polarimetric radar measurements: Development of fuzzy logic and neuro-fuzzy systems, and in situ verification. *J. Atmos. Oceanic Technol.*, **17 (2)**, 140–164.

Marzano, F., D. Scaranari, and G. Vulpiani, 2007: Supervised fuzzy-logic classification of hydrometeors using C-band weather radars. *IEEE T. Geosci. Remote Sens.*, **45 (11)**, 3784–3799, doi:10.1109/TGRS.2007.903399.

Marzano, F. S., G. Botta, and M. Montopoli, 2010: Iterative bayesian retrieval of hydrometeor content from X-band polarimetric weather radar. *IEEE T. Geosci. Remote Sens.*, **48 (8)**, 3059–3074.

Matrosov, S. Y., 1992: Radar reflectivity in snowfall. *IEEE T. Geosci. Remote Sens.*, **30 (3)**, 454–461.

Miniscloux, F., J.-D. Creutin, and S. Anquetin, 2001: Geostatistical analysis of orographic rain bands. *J. Appl. Meteor.*, **40 (11)**, 1835–1854, doi:10.1175/1520-0450(2001)040⟨1835:GAOOR⟩2.0.CO;2.

Mott, R., D. Scipion, M. Schneebeli, N. Dawes, A. Berne, and M. Lehning, 2014: Orographic effects on snow deposition patterns in mountainous terrain. *J Geophys. Res. Atmos.*, **119 (3)**, 1419–1439, doi: 10.1002/2013JD019880.

Otto, T. and H. W. J. Russchenberg, 2011: Estimation of specific differential phase and differential backscatter phase from polarimetric weather radar measurements of rain. *IEEE Geosci. Remote Sens. Lett.*, **8 (5)**, 988–992, doi:10.1109/LGRS.2011.2145354.

Pruppacher, H. R. and R. L. Klett, 1997: *Microphysics of clouds and precipitation.* 2d ed., No. 18 in Atmospheric and Oceanographic Sciences Library, Kluwer Academic Press, 954 pp.

Ryzhkov, A. V. and D. S. Zrnic, 2007: Depolarization in ice crystals and its effect on radar polarimetric measurements. *J. Atmos. Oceanic Technol.*, **24 (7)**, 1256–1267, doi:10.1175/JTECH2034.1.

Schleiss, M., J. Jaffrain, and A. Berne, 2012: Stochastic simulation of intermittent DSD fields in time. *J. Hydrometeor.*, **13 (2)**, 621–637, doi:10.1175/JHM-D-11-018.1.

Schneebeli, M. and A. Berne, 2012: An extended Kalman filter framework for polarimetric X-band weather radar data processing. *J. Atmos. Oceanic Technol.*, **29 (5)**, 711–730, doi:10.1175/JTECH-D-10-05053.1.

Schneebeli, M., N. Dawes, M. Lehning, and A. Berne, 2013: High-resolution vertical profiles of polarimetric X-band weather radar observables during snowfall in the Swiss Alps. *J. Appl. Meteor. Clim.*, **52 (2)**, 378–394, doi:10.1175/JAMC-D-12-015.1.

Scipion, D., R. Mott, M. Lehning, M. Schneebeli, and A. Berne, 2013: Seasonal small-scale spatial variability in alpine snowfall and snow accumulation. *Water Resour. Res.*, **49 (3)**, 1446–1457, doi:10.1002/wrcr.20135.

Snyder, J. C., H. B. Bluestein, G. Zhang, and S. J. Frasier, 2010: Attenuation correction and hydrometeor classification of high-resolution, X-band, dual-polarized mobile measurements in severe convective storms. *J. Atmos. Oceanic Technol.*, **27 (12)**, 1979–2001, doi:10.1175/2010JTECHA1356.1.

Straka, J. M., D. S. Zrnic, and A. V. Ryzhkov, 2000: Bulk hydrometeor classification and quantification using polarimetric radar data: synthesis of relations. *J. Appl. Meteor.*, **39 (8)**, 1341–1372.

Testud, J., E. Le Bouar, E. Obligis, and M. Ali-Mehenni, 2000: The rain profiling algorithm applied to polarimetric weather radar. *J. Atmos. Oceanic Technol.*, **17 (3)**, 332–356.

Thompson, E. J., S. A. Rutledge, B. Dolan, V. Chandrasekar, and B. L. Cheong, 2014: A dual-polarization radar hydrometeor classification algorithm for winter precipitation. *J. Atmos. Oceanic Technol.*, **31 (7)**, 1457–1481, doi:10.1175/JTECH-D-13-00119.1.

Tyynela, J., J. Leinonen, D. Moisseev, and T. Nousiainen, 2011: Radar backscattering from snowflakes: comparison of fractal, aggregate, and soft spheroid models. *J. Atmos. Oceanic Technol.*, **28 (11)**, 1365–1372, doi:10.1175/JTECH-D-11-00004.1.

Vivekanandan, J., V. N. Bringi, M. Hagen, and P. Meischner, 1994: Polarimetric radar studies of atmospheric ice particles. *IEEE T. Geosci. Remote Sens.*, **32**, 1–10.

Westbrook, C. D., R. C. Ball, and P. R. Field, 2006: Radar scattering by aggregate snowflakes. *Q. J. Roy. Meteor. Soc.*, **132 (616)**, 897–914, doi:10.1256/qj.05.82.

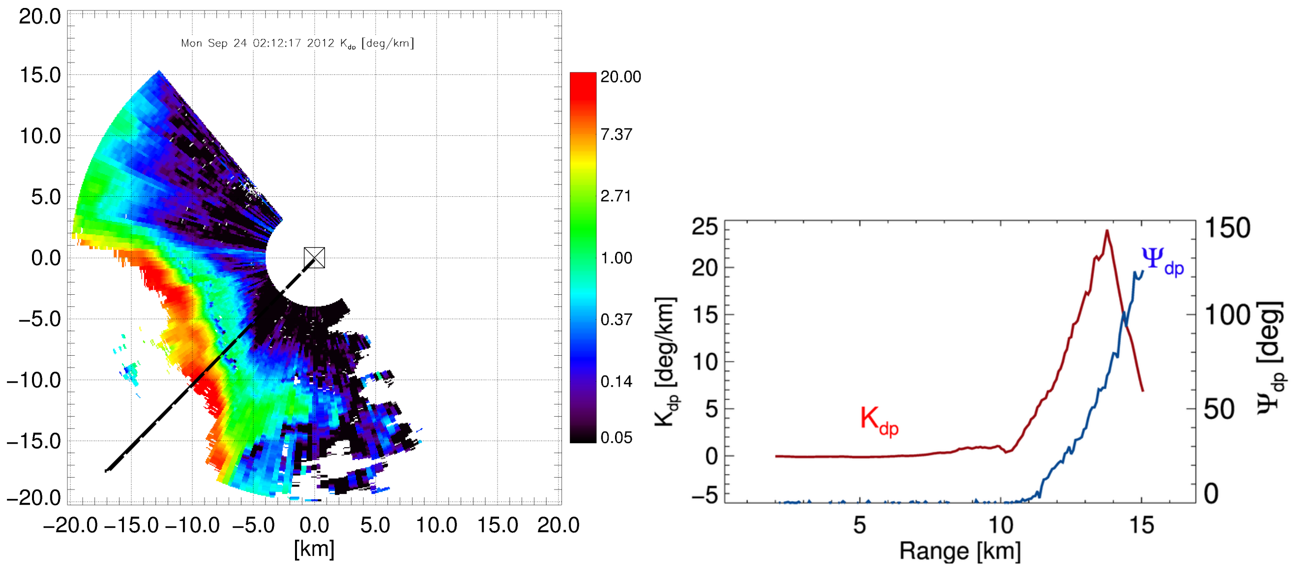

Figure 1: An example of large values of K_{dp} observed in real data. **Left:** PPI collected during the HyMeX campaign on the $24th$ September 2012, 0212 UTC showing K_{dp} [$^{\circ}\mathrm{km}^{-1}$]. **Right:** details of a profile of K_{dp} and Ψ_{dp} extracted from the PPI image along the dashed black line.

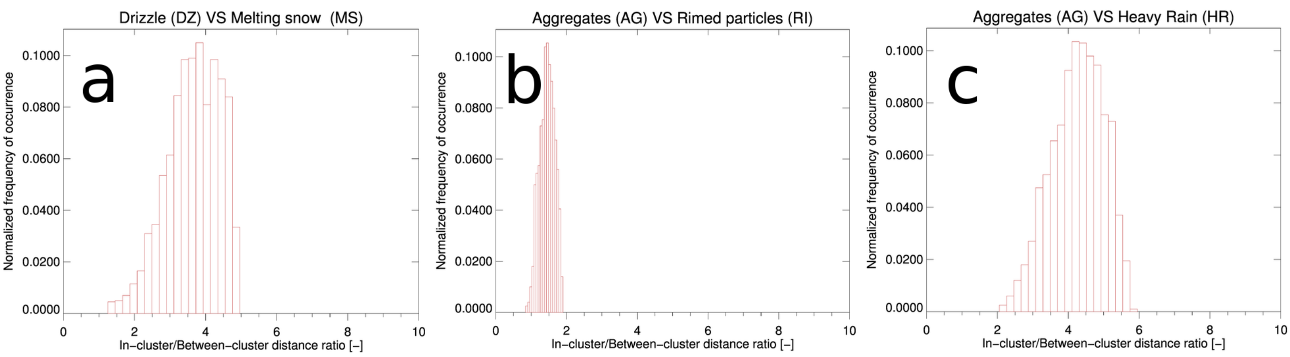

Figure 2: Distribution of the ratio $D(\mathbf{x}, j)/D(\mathbf{x}, i)$ for three couple of clusters. (a): $D(\mathbf{x}, MS)/D(\mathbf{x}, DZ)\,, \mathbf{x} \in DZ$. (b): $D(\mathbf{x}, AG)/D(\mathbf{x}, RI)\,, \mathbf{x} \in AG$. (c): $D(\mathbf{x}, AG)/D(\mathbf{x}, HR)\,, \mathbf{x} \in AG$. The notation is the same as in the originally submitted AMTD manuscript.

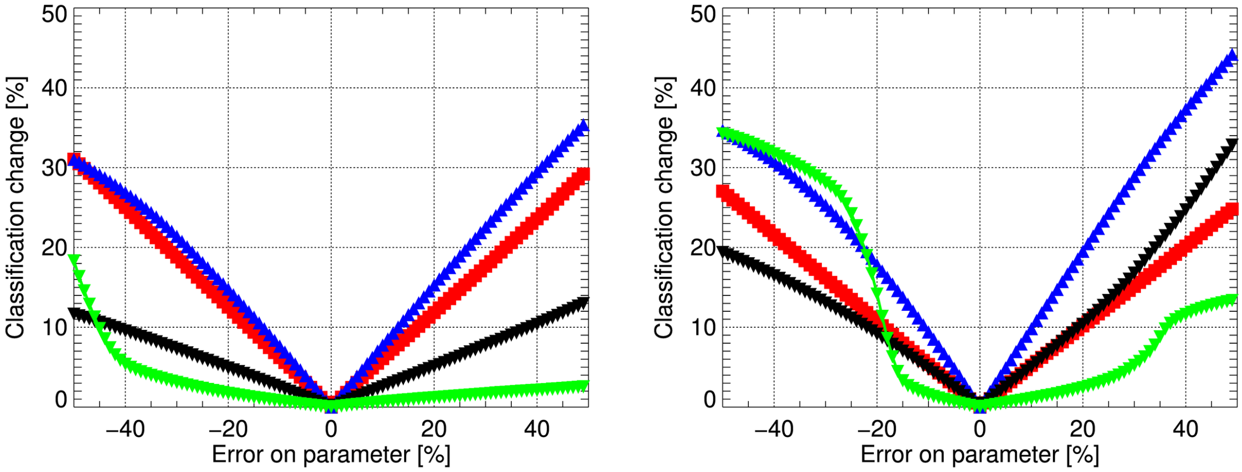

Figure 3: Examples showing the effect of biases on the results of the classification. On the $x-$ axis is shown the percentage of bias on the parameters and on the $y-$axis the percentage of observations that change their cluster (hydrometeor class), in reaction to the bias. **Left:** the parameters are the boundaries of standardization of the polarimetric variables (see Sec. 4.1 in the manuscript). **Right:** the parametrs are the polarimetric variables themselves. Z_H is blue , Z_{DR} is red , K_{dp} is black and Δz is green . ρ_{hv} has an intermediate behaviour between K_{dp} and Z_H and it is not shown here.