Atmospheric Measurement Techniques

amt-2016-105: Responses to Reviewers

Hydrometeor classification through statistical clustering of polarimetric radar measurements: a semi-supervised approach

N. Besic, J. Figueras i Ventura, J. Grazioli, M. Gabella, U. Germann, and A. Berne

Dear reviewers, Dear Editor,

We would like to thank the reviewers for their constructive comments. We did out best to address all their questions in the revised version of the manuscript.

The questions (**M** - major, **m** - minor, **S** - specific) and comments (**C**) of reviewers are reported in italic font. Quoted modifications of the manuscript are highlighted in bold font.

1 Anonymous reviewer #1

Hydrometeor classification through statistical clustering of polarimetric radar measurements: a semi-supervised approach. This paper presents a novel hydrometeor classification algorithm which combines the unsupervised cluster approach and the classical fuzzy logic in an iterative approach using 2 statistical methods: the unsupervised k-medoids clustering and the Kolmogorov-Smirnov statistical test. The paper is well written, the results are clear, the explanations of the method are simple and pedagogic. BUT again the main issue of this article is the validation part which, for me, is not sufficient. My recommendation is to resubmit the paper after Minor revision.

C1: *I have 2 independent points: First of all, perhaps it is a philosophical question or comment. With my experience with operational met services, I think a hydrometeor classification algorithm should answers these points (I will answer on each question):*

- *Is it simple to be implemented? (It is a very clever and nice approach but it is not simple!)*

  Actually, the operational implementation is fairly simple. Namely, the centroids derivation does not have to be done in real time, but is considered as part of the configuration file of the algorithm for each radar. This reduces the operational routine to the pixel assignment part, which comes down to the calculation of Euclidean distances.

  Page 10: **Once we have obtained the set of centroids characterizing a particular radar, the operational implementation of hydrometeor classification comes down to the calculation of Euclidean distances in a five dimensional space, formed by four parametric radar parameters and one external parameter (Ind). The configuration file containing centroids is supposed to be updated at regular intervals of several months in order to account for potentially occurring systematic errors in radar measurements, with the prospective of making this operation continuous.**

- *Is it fast, i.e. can we use it in operational real time? (Using iterations, convergence test . . . with super computer and some paralleling it will be manageable).*

  The calculation of distances in the Euclidean space, which is actually done in operational real time, is reasonably fast, even with a standard desktop PC. Therefore, paralleling is not really needed. In fact, the presented method is currently being implemented in the MeteoSwiss Rad4Alp network processing chain, and is already running in real-time on mobile X-band radar with limited computational resources.

  Page 14: **An interesting remark, related to the operational implementability, is that the calculation of the semi-supervised classification of the illustrated reconstructed RHI along with its measure of entropy (Fig. 11e and Fig. 11g), takes 6 times less than the calculation of the corresponding fuzzy logic counterpart.**

  Page 14: **...An advantage of the proposed method however remains its computational efficiency: in this particular case, which assumes classifying a 7×7 volumes around the station of interest, the semi-supervised method output is obtained in average 4.6 times faster.**

  Page 16: **The presented method is already running in real time on a mobile X-band radar (DX50), while the operational implementation in the processing chain of Rad4Alp network is under progress.**

- *Is the results are on PPI or on ground, i.e. do we have classification on what is falling on ground? (It is more on PPIs or RHI)*

  The classification output is volumetric. Namely, we classify all the observed radar sampling volumes and, as for the moment, do not extrapolate measurements down to the ground level.

- *Is the algorithm is validated using REAL captured cases? (Well, not really)*

  Though the actual validation of the hydrometeor classification (HC) is an exceedingly complex topic, the algorithm was evaluated through the comparison with the selected supervised and (or) unsupervised classification prototypes. This was done both directly (with hydrometeor types measured at the ground level by a 2DVD), or indirectly (e.g. using Probability of Hail (POH) product). In all cases it provided equal or better results.

- *Does it take the measurement errors into account (attenuation, SNR . . .)? (No)*

  Actually, it does take into the account systematic measurement errors, which is the core idea behind the semi-supervised approach. Concerning random measurement errors, as it is the case with any use of radar measurements, we minimize them in the processing part. A comparison with the conventional method (e.g. Fig. 11 in the article) shows a higher degree of robustness with respect to very typical measurement errors, like the ones cause by differential attenuation.

  Pages 13-14: **The increased presence of the vertically aligned ice could be explained by the reported atmospheric lightning, although a straightforward relation cannot be assumed (Hubbert et al., 2014; Roberto et al., 2016). In terms of vertical ice detection, a comparison with the conventional (fuzzy logic) approach (Fig. 11b) could serve as an indicator of a certain robustness of the semi-supervised method with respect to the differential attenuation. That is to say, despite the reported atmospheric lightning, after analyzing Z_H it seems more plausible that the observed negative Z_{DR} (Fig.11b) is partly a result of the differential attenuation, and therefore should not be labeled as vertically aligned ice.**

  The same thing can be observed at X band, in Fig. 9.

- *What is the sensibility of the algorithm when the measurement errors increase? (Not shown)*

  Diagnosing the incrementation in measurement errors is supposed to lead to the recalculation of centroids, which ought to ensure the stability of the HC performance.

- *Can be easily adjusted, i.e. adding or removing hydrometeor types? (So far, yes).*

  Indeed, new hydrometeor types can be easily added. This sort of flexibility is one of the important properties we would like to highlight. Furthermore, the assumptions about the polarimetric properties of different hydrometeor types are, in fact, one of the inputs. Therefore, any advancement in simulating or observing these properties should induce improvement in the HC performance.

  Page 5: **Nevertheless, this selection of classes is not mandatory, because the proposed approach can be used with any set of hydrometeor classes i.e. new classes can be added quite easily.**

- *Is there any long term statistical verification study (Scores)? (Not really)*

  Indeed, we did not yet performed a long term statistical verification study. The idea of this paper is to present the method along with some illustrations of its potential performance, with the latter being a sort of limited validation. As emphasized in the perspectives, a long term comprehensive statistical verification will be the subject of subsequent publications.

  Page 16: **The presented method is already running in real time on a mobile X-band radar (DX50), while the operational implementation in the processing chain of Rad4Alp network is under progress. This constitutes the foundation for an envisaged longterm statistical evaluation of the method.**

C2: *The second philosophical point: I hope it is not offensive, I am not trying to be pessimist or difficult, but for me it is really important. I always enjoy reviewing papers when there is a new approach or algorithm presented and here it is the case. BUT in hydrometeor classification, do we really need a* *new algorithm? If I am not wrong this is the fourth or the fifth article from the coauthors about the same topic, is it more important to more focus about what is written in the conclusion and future work (about the challenge of the hydrometeor mixture . . .) or to have reliable verification? Sorry for this point it is just an independent open question, I hope the authors take it as it is; it is ok if they don't answer and again I don't mean to offence or criticize the work, I respect the exceptional effort and the high quality of this paper.*

Indeed, there is a number of HC methods already present in the literature. Some of them are being mentioned in the introduction section of the paper. The motivation behind our efforts to contribute to this family of methods, is intrinsically linked to the idea of accounting for the peculiarities of radar data instead of entirely relying on assumptions about the polarimetric properties of different hydrometeor types. The reviewer will probably agree that our method should not be considered redundant in this context.

This is the second paper of this group of authors on hydrometeor classification based on radar polarimetric measurements, and both involved comparison with ground level measurements from a 2DVD for evaluation.

Back to the paper, my major comments are:

M1: *For the validation, I don't think it is reliable in this paper, for example, the authors validate the algorithm by compare it with an algorithm which was validated using video disdrometers. I don't think it is work like that, I would prefer to see a revalidation of this approach as done for the unsupervised approach.*

Following the reviewer's suggestion, the validation part is reinforced with the comparison with the 2DVD classification.

Pages 12-13: **Nevertheless, we have analyzed the performances of the proposed semi-supervised method, using the same 2D video disdrometer dataset (Davos, 2010-2011). The analysis was done in the same framework as the one used in Grazioli et al. (2015). The hydrometeor labels assigned to the sampling volumes above the DVD are compared to the corresponding labels of hydrometeor populations, as designated by the 2DVD classification method (Grazioli et al., 2014). After appropriate aggregation of classes (CR - crystals, AG - aggregates, RP - rimed ice particle+ice hail/high density graupel), we obtained the confusion matrix given in Table 1 (responses enumeration).**

Table 1: Confusion matrix (semi-supervised method vs. 2DVD).

		Semi-supervised classification		
		CR	AG	RP
2DVD classification	CR	34.7	6.8	0.2
	AG	7.4	16.0	0.5
	RP	21.3	12.1	1.0

The comparison of Cohen's kappa and overall accuracy with respect to the results given in Grazioli et al. (2015), quantified in Table 2 (responses enumeration), shows a slight improvement over the unsupervised and an important improvement over the supervised method. The vertical distance of 400m between the instrument and the lowest radar sampling volume could explain the generally low score values.

The operational radars do not have a sufficient visibility above the location at which

Table 2: Cohen's kappa and overall accuracy.

Method	Cohen's kappa	overall accuracy
Fuzzy logic (Grazioli et al., 2014)	0.08	0.38
Unsupervised (Grazioli et al., 2014)	0.23	0.49
Semi-supervised	0.25	0.52

the 2DVD was deployed, so a comparison to the ground level reference is not possible. Therefore, we rather concentrate on evaluating the detection of hail and liquid precipitation using C-band data.

M2: *I believe that there is tremendous efforts invested on the calibration and monitoring of the network and the probability of radar errors is reduced. But it is not enough to not taking the errors into account in the algorithm. The reason is: what if another operational service wants to use this approach? What if its network doesn't have the same degree of calibration and monitoring?*

The core idea of the proposed method is exactly this sort of adaptability. If the method is applied on data of different quality (different nature of systematic errors), it will be reflected through the positions of centroids. This is illustrated in Fig. 6 of the article.

M3: *I think here it is important to decrease the quality of the data and see what is happened with the results.*

Actually, centroids before and after data corrections (Fig. 6 in the article), show the influence of the data quality on the position of the centroids. We as well enclose the analysis which compares through the matching matrix the results of classification before and after data corrections, with the respective sets of centroids (Fig. 1 responses enumeration). It illustrates some robustness due to the use of centroids adapted to the quality of data.

Page 10: **The comparison with the centroids derived from unprocessed data (before attenuation and noise corrections), illustrate the rather significant influence the processing can have on the position of centroids. Despite this, the matching analysis between the classification performed with the "uncorrected" centroids on unprocessed dataset, and the one performed with the "corrected" centroids on the same dataset after processing, indicates some skills with respect to the data quality (65.3% of observations on the matching matrix diagonal).**

M4: *It is important to validate the hail detection, but it is the simplest way. The detection of hail is simple but the difficulty is to know its size and/or its percentage in the scanned volume. I think the comparison with POH is important but it is not really relevant.*

We agree that a sort of a more continuous classification, being capable of distinguishing between different hail stone sizes, would be very useful and it is actually one of our future objectives. However, at the moment, the successful detection of hail without considering a vertical structure is a relevant indicator of quality.

Nevertheless, once a hydrometeor type is identified, specific processing methods can be applied. For hail size the method proposed in Ryzhkov et al. (2013) could be used.

M5: *I will like to see how fast (in function of the CPU and RAM) the algorithm is.*

As we elaborated in #2.C1, the classification itself is based on simply calculating the Euclidean distance in the five-dimensional space. Therefore, the computational performance cannot possibly be worse than the ones of existing methods.

The method is developed using CSCS (Swiss National Supercomputing Centre). If using only one CPU (2.6 GHz), with the shared RAM memory, the time necessary to compute fuzzy logic classifi-

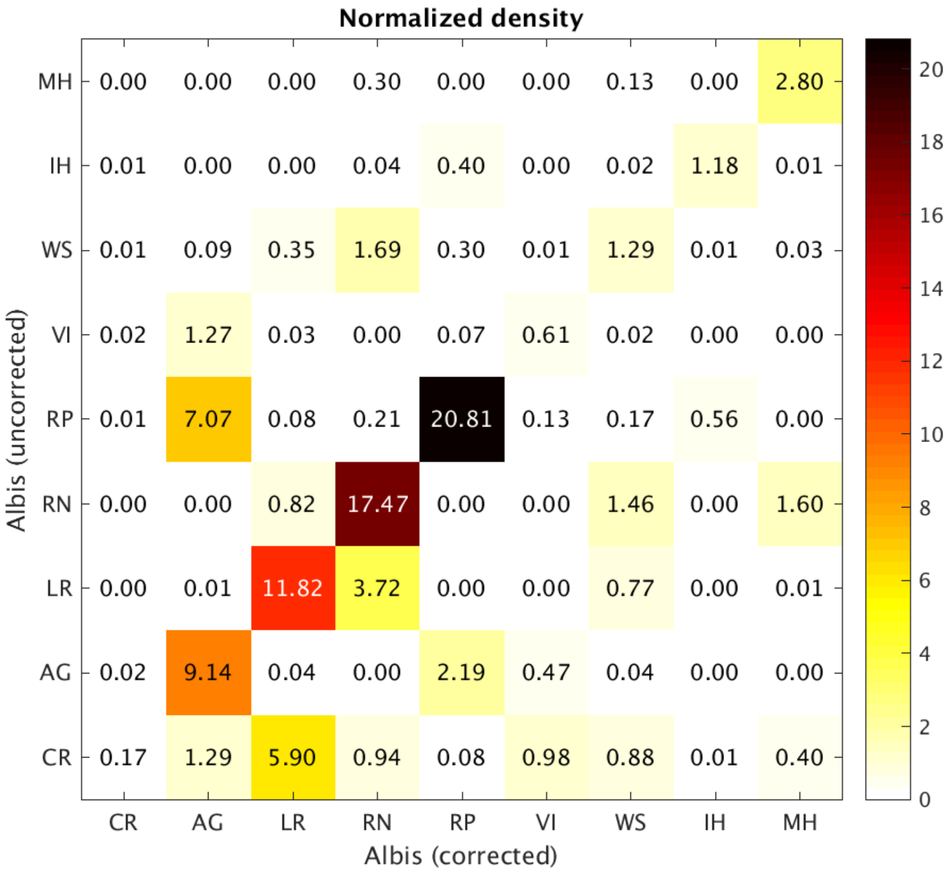

Figure 1: Albis C band radar, 12/06/14, 15h-18h30: matching matrix of classification before and after data corrections, with the respective sets of centroids.

cation shown in Fig. 11e in the article would be 5.4158 seconds, while time required to compute the semi-supervised classification along with the entropy measurement would be 0.8890 seconds.

Page 14: An interesting remark, related to the operational implementability, is that the calculation of the semi-supervised classification of the illustrated reconstructed RHI along with its measure of entropy (Fig. 11e and Fig. 11g), takes 6 times less than the calculation of the corresponding fuzzy logic counterpart.

Page 14: ...An advantage of the proposed method however remains its computational efficiency: in this particular case, which assumes classifying a 7×7 volumes around the station of interest, the semi-supervised method output is obtained in average 4.6 times faster.

Minor points:

m1: *Just for clarification: I know that most of the researcher called the Dual-pol radar: polarimetric radar. This nomination is not really correct, every electromagnetic wave is polarimetric, so simple-pol radar is a polarimetric radar but not Dual-pol!*

Acknowledged, we specified that the nomination polarimetric is used as a synonym for dual-polarization + Doppler.

Page 2: ...incorporating other Doppler dual-polarization (called polarimetric hereafter) parameters...

m2: *P3, L16: Redefine EPFL.*

Acknowledged and done.

Page 3: ... MXPol - belonging to École Polytechnique Fédérale de Lausanne (EPFL), and DX50 - operated by MeteoSwiss.

m3: *P3, L26-27: Expert readers: I prefer to remove this classification of the readers; I think we don't need to know these two statistical methods to be expert or no! If we don't use them before, it is simple we don't know or we forgot them.*

The phrase is reformulated.

Page 4: The proposed semi-supervised algorithm mainly relies on two statistical tools, elaborated in the following subsections: the unsupervised k-medoids clustering and the Kolmogorov-Smirnov statistical test.

m4: *P7, L3: do you have any statistical percentage about the error?*

The calibration accuracy of the Rad4Alp network is elaborated in the paper of Gabella et al. (2016), which was recently accepted for publication and is now included as a reference. In the included table we provide the standard deviation of H/V, H/Sun and V/Sun errors and invite the reviewer to consult the paper for further details:

Page 19: In what concerns the C band data, due to the tremendous efforts invested in automatic calibration and monitoring of the network, we are confident that the probability of radar errors is significantly reduced (Germann et al., 2015). This is demonstrated in the analysis of H/V, H/Sun and V/Sun errors elaborated in (Gabella et al., 2016) and briefly summarized in the following table:

Table 3: Standard deviation of the error, having the Sun radiation as a reference, without noise subtraction.

Radar	H/V (dB)	H/Sun (dB)	V/Sun (dB)
Albis	±0.04	±0.12	±0.10
Monte Lema	±0.07	±0.12	±0.13
Plaine Morte	±0.21	±0.18	±0.07

The same case is with the X band radars, a careful calibration and monitoring being very important due to their very specific research and operational tasks.

m5: *P7, L4-8: for the information from the model, is it one altitude by radar or it is a full layer?*

The exact altitude is used. It is the case both when there is a possibility to estimate the melting layer by means of radar (stratiform precipitation), and where we adopt information from the model (re-sampled for each radar bin).

m6: *P12, L5: Do you have any scores for this test?*

Score is illustrated in Figure 10d, with the average value now also indicated in the text.

Page 13: Quantification of the comparison over a period of time (averaged over 1°, 1.6° and 2.5° elevations), through Heidke-Skill score (ΔHSS=HSS(POH,SS)-HSS(POH,FL)), shows an advantage of the proposed method in terms of matching, during the span of the convective event (average value of ΔHSS=0.05).

m7: *P12, L28: "reasonably simple", here I think we don't have the same reason!*

In case of implementation, it is indeed reasonably simple, which is hopefully now more obvious in the revised version. The presentation of the algorithm in the form of the "off-line" and the "on-line" part, should clarify the simplicity argument.

2 Anonymous reviewer #2

General comments

The paper shows an hydrometeor classification algorithm (HCA) based on a semi-supervised approach using statistical clustering and fuzzy logic method. An iterative approach examines clusters of polarimetric observations by compared them to fuzzy logic membership function. The comparison is performed using the Kolmogorov-Smirnov test for each variable. Each set of parameters is depicted in a set of centroids which are employed in operational labeling of different hydrometeors. The methods is applied to C-band and X-band radar measurements of operational MeteoSwiss radars and research radars. This classification algorithm is in the frame of the new concept of HCA in which the classification is not based exclusively on physical assumption such as in Fuzzy Logic scheme but statistical features are exploited. In general this paper is interesting and well written. However, there are some major points, mainly in results and validation section which need to be extended and improved, furthermore other analysis should be added to test and to validate the methods. My recommendation is to resubmit the paper after MAJOR REVISION.

Major comments

The statistical classification method is described in detail for which the most of the paper is spent (about 7 pages). Unfortunately the performance of the method are not adequately shown: for the results and validation a very short part (about 2 pages) is devoted. Results showed in the manuscript do not convince me that this method is better than other available classification method (supervised and unsupervised). Please discuss the following points.

The paper is conceived to present a novel approach and illustrates through diverse examples both the obvious and the potential advantages of this idea. Following the concerns of the reviewers, the validation part is updated in the revised version, although the longer term comprehensive validation will be investigated in future work.

M1: *At C-band the sampling of measurements are restricted in range between 3 and 40 km (at Xband the range is not specify). This choice is not adequately justify in the text (lines 27-28, page 5). What happens to the classification over 40 km when the sampling volume is larger?*

The maximum range of mobile MXPol band radar is 30-35 km. Therefore, all the representative observations belong to this range. In case of DX50 radar, it would be $\sim$ 50 km.

The motivation for restricting the observations to less than 40 km range comes down to the issue of the possibility to detect weaker echoes corresponding to crystals and vertically aligned ice. Namely, the noise threshold increases with range and at the 40km distance from the radar it is already impossible to detect anything below $\approx$ 7 dBZ, which would make fairly difficult obtaining platykurtic distribution of representative observations, the latter being supposed to ensure the equal presence of different hydrometeors. As well, by considering relatively low range in centroids derivation part, we minimize the proportion of hydrometeor mixtures (smaller radar sampling volumes).

On the other side, the centroids derived after including data $40 < R < 80$km do not differ significantly, as it can been in the Fig. 2.

The probability measure is replaced by the entropy measure in the revised version, with a new section dedicated to the latter.

Page 11: **The obtained classification map is accompanied by a corresponding entropy estimation. The entropy indicates a level of uncertainty with which a hydrometeor class is assigned to an observation. Lower entropy reflects significant confidence in the**

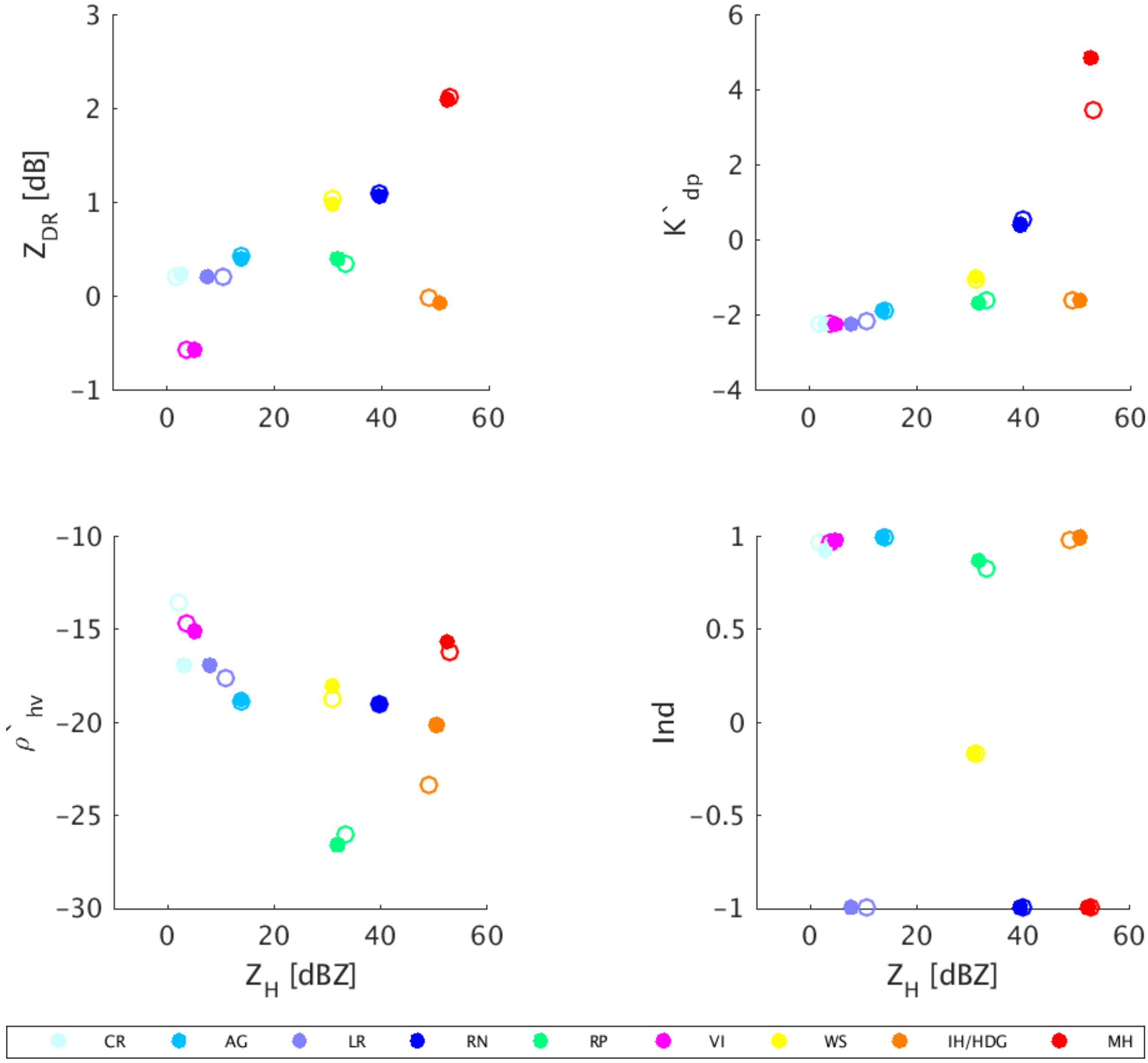

Figure 2: Albis C band radar: comparison of centroids obtained using $R < 40$km range (full), and $40 < R < 80$km range (empty).

assigned hydrometeor label, while higher entropy should be interpreted as an uncertain decision and a potential mixtures indicator.

Out of Rényi's entropies (Rényi, 1960), we have chosen the min-entropy as a measure of uncertainty. The distances (arrow lines in Fig. 8) of the observation with respect to all centroids are converted to the respective probabilities assuming exponential distribution (emphasizing lower distances):

$$p_i = 3\exp(-3d_i), \quad i = 1, \cdots 8(9), \tag{1}$$

which are further on normalized so that $\sum_i^{8(9)} p_i = 1$. Finally, min-entropy is calculated as:

$$H = -\log_{8(9)} p_{max}, \tag{2}$$

with p_{max} corresponding to the distance with respect to the nearest centroid (bold arrow line in Fig. 8).

Page 12: The entropy estimate illustrated in (Fig. 9d) shows the potential of this parameter to detect radar sampling volumes without dominant hydrometeor type (hydrometeor mixtures): particularly elevated values above the detected melting layer.

As we increase the range the uncertainty rises.

Page 14: **As it was the case with Fig. 9d, Fig. 11g shows an increase in the entropy estimate with the increase in sampling volume size, confirming the potential role of this parameter in hydrometeor mixtures identification.**

M2: *In this work a unique methodology that should be applied to all the datasets is not identify to validate the method, somewhat validation is sparsely presented and different feature is validate in each dataset, such as spatial homogeneities, POH and RR. In particular, spatial homogeneity in term of coefficient C (never defined. In the equation (12), C(p,q) is the cooccurrence matrix) between FL, unsupervised and semi-supervised methods (table 2) is shown for one RHI at X-band. Although the number of observations in a RHI should be enough for the statistical point of view, while one RHI could not be enough to support the spatial homogeneity validation. Furthermore, C values (in table 2) shows better performance for unsupervised method respect to semi-supervised, the explanation of this point is not clear (line 19, pag 11). I suggest to extend this type of validation to other classification map also at C-band and to find a common validation for all datasets.*

The spatial homogeneity in terms of coefficient C is now explicitly defined. We have added the generalization of the analysis based on this feature at X band (average over a series of RHIs) and a similar measure between FL and SS at C band (over the convective event illustrated in Fig. 10 and 11).

Pages 14-15: **The comparison with the output of a fuzzy logic algorithm which uses similar membership functions to the ones employed in constraining our clustering, was quantified using the spatial homogeneity feature, derived from the co-occurrence matrix:**

$$C(p,q) = \sum_i \sum_j \begin{cases} 1, & \textbf{if } I(i,j) = p \textbf{ and } I(i \pm 1, j \pm 1) = q, \\ 0, & \textbf{otherwise}, \end{cases} \tag{3}$$

with i, j being the position indices and p, q pixel values (in our case number of a label - from 1 to 9) (Haralick et al., 1973). It is actually a measure of the co-occurrence matrix diagonality:

$$SH = \sum_{p,q} \frac{C(p,q)}{1 + |p - q|}, \tag{4}$$

estimated at X band for two RHIs (226.8° and 316.7°), during the event around the instant illustrated in Fig. 9 (14h-16h). As illustrated in Table 4 (responses enumeration) there is a non-negligible increase in averaged spatial homogeneity with respect to fuzzy logic and very small decrease with respect to the unsupervised approach which nevertheless contains spatial information and has one class less (advantageous for this sort of comparison).

Table 4: MXPol X band rada, comparison of semi-supervised approach with its supervised and unsupervised counterpart: spatial homogeneity.

Freq. band	Method	Spatial homogeneity (SH)
X	Supervised	0.8037
	Unsupervised	0.8795
	Semi-supervised	0.8748

The spatial homogeneity analysis in an analogous framework was applied in case of C-band data, as well. Namely, the spatial homogeneity scores (Eq. 15) are derived for the reconstructed RHIs around 187° azimuth angle, presented in Fig. 11 ($186^\circ - 188^\circ$), for each 5 minute instant during the event presented in Fig. 10 (15h-18h30). To ensure

a fair comparison (equal number of classes), LDG and HDG classes in the fuzzy logic approach are merged to one. After averaging, the results show an advantage of semi-supervised method over its bin-based counterpart (Table 5, responses enumeration).

Table 5: Albis C band radar, comparison of semi-supervised approach with its supervised counterpart: spatial homogeneity.

Freq. band	Method	Spatial homogeneity (SH)
C	Supervised	0.6506
	Semi-supervised	0.7179

Aside from the comparison in the context of spatial homogeneity, which covers both C and X band, due to the practical limitations (availability of data) we had to opt for different means of illustration and validation with different datasets. However, in our forthcoming measurement campaign (which should take place in the canton of Valais), we plan a more careful placement of our ground instruments (disdrometers, 2DVD and MASC) in order to be able to use them both with the mobile X band radar and with the nearby C-band MeteoSwiss radar - Plaine Morte.

M3: *The comparison of classification maps obtained from different methods are shown with not identical hydrometeor labels (see Fig.8 and Fig.10) this does not allow a good comparison. In particular in Fig. 10 it is used the Fuzzy Logic scheme of Dolan et al (2013), the original version of this scheme classify 10 hydrometeor classes including the class "Melting hail and large drops" that is not used in this work. Why do you eliminate this class? In Figure 10 that shows the POH comparison Melting hail of Brenda et al (2013) can match your melting class. In this frame I suggest to show classification maps from different methods with the same labels, when is possible. The classes that are not identical matched are need to adequately describe in term of microphysical properties and the best match to a similar class need to be find in order to appropriately comparing the results. I guess that using similar or identical hydrometeor classes comparison results in particular Fuzzy Logic comparison in Fig.10 could be improved...*

As suggested, the class big drops/melting hail is included in the comparison (Fig. 10 and 11 in the article) in the revised version of the article. This change indeed decreases the quantitative advantage of the proposed method in terms of ΔHSS. As for the other classes, there is a clear correspondence.

M4: *Is the semi-supervised algorithm need to be run for each dataset of observations? Is it possible to use this method to obtain an operational product? How do you implement an hypothetical chain of radar processing that include semi-supervised classification?*

After the algorithm is being run on the representative set of observations for a particular radar, the positions of derived centroids are used in a configuration file for that radar. The classification is included in the chain of radar processing through a fairly simple script which assigns a hydrometeor type label by comparing the Euclidean distance of different centroids with respect to the concerned radar volume. The configuration file is supposed to be updated at regular intervals of several months in order to account for potentially occurring systematic errors in radar measurements.

Page 10: Once we have obtained the set of centroids characterizing a particular radar, the operational implementation of hydrometeor classification is reduced to the calculation of Euclidean distances in a five dimensional space, formed by four parametric radar parameters and one external parameter (Ind). The configuration file containing centroids is supposed to be updated at regular intervals of several months in order to account for potentially occurring systematic errors in radar measurements, with the prospective of making this operation continuous.

Page 16: The presented method is already running in real time on a mobile X-band radar (DX50), while the operational implementation in the processing chain of Rad4Alp

network is under progress.

M5: *Microphysical description of the hydrometeor classes is completely omitted. Is not clear in which way the results of Dolan and Rutledge (2009) and Dolan et al (2013) are adopted and modified. How the MBF shown in Appendix A (table 3, 4, and 5) are obtained?*

Appendix A has been elaborated in the revised version, including the rationales behind the modifications of the referent membership functions.

Pages 16-17: **Wet snow class for X band, not defined in Dolan and Rutledge (2009), was adopted from Grazioli et al. (2015).**

Ice and melting hail classes for both considered frequency bands were defined using scattering simulations with a single and a double layer T-matrix method (Mishchenko et al., 1996), as indicated in Section 3. The same simulator was used for modifications introduced in crystals and vertically aligned ice classes. The modified parameters are emphasized using bold font in Tables 6 and 7. The rationales behind the modifications are the following:

- **CR and VI: The very broad Z_H distributions of these two classes (e.g. reflectivity of vertically aligned ice in C band ranges roughly from -26 to 24 dBZ) are in discordance with our scattering simulations (indicating significantly more contracted ranges). This could be a consequence of the considered particle size distributions. Basically, in order to avoid an important impact of such broad assumptions on the convergence behavior of the centroids derivation algorithm, we decided not to account for very small crystals (almost undetectable at C band) and very big ($D > 30$ mm) dendrites.**
- **IH and MH: Given the specific nature of the IH class, which is directed by the fifth parameter to the high altitudes, this class has more specific properties with respect to the conventional hail class (insignificant liquid water content). Consequently, there is a need for the complementary MH class.**

With the exception of the modifications stated above, the microphysical properties of the classes correspond to the definitions found in (Dolan et al., 2013; Dolan and Rutledge, 2009; Grazioli et al., 2015).

M6: *At page 6 lines 1-10 a selection of observations is made constraining in determinate ranges each parameter with a platykurtic distribution. If we need to classify all the observations of a map without the possibility to select data (for instance in operational products), how could you take into account the hydrometeor classes that are scarcely populated? For example, hail class in initial stages of thunderstorm development could be detected in few pixels and maybe a cluster of this class could not be found. This point should be critical if you consider large sample volume (over 40km of range).*

As indicated in the answer to the #2.M4, and made clearer in the revised article, the selection of the data is not being done in real-time (operationally). Therefore, there is no risk of failing to include all the classes of interest, because there should be always a possibility to select data from the vast database of measurements. In the worst case scenario, if the new set of centroids is supposed to adapt to the recently determined systematic error, the database is limited to the observations since the systematic error started occurring.

M7: *Key figures (from 7 to 11) that should demonstrate the validation and the performance of this classification methods are scarcely commented and descript. In my opinion the results showed are few and weak.*

As suggested by the reviewers, in the revised version these figures are commented in a more appropriate manner and the performance analysis is reinforced, primarily through the inclusion of the comparison with 2DVD-based classification of ice-phase hydrometeors, but also through a sort of generalization of results presented in the original version of the paper (e.g. spatial homogeneity, robustness analysis).

The aim of the paper remains introducing the concept, and illustrating its plausibility, leaving the systematic long-term validation as one of the objectives for the future. However, we are pretty confident that the presented measures of plausibility (validation), are not less convincing with respect to the ones found in the referenced publications. This is particularly the case about comparisons with other methods, which are not very often found in hydrometeor classification papers.

Minor points

m1: *Pag 4 lines 10-19: What is the implementation used in this work? Is the third implementation (line 18) since the observations of dataset selected is very large? Please detail this point in the text.*

Implementation depends on the size of the set. Therefore, we use all of them (iterative splitting is making sets smaller).

m2: *Pag 11 lines 11-16 The 9 classes used is "ispired by Dolan and Rutledge (2009) and Dolan et al (2013)". What do you mean with the term "ispired"? Do you use their MBF? Do you run simulations with their parameters?*

We use their MF's, altered by results of our own simulator in case of altered microphysical properties of the classes (please refer to #2.M5).

m3: *Pag 11 lines 21-26 How many full scans did you consider in total for the 20 days selected for the three C-band radars (8+8+4 days)? The number of observations in table 1 should be greater if you consider 280 full scan per day for 20 days.*

The total number of observations is reduced by a sort of a constrained sampling, which ensures quasi platykurtic distributions of Z_H and ΔH (Fig. 3).

m4: *Pag 7 lines 15-16: The phrase "The value of N is set to nine which corresponds to the number of hydrometeor classes we eventually seek (see section 3), though a different value does not alter the convergence of the algorithm." is not clear. Please rewrite it.*

Acknowledged and done.

Page 7: **This is done by means of the k-medoids clustering algorithm, which divides the set of representative observations into N initial, distant sets, by using the standardized Euclidean distance as a criterion. The number of initial sets is set to nine which corresponds to the number of hydrometeor classes we eventually seek (see introduction to section 3.1). A different number of initial sets does not alter the convergence of the algorithm.**

m5: *Pag 7 line 26 In the phrase "The resultant test statistic is finally compared with the threshold defined by a chosen test significance..." what is the threshold?*

The threshold is a critical value, defined by a hypothesis test. It is the value of the test statistic for which the p-value equals the test significance α.

Page 8: **The resultant test statistic is finally compared with the critical value defined by a chosen test significance (α) and a number of samples (β), following Pearson and Hartley (1972).**

m6: *Pag 9 line 27 In the sentence "The results a X band match to a significant degree those obtained unsupervisely derived centroids (Fig.8)" what "results" do you refer to?. Please replace "(Fig.8)" with "(Fig.8a)". The comment to Fig.8 is not sufficient to describe the results obtained.*

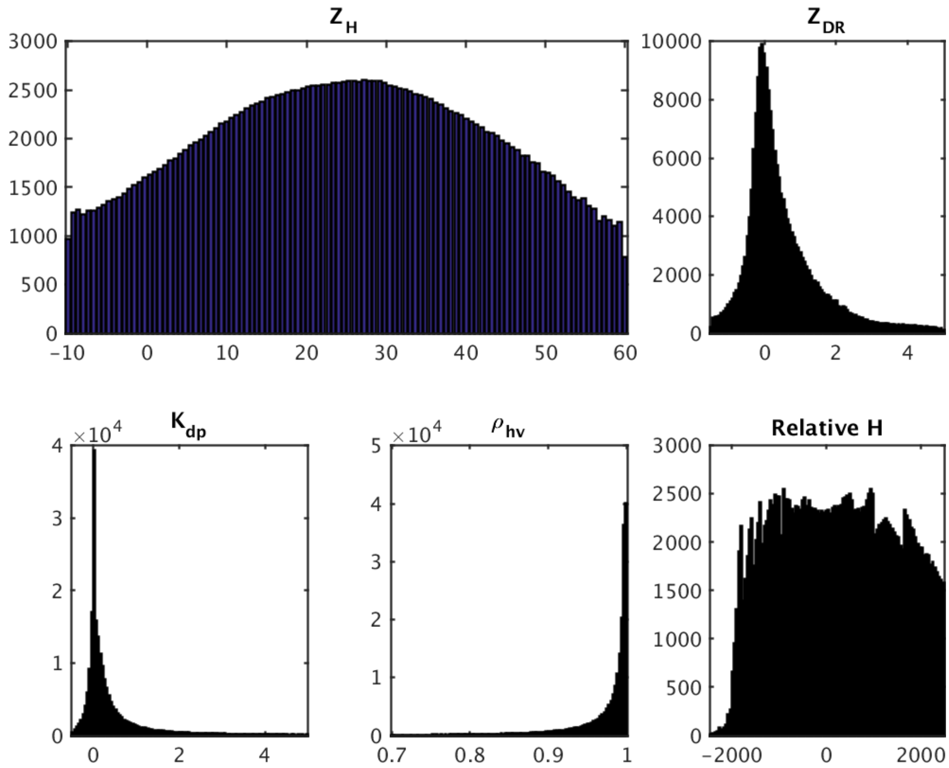

Figure 3: Representative observations acquired by Monte Lema, C band: distribution of polarimetric parameters.

We refer to the RHI of the semi-supervised classification (Fig. 9c in the revised article).

Page 12: The results of semi-supervised classification at X band somewhat differ from those obtained by the supervised method (54% of matching after the suitable aggregation of classes) and match to a significant degree those obtained by employing the centroids derived from the unsupervised method (83% of matching).

m7: *Pag 11 line 15 Which is the spatial homogeneity feature derived from the co-occurrence matrix? In table 2 C is a coefficient, while in the text is the Co-occurrence matrix. Please clarify this point.*

C is indeed the co-occurrence matrix, while the spatial homogeneity (SH) took the place of the wrongly indicated C in Table 2 (please refer to #2.M2).

m8: *Pag 11 line 22 How is calculated the normalized matching matrix?*

Matching matrix is normalized with respect to the total number of observations.

Page 15: The analysis is performed using normalized matching matrix (otherwise called confusion matrix), which provides information about the matching of hydrometeor labels in the overlapping sampling volumes of two radars. It is normalized with respect to the total number of observations.

m9: *Pag 12 line 4 What is the "classical fuzzy logic approach used"?*

For the purpose of a more rigorous comparison, in the revised article it is the fuzzy logic scheme based on the inference proposed in Dolan et al. (2013), instead of the product inference applied in the first version of the paper. The MFs we have been using are the original ones found in Dolan and Rutledge (2009); Dolan et al. (2013), with the wet snow class for X band as well as the trapezoid functions for the relative altitude with respect to the 0°C level, taken from Grazioli et al. (2015).

Page 18: The employed inference is:

$$\mu_i = \frac{w_{Z_{DR},i}\beta_{Z_{DR},i} + w_{K_{dp},i}\beta_{K_{dp},i} + w_{\rho_{hv},i}\beta_{\rho_{hv},i}}{w_{Z_{DR},i} + w_{K_{dp},i} + w_{\rho_{hv},i}} \beta_{Z_H,i} T_{\Delta H,i}, \tag{5}$$

with three weighting coefficients having values: $w_{Z_{DR},i} = w_{K_{dp},i} = 1, w_{\rho_{hv},i} = 0.75$**. This does not correspond to the weighting coefficients found in the original publications, and was chosen for the purpose of the coherent comparison with the presented method. Namely, these are the weights used in the semi-supervised approach, except for the non-radar (fifth) parameter, whose influence is not minimized in the fuzzy-logic approach.**

m10: *Pag 12 lines 9-10 The method used to identify the convective core of storm (Fig.11) is not adequate. During the convective events ice particles grow due to intense updraft and when reach a certain size precipitate. Hail can be found at ground also as "ice hail" (since no microphysical description of hydrometeor are provided in the text, I suppose that for "ice hail" you mean no wet particle). Furthermore due to strong updraft wet particles can be found also above the 0° C level. Typical radar signature of convective events are the absence of bright band . The two type of hydrometeor that you consider "ice hail" and "high density graupel" follow an horizontal layered scheme typical of stratiform events. In the scheme proposed an hydrometeor that can be found below and above the 0° C is not included, using the scheme proposed in the text is not possible to correctly identify the convective core.*

Actually, the proposed method allows for the"ice hail" to be detected below 0°C isotherm level, as well as it allows for the "melting hail" to be detected above 0°C isotherm. Namely, there is only one parameter (out of five) which "forces "ice hail" above and "melting hail" below 0°C isotherm level, and it has a maximum weight of 0.5 (out of 1) in the decision making. If the "ice hail" indeed occurs below, the four other parameters (particularly Z_{DR} and K_{dp}, both weighted 1 out of 1) will make it detectable. If we have a closer look at the values of Z_{DR} and K_{dp} parameters in the given example, we can see that the presence of ice hail below 0°C isotherm is very improbable.

m11: *Pag 12 line 10. The relation between ice detection by radar and lightning activity is not well shown and described. Fig. 11 are not able to provide the "potential of properly detecting the presence of vertical ice". This is a challenging issue that need appropriate methodology to be addressed. Lightning map (Fig. 11f) that represents 24 hours of lightning in a large area is too general to extract qualitative information to be related to the pseudo RHI observed in a small area in few minutes. In order to clear this aspect I suggest to refer to the recently work by Roberto et al (2016), in which a quantitative relation between ice mass detected by weather radar and lightning activity is found.*

This is rather thought to be a qualitative indicator, which should justify the presence of identified vertical ice. It does not represent a veritable lightning validation. With the help of the suggested referencew (Roberto et al., 2016; Hubbert et al., 2014), this is made clear in the revised version.

Pages 13-14: The increased presence of the vertically aligned ice could be explained by the reported atmospheric lightning, although a straightforward relation cannot be assumed (Hubbert et al., 2014; Roberto et al., 2016). In terms of vertical ice detection, a comparison with the conventional (fuzzy logic) approach (Fig. 11b) could serve as an indicator of a certain robustness of the semi-supervised method with respect to the differential attenuation. That is to say, despite the reported atmospheric lightning, after analyzing Z_H **it seems more plausible that the observed negative** Z_{DR} **(Fig. 11b) is partly a result of the differential attenuation, and therefore should not be labeled as vertically aligned ice.**

m12: *Pag 12 lines 14-18 Why the comparison is not shown for FL classification? The comment of this figure, that is consider one of the key point for validation of the method is scarce and inappropriate.*

Supervised classification = FL classification, as it is clarified in the revised article. This figure is rather meant to be a sort of illustration, given that we're aiming to compare the detection of each hydrometeor type with some other sort of information. Furthermore, as it is indicated in the article, this figure demonstrates very similar performances with the fuzzy logic approach when it's about the light rain/rain discrimination, the processing time being an advantage of the proposed method.

Page 14: **An additional comparison with the corresponding supervised routine concerns liquid precipitation (Fig. 12). Namely, 10 minute rain gauge measurements at two MeteoSwiss stations (in the vicinity of Monte Lema radar) are compared to the rain vs. light rain output of semi-supervised and supervised (fuzzy logic) classification. Although drawing a border line between light rain and rain is indeed somehow debatable, by observing ground measurements, one can perceive a very large plausibility of the results obtained with both methods. An advantage of the proposed method however remains its computational efficiency: in this particular case, which assumes classifying a 7×7 volumes around the station of interest, the semi-supervised method output is obtained in average 4.6 times faster.**

m13: *Pag 13 line 1-5 The performance and the validation of the method are just mentioned in the conclusion. I think that conclusions need to be enriched in the light of more robust results that should be shown in the revised manuscript.*

We did our best to make the validation and the included illustrations of method's plausibility more robust and comprehensible. This part is appropriately emphasized in the revised conclusions.

Page 16: **We illustrated some noteworthy properties of the proposed classification, which concern: discrimination between different solid phase particles, hail detection performance and spatial homogeneity of the classification output. Simultaneously, we demonstrated a very low dependency on non-radar inputs and a significantly enhanced computational efficiency.**

m14: *Table 2 What is the coefficient "C" ?*

C is the co-occurrence matrix. SH - spatial homogeneity (a measure of co-occurrence matrix diagonality) took its place in Table 4 in the revised version (please refer to #2.M2).

m15: *Figure 5 Why any wet class are shown in this plots? For example should be interesting showing PDFs of Rain class.*

The revised version of the article contains the PDFs of aggregates, rain and wet snow. This time both the PDFs derived with the MXPol and the ones derived with the DX50 radar.

Page 9: **A comparative illustration of the PDFs obtained by MXPol and DX50 radars (Fig. 5), shows a fairly similar two-dimensional distributions except in what concerns ρ_{hv} where the obtained PDFs and consequently the representative centroid reflect the systematic underestimation of this parameter by the DX50 radar. This sort of adaptability is the basic idea of the proposed classification method.**

m16: *Fig. 8 Rewrite the caption. Please insert the data and time of radar measurements. The probability map is never mentioned in the text. What is it represent?*

Acknowledged and done. Entropy, replacing probability in the revised version of the article, is defined in #2.M1.

Page 32: **MXPol RHI profile $226.8°$ azimuth, 14h25, 29/09/13, Ardèche, France: (a) Z_H, (b) Z_{DR}, (c) K'_{dp}, (d) ρ'_{hv}; Comparison of semi-supervised approach (g) with its supervised (e) and unsupervised counterpart (f), along with the entropy estimate of the semi-supervised approach (h).**

m17: *Figure 9 What do these comparisons (obtained by normalized matching matrix) show in terms of classification results?*

In the revised version, in the same framework (normalized matching matrix) we present the comparison between the semi-supervised and the fuzzy logic method for a day when one of the radar (MXPol) had an issue with the magnetron. The idea behind these comparisons is to evaluate the matching of classifications of the same volume, using two different radars, which should be an indicator of the robustness of the classification method. The comparisons before and after reflectivity bias correction, are used to demonstrate the stronger reliance of the semi-supervised method on the dual-pol radar parameters.

Page 15: **Finally, we performed the matching analysis by applying both our classification, based on MXPol and DX50 centroids, and the fuzzy logic approach, on two X-band radars pointing toward the same volume (Fig. 13). The peculiarity of the analyzed day is that the MXPol radar had an issue with the magnetron performance, causing a significant reflectivity bias.**

Before bias correction, the semi-supervised method shows better matching performances: 55% **of observations on the diagonal vs.** 50% **in case of the fuzzy logic approach. However, more significant is the fact that after bias correction, the performances of the presented method improve significantly more than those of the considered fuzzy logic approach:** 61% **vs.** 51%**. We deduce that this difference indicates a stronger reliance of the semi-supervised method on the polarimetric radar parameters, whereas the decision of the supervised approach depends significantly on the external (fifth) parameter.**

m18: *Fig 10 What is the ΔHSS?*

As indicated in the caption, ΔHSS is a difference in Heidke-Skill Scores (ΔHSS=HSS(POH,SS)-HSS(POH,FL)). Positive value indicates a better matching of the semi-supervised method with POH product, while negative value point toward better matching of fuzzy logic approach.

3 Anonymous reviewer #3

General comments:

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

C1: *The first concern I have relates to the use of the scattering calculations in the algorithm. There needs to be more information provided to the reader about the microphysical assumptions that went into the scattering calculations (and thus the resulting membership functions), most importantly the aspect ratios and particle size distributions of the hydrometeors. These choices will influence the algorithm and therefore need to be justified.*

Please refer to #2.M5.

Page 17: **In what concerns the choice of the particle size distribution (PSD) and the aspect ratios for the simulated particles, we opted for:**

- **CR and VI: PSD as the simplified (exponential) version of the empirical, temperature (T) dependent, distributions found in Heymsfield et al. (2013):**

$$N(D) = N_0 \exp -\lambda D, \tag{6}$$

with $N_0 = 3.304 \exp(-0.04607T)$, and $\lambda = 15.3 \exp(-0.053T)$ for stratiform and $\lambda = 3.4 \exp(-0.083T)$ for convective events.

Aspect ratio follows the empirical relation $h = aD_{max}^b$, with h being the smallest dimension and D_{max} the largest: spherical ice crystals ($a = 0.9,\ b = 1$), solid thick plates ($a = 0.23,\ b = 0.778$), dendrites ($a = 0.0418,\ b = 0.377$).

- **IH and MH: PSD as defined in Cheng et al. (1985):**

$$N(D) = C\Lambda^{4.11} \exp(-\Lambda D), \tag{7}$$

with Λ varying from 0.1 to 1 mm^{-1} and C varying from 60 to 300.

Aspect ratio as defined by Ryzhkov et al. (2011) (dh - dry/ice hail, mh - melting hail):

$$r_{dh} = \begin{cases} 1 - 0.02D, & D \leq 10mm, \\ 0.8, & D \geq 10mm. \end{cases} \tag{8}$$

$$r_{mh} = \begin{cases} r_{dh} - 5(r_{dh} - 0.8)f_{mw}, & f_{mw} < 0.2, \\ 0.88 - 0.40 f_{mw}, & D0.2 \leq f_{mw} \leq 0.8, \\ 2.8 - 4r_w + 5(r_w - 0.56)f_{mw}, & Df_{mw} > 0.8, \end{cases} \tag{9}$$

with f_{mw} being the mass water fraction, and r_w the aspect ratio of the equivalent raindrop.

C2: *Given that the authors' intention is to minimize the reliance of their HCA on the uncertain microphysical and electromagnetic properties of hydrometeors, the impact of these assumptions on the resulting algorithm needs to be shown. Therefore, I recommend the authors perform some sort of sensitivity analysis where the parameters used in the scattering calculations are modified and resulting changes in the cluster centroids are examined.*

The variation of MF parameters in the external loop increased from 5% to 10% and 20% around the values given in the Appendix. The evoked impact is analyzed by observing the changes in the positions of final centroids and the dispersion of centroids at the output of the internal loop.

The averaged changes in the positions of centroids are given in Table 6, while the averaged changes in the interquartile coefficient of dispersion of centroids are given in Table 7.

Table 6: The averaged changes in the positions of centroids after increasing the variability of MF parameters from 5% to 10% and 20%.

Parameter X	Z_H (dB)	Z_{DR} (dB)	$K`_{dp}$	$\rho`_{hv}$	Ind
ΔX (5%)	0.5276	0.0559	0.1740	0.4150	0.0074
ΔX (10%)	0.6550	0.0924	0.2062	0.7855	0.0092

By observing the obtained results, we can deduce very small resulting changes in the position of the centroids and what is even more important, negligible changes in the convergence of the algorithm. Namely, changing up to 10% the MBF (which are a sort of clustering constraints) from one iteration to another in the external loop, and not introducing any dispersion, demonstrates an emphasized dependence on the

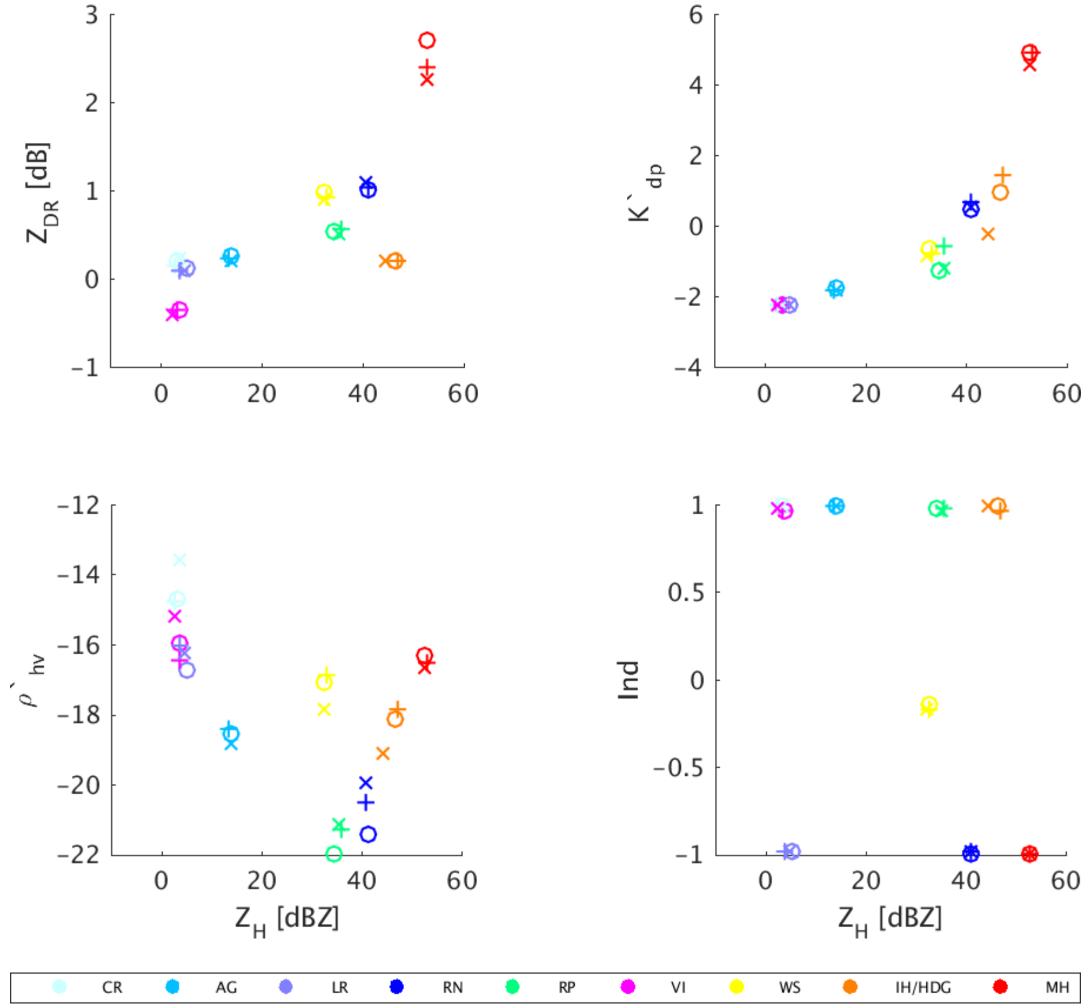

Figure 4: Monte Lema C band radar: comparison of centroids obtained using 5% variations of MF parameters (∘), and 10% (+) and 20% (×) variations of MF parameters.

Table 7: The averaged changes in the interquartile coefficient of dispersion of centroids after increasing the variability of MF parameters from 5% to 10% and 20%.

Class	CR	AG	LR	RN	RP	VI	WS	IH/HDG	MH	Average
Δc (5%)	0.2970	0.0294	0.0776	0.0188	0.0338	0.1337	0.0273	0.2322	0.0168	0.0963
Δc (10%)	0.0698	0.0617	0.0617	0.0287	0.0918	0.0591	0.1132	0.0568	0.0398	0.0647

information contained in the radar data, and thus a decreased dependence on the information coming from the reference observations.

Page 9: **An important remark is that the increase of the variation range of reference parameters in the external loop, from 5% to 20% around their initial values, results in a very small change in the positions of centroids (e.g. average $\Delta Z_{DR} = 0.0924\ dB$), and a negligible change in their dispersion (average $\Delta c = 0.0647$). This clearly indicates a limited sensitivity of the semi-supervised method on the employed reference assumptions.**

C3: *Another concern I have is the lack of discussion about the polarimetric variables used in the algorithm. It should be stated how these variables depend on the microphysical properties of the hydrometeors. It would also help to introduce in this section the hydrometeor types that will be used in the HCA. The degree of uncertainty in the polarimetric radar signatures associated with each hydrometeor type should* *then be discussed to motivate the addition of clustering to the HCA.*

The section 3.1 is reinforced with a concise introduction of polarimetric parameters and a comment about the uncertainty in measurements and polarimetric radar signatures.

Page 6: The values of selected polarimetric parameters can be associated with different microphysical properties of the particles: Z_H - concentration, size and density, Z_{DR} - shape, orientation and density, K_{dp} - concentration and shape, ρ_{hv} - homogeneity. Therefore, information they provide is to a degree complementary, and should lead to the inference of a particle type. However, given that radar measurements are in fact estimates of these parameters, these values are characterized with a certain level of uncertainty, as well as are the presumed polarimetric signatures of different particles. The proposed method is actually designed to adapt to these uncertainties.

C4: *Finally, there are a number of instances of somewhat confusing wording, compounded in some cases by grammatical errors. Addressing these wording issues and grammatical errors will make the points highlighted by the authors clearer. Based on these concerns, I recommend that this manuscript undergo major revisions. My specific and technical comments follow.*

The wording is hopefully improved and grammatical errors eliminated.

Specific comments:

S1: *Page 1, Line 11-12: It is unclear what is meant by "merging" here.*

Bringing together different clusters which are identified as belonging to one particular class.

S2: *Page 1, Line 19-20: Add references to support the uses of hydrometeor classification.*

Acknowledged and done, (Bringi et al., 2007).

S3: *Page 2, Line 16-21: Re-work this sentence or split it into two sentences. Also, define the polarimetric variables mentioned here and describe how they depend on the microphysical properties of the hydrometeors.*

The sentence was split in two sentences. The polarimetric variables are defined in the section 3.1 (please refer to #2C.3).

Page 2: Among the number of methods developed for S, C and X band we can distinguish between: those using reflectivity at horizontal polarization (Z_H), differential reflectivity (Z_{DR}), specific differential phase shift (K_{dp}) and correlation coefficient (ρ_{hv}) (e.g. Dolan et al., 2013), and these using linear depolarization ratio (LDR) (e.g. Straka et al., 2000). Also, we can discriminate between methods using temperature as an external parameter (e.g. Zrnic et al., 2001), rather than relative altitude with respect to the $0°C$ isotherm (e.g. Lim et al., 2005), as well as between those using two or more dimensional membership functions (e.g. Marzano et al., 2007), rather than one dimensional ones (e.g. Liu and Chandrasekar, 2000).

S4: *Page 2, Line 26: Add references and discuss the uncertainty in the polarimetric signatures of different hydrometeor types.*

Uncertainty in the polarimetric signatures of different hydrometeor types is evoked at several places in the manuscript (e.g. section 3.1), though we opted not to elaborate the discussion, already present in the majority of the quoted papers, due to the obvious constraints with space (revised version is already significantly longer with respect to the original one).

S5: *Page 3, Lines 10-12: The authors need to discuss further the differences between their proposed semisupervised method and the method proposed by Bechini and Chandrasekar (2015) (i.e., why the method proposed in this manuscript is an improvement over the previous method).*

Though as well named semi-supervised, the method of Bechini and Chandrasekar, 2015 is conceptually significantly different. Namely, its emphasis is on improving the output of fuzzy logic schema by applying clustering with a significant spatial (regional) constraint, which increases the robustness with respect to the inevitable noise in the measurements.

Rather than demonstrating improvement in performance, we aim to emphasize theoretical and practical advantages of the proposed methodology: we reduce the problem of hydrometeor classification to the pretty intuitive Euclidean geometry in the five-dimensional space, where the reference points in this space (centroids) result from the synthesis of microphysical assumptions and information coming from the data. By doing so, aside from intrinsically accounting for the uncertainty both in the data and in the assumptions, we ensure a remarkable computational efficiency.

Page 3: **The first notion of the semi-supervised method is found in Liu and Chandrasekar (2000), where employing neural networks in updating the membership functions, reduces classification algorithm to a neuro-fuzzy system. However, the most representative for this category would be a region-based method proposed in (Bechini and Chandrasekar, 2015), its emphasis being on improving the output of fuzzy logic schema by applying clustering with a significant spatial (regional) constraint, which introduces an important robustness with respect to the inevitable noise in the measurements.**

Page 3: **This is achieved in a somehow different way with respect to the state-of-the-art semi-supervised approach (Bechini and Chandrasekar, 2015), by involving, in a computationally efficient bin-based approach, two classical data processing tools: k-medoids clustering and Kolmogorov-Smirnov test.**

S6: *Page 4, Line 6: How can the standard deviation for each cluster be a scalar? Shouldn't this quantity be a vector since each polarimetric variable will have a different standard deviation for a given cluster?*

Indeed, there was a mistake in Eq. 2. It should have been:

$$d_i = \left\| \frac{\mathbf{x} - \boldsymbol{\mu}_i}{\boldsymbol{\sigma}_{S_i}} \right\|_2, \tag{10}$$

S7: *Page 4, Lines 11-12: Does the choice of algorithm affect the resulting clusters?*

Please refer to #2.m1.

S8: *Page 5, Line 16: How are the authors sure that no vertically aligned ice was present in the X band radar observations without assuming some signature for this class of particles a priori?*

The interquartile dispersion coefficient serves as the indicator of a very weak presence of vertical ice, in a way that the vertical ice centroids at the output from the internal loop show significant divergence. This is, a posteriori, confirmed by observing very few volumes with negative Z_{DR} and positive Ind in the representative observations dataset.

S9: *Page 5, Line 23-24: Is this sentence implying that the four chosen cases from the Monte Lema radar were selected because more hail was reported during these events then the climatological frequency of hail events in the region?*

Lapsus calami, it should be Plaine Morte radar. Actually not, we wanted to keep the shares of different hydrometeor types proportional.

Page 6: **The reason behind the smaller initial set for Plaine Morte radar (four days) is the lower regional frequency of hail storms and our desire to keep the proportions of different hydrometeors similar for all radars.**

S10: *Page 5, Line 25: Are these plan position indicator (PPI) scans? Please clarify.*

Yes, this point is clarified in the revised version.

Page 6: All three considered operational radars have the same scanning pattern covering the entire azimuth revolution with 20 elevations (Plan Position Indicators - PPIs, from -0.2° to 40°) in 5 minutes, resulting in 288 full scans per day.

S11: *Page 6, Line 15: Why is ΔH the relevant parameter for phase composition of the hydrometeors? Doesn't the degree of melting depend on both the height above or below 0 $\hat{A}^o C$ and the lapse rate?*

Indeed, it depends. However, our idea is to rely exclusively on radar polarimetry when distinguishing between different hydrometeor types inside either liquid or ice phase. Therefore, we apply the sigmoid transformation, which has the aim to diminish the influence of this information inside liquid or ice phase.

S12: *Page 6, Line 24: I suggest adding the number of scans from each radar in addition to the number of pixels to table 1.*

Actually, all scans from 8 considered days for one C band radar are used. After eliminating the non-precipitation volumes, we perform a constrained sampling which results in the given number of pixels, distributed according to the Fig. 3. We could have retain the position of each pixel and state from how many different scans day originate, but given that we rather opt for bin-based approach (without considering neighboring effects), this did not appear as very relevant.

S13: *Page 7, Line 3: Does this include correcting for non-uniform beam filling?*

Unfortunately, the correction of non-uniform beam filling is not performed.

S14: *Page 7, Line 5-6: I am confused by what is being done for the pixel assignment step to get the altitude of the 0°C isotherm. Why use a 6.5°C/km lapse rate when the temperature levels should be available in the model?*

Given that in the stratiform events we have the possibility to rely on the radar estimate of the melting layer height, we have opted to use the altitude rather than the temperature as the input for the derived 5th parameter. Therefore a need to use a standard atmospheric lapse, in order to get the local altitude with respect to the 0°C isotherm for a radar sampling volume for which we already have temperature (after the re-sampling of COSMO temperature profile to the radar grid). Alternative (as well used with the presented method) is to use directly the information about the 0°C isotherm, together with the calculated height of a radar sampling volume.

Page 7: The information concerning the altitude of the 0°C isotherm has been collected from the numerical weather prediction model COSMO (Baldauf et al., 2011), by relying on the 0°C isotherm product or by applying standard atmosphere lapse rate in the troposphere (6.4 °C/km) on the temperature profiles.

S15: *Page 7, Line 7: How well do the 0°C isotherms altitudes derived from the polarimetric melting layer signature correspond to the 0°C isotherms altitudes from the model?*

A detailed description of the performances of the radar based melting layer detection are provided in the referred article (Wolfensberger et al., 2015).

S16: *Page 7, Line 12: Switch order of Figs. 2 and 3; as it is now, Fig. 3 is referenced in the text before Fig. 2.*

Acknowledged and done.

S17: *Page 7, Line 20: This discussion of the scattering calculations is vague. The details of the scattering calculations need to be presented if they differ from the calculations from the Dolan et al. (2013) and Dolan and Rutledge (2009) studies. There is some limited discussion of the scattering calculations in the Appendix that should be expanded upon and referenced in this section of the text.*

Please refer to #2.M5.

S18: *Page 7, Line 28: State the H0 hypothesis being tested.*

Acknowledged and done.

Page 8: Clusters which satisfy the H_0 hypothesis (distribution similar enough to one of the reference classes) exit the iteration as labeled observations, while the rest proceeds to the additional clustering, this time divided into two sets.

S19: *Page 8, Line 5: In general, most radar sampling volumes will contain a mixture of different particle types. In many cases, one of these particle types will dominate the polarimetric variables. Therefore, in this section of the text, it would be better to say that the unlabeled clusters are mixtures of particle types that have more equal contributions to the polarimetric variables.*

Acknowledged and done.

Page 8: On the other side, unlabeled clusters are assumed to be mixtures of hydrometeor types that have more equal contributions to the polarimetric variables and therefore are not further analyzed in this phase of our research.

S20: *Page 8, Line 6-7: What is the justification for limiting the observations to a range of 40 km?*

Please refer to #2.M1.

S21: *Page 8, Line 19: State what the number of samples S refers to specifically in the algorithm.*

Acknowledged and done.

Page 9: Namely, in order to keep the $\alpha\beta$ product quasi-constant, we only slightly vary the number of samples $S \in [30, 35, 40]$ and therefore the β parameter, whereas we keep the low value of the test significance $\alpha = 0.01$.

S22: *Page 8, Line 30: Clarification: these 30 centroids come from the 30 iterations of the external loop of the algorithm?*

Exactly.

S23: *Page 11, Line 7-9: Rework this sentence.*

Acknowledged and done.

Page 12: Here, by using the simpler k-medoids clustering method and without considering at all the texture information we are obtaining very similar results. Further more, the identification is performed automatically, modifiable theoretical assumptions are used at the input, and more classes are derived.

S24: *Page 11, Line 8: Determining the sensitivity of the classification algorithm to the assumptions in the scattering calculations/membership functions might be another way to illustrate the differences between semisupervised and unsupervised methods. Doing this would also illustrate just how dependent the semisupervised scheme is on the scattering calculation assumptions.*

Please refer to #3.C2.

S25: *Page 12, Line 6: It would be useful here to discuss the skill of the POH algorithm.*

Instead of elaborating the performances of the POH algorithm, we prefer referring the reader to the very recent and very systematic paper (Nisi et al., 2016).

S26: *Page 12, Line 10-11: Does the algorithm identify vertically-oriented ice during all cases where lightning is observed? One example where vertically-oriented ice with lightning is detected is not sufficient to suggest that vertically-oriented ice is actually present. Observations of depolarization streaks in ZDR may be another way to support the presence of vertically-oriented ice during this period. See the paper by Hubbert et al. (2014).*

Indeed, this is rather meant to be an illustration. The suggested article (Hubbert et al., 2014) is used to clarify the point. Please refer to #2.m11.

S27: *Page 12, Line 29: The microphysical assumptions, especially the assumed particle size distributions and aspect ratios used to generate the membership functions, need to be revealed either in section 3 or the Appendix. It is also important to summarize how these assumptions affect the centroids, given that the proposed method should not be overly dependent on the scattering calculations.*

Please refer to #2.M5 and #3.C2.

S28: *Page 13, Line 15: How does considering only small sampling volumes eliminate hydrometeor mixtures? Microphysical situations where mixtures of hydrometeors are likely (e.g., aggregates and pristine ice particles, rimed and unrimed ice particles, melting hail and rain) are still going to be encountered by the radar, regardless of the size of the radar sampling volume.*

Indeed, our entropy parameter potentially suggest the same thing. Please refer to #2.M1 for further explanations.

The last sentence is appropriately rephrased.

Page 16: Therefore, the plan is to deal with the volumes characterized with high entropy (potential hydrometeor mixtures), either through their decomposition or through defining a new set of mixed classes, dominantly for far ranges.

Technical corrections:

- *Page 1, Line 5: Remove "eventually."* ✓
- *Page 1, Line 13: Remove "finally."* ✓
- *Page 2, Line 26: Remove "whereas" and split this sentence into two sentences.* ✓
- *Page 2, Line 32: The word "distant" should be "distinct."* ✓
- *Page 3, Line 1: Remove "whereas" and split into two sentences.* ✓
- *Page 3, Line 4-9: Split this sentence into multiple sentences.* ✓
- *Page 3, Line 13: Remove "whereas" and revise.* ✓
- *Page 5, Line 26: Change "pointed" to "indicated."* ✓
- *Page 6, Line 10-11: Remove parentheses and change "precipitations" to "precipitation."* ✓
- *Page 6, Line 27: Remove "in particular."* ✓
- *Page 8, Line 1: Revise this sentence.* ✓
- *Page 10, Line 2: Change "pixels" to "pixel."* ✓
- *Page 10, Line 28: Change "unsupervisely derived centroids" to "the centroids derived from the unsupervised method."* ✓
- *Page 11, Line 7: Change "unsupervisely."* ✓
- *Page 13, Line 23: Change "was" to "were."* ✓
- *Figure 2: Add yes/no text to decision branches in the flow chart.* ✓

- *Figure 6: I recommend transforming $K`_{DP}$ and ρ'_{hv} toK_{DP} and ρ_{hv} before plotting to make their interpretation easier. Change "CS" in the caption to "CR."* ✓

  Actually, we decided to illustrate in Fig. 9 and 11 the $K`_{DP}$ and ρ'_{hv} rather than K_{DP} and ρ_{hv}, in order to facilitate the interpretation of the resulting classification.

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

- *Figure 11: Change the color bars of panels b and c so that the negative ZDR and KDP values are distinguishable. Also, make the +, -, and x symbols in panel f larger.* ✓

  Unfortunately, we could not make the symbols +, - and x larger, cause we cannot modify the output of ©Météorage product.

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

4 Anonymous reviewer #4

General comments:

M1 *This paper illustrates a semi-supervised technique for the hydrometeor classification from dual-polarization radar observations. The authors suggest that by introducing a degree of (unsupervised) adaptability by means of K-medoid cluster analysis, the classification can be improved upon standard supervised techniques relying on fuzzy-logic. In general I enjoyed reading this manuscript and I liked the idea proposed. In particular the implementation of the double loop for the cluster analysis with "successive refinements" in combination with Kolmogorov-Smirnov test appears a clever solution for the realization of the semi-supervised method. The technique is in general well described, although some parts need further clarification (see specific comments below). In addition, the generalization of the technique is affected by few arbitrary assumptions, in particular regarding the choice of the distance measure, which is different for the cluster analysis and for the pixel assignments (there may be a reason, but I did not got it). Another issue in the technique description is related with the choice of the membership functions. These are said to be "appropriately modified and enriched by means of scattering simulations based on double layer T-matrix method". This needs to be discussed with more detail, indicating how the electromagnetic scattering simulations are performed and which hydrometeor classes have been modified and how.*

The reason of not opting for the standardized Euclidean distance also in the pixel assignment is the dependence of the result on the amount of pixels considered (via estimated standard deviation). Basically, the result of the classification would differ if we consider the entire volume or only one RHI profile.

Concerning the modification of the membership functions, please refer to #2.M2.

M2 *Although the description of the method can be easily improved, my major concern is about the validation, which is actually quite weak. For the C-band classification in particular the discussion is superficial and I do not really understand how a single polarization product (hail) could be used to validate a multiple-parameter classification (which is expected to provide better information about the cloud microphysics). In addition, only the classification results are presented, preventing a comprehension of the reasons for the very different results using the fuzzy logic and the semi-supervised approach (figure 10). Fig. 12 does not really add a significant contribution to the validation (light vs. rain distinction) and I suggest to drop it. On the other hand, the X-band discussion appears much* *more meaningful to me and I suggest expanding this part. In particular, the results illustrated in fig. 9 help showing the reliability of the classification (not properly "validation") should be expanded (also showing the radar dual-polarization moments). In particular you may add the same comparison using standard fuzzy logic. If you can show that the classification from two different radars produce more similar results using the semi-supervised approach in comparison with fuzzy logic, this may be a very good result demonstrating the robustness of the approach. This should be discussed more extensively. Could you also do the same comparison between C-band and X-band classifications?*

- Though it is indeed a single polarization product, the fact that it considers the vertical structure of the hail cell (the most reliable indicator), makes it very useful in verifying the decision of the algorithm which considers only one sampling volume (bin-based algorithm).
- Fig. 12 indeed serves rather as an illustration than a validation and its purpose is therefore adequately rephrased in the revised version of the article. Please refer to #2.m12.
- Fig. 9 is remade and now contains a more elaborated comparisons (please refer to #2.m17). Similar comparison between C-band and X-band classifications is indeed a very good idea. Unfortunately, we did not yet expand the classification to the nearby C-band radar (La Dole), and therefore are still not able to perform the comparison. We will try to provide this comparison in the forthcoming article, which aims at comparing general performances of MXPol, DX50 and La Dole C-band radar, if the configuration of two X band radars with respect to the C-band radar allows us to do so.

M3 *Finally, the use of the 0C level only through deltaH (eq. 4) appears to me somehow "dangerous" for general applications. In fact, this implies the assumption of a standard temperature profile. It may work well for selected "standard" cases, but I anticipate problems with situations where the real profile is more complex, showing inversions or even multiple freezing levels. In these cases, consideration of the whole temperature profile would be more adequate. More specific comments are listed below.*

Indeed, that is exactly the reason why the impact of this parameter is decreased by both the introduced sigmoid transform and the lower weight.

Specific comments/Technical corrections:

S1 *P.3, L.28: "role of adhesive". I'm not sure this is proper English wording for this context.*

Acknowledged and rephrased.

Page 4: These two methods serve as a sort of link between the polarimetric radar measurements and the hydrometeor scattering hypotheses.

S2 *P.5, L.23: "The reason behind the smaller initial set for Monte Lema radar..". Do you mean Plaine Morte here instead of Monte Lema?*

Exactly, lapsus calami.

S3 *Data preparation: the selection criteria, i.e. elevation above 3.5deg and range below 40km, poses the question of how the method could be actually useful for operational implementation, where the larger amount of observations usually comes from lower elevations and ranges up to 150-200km.*

Lower elevations have been avoided due to the increased risk of contamination by residual clutter. Given that representative observations come from all along the year (different altitudes of 0°C isotherm), the observations at slightly higher elevations ($3.5°, 4.5°$) are equally representative for the lowest elevations.

Please refer to #2.M1 in what concerns the range issue.

S4 *P.6, L.15-18: it is not clear to me why you need to arbitrarily define two different transformations, one for the centroids derivations and another one for the assignments.*

The stated rationale is hopefully made more comprehensible.

Page 7: **The former one is applied in centroid derivation, while the latter is used in the assignment. The rationale behind the less steep slope applied on the representative observations is preserving a sort of continuity, for the purpose of the coherent statistical testing of five continuous distributions. Namely, the radar polarimetric variables are continuous, both in the centroid derivation and the assignment part. Therefore, given that we combine the scores of five-dimensional Kolmogorov-Smirnov (KS) tests (as it will be elaborated in the following subsection), we decided to soften the "blue" transformation, in order to make this parameter also continuous in the part where the KS test is used. This softer transformation still however limit the influence of ΔH on the classification.**

S5 *P.7, L.1: please provide more detail about the "standard operational procedure" for noise correction in the correlation coefficient. Also, specify whether the same general processing of the radar moments (including Kdp estimation) is applied to all the radars considered in this study or there are differences between the Rad4Alp network and the X-band systems.*

The revised version contains an appendix which contains concise information concerning data processing.

Pages 18-19: **Attenuation and differential attenuation for all datasets were corrected in the entire volume using ZPHI method from Testud et al. (2000), while noise in correlation was corrected according to the standard operational procedure:**

$$\rho_{hv}^{corr} = \rho_{hv}\sqrt{\left(1+\frac{1}{SNR^{lin}}\right)\left(1+\frac{Z_{DR}^{lin}}{\alpha SNR^{lin}}\right)}, \quad (11)$$

with α being the noise ratio between two channels.

At C band, instead of using the specific differential phase shift derived as a product of operational Rad4Alp radar network, in order to minimize the number of outliers, this parameter was estimated in the presented study by rigorously employing a multi-step approach (Vulpiani et al., 2012), reinforced by median filtering. In case of the MXPol radar, the method proposed in Schneebeli et al. (2014) is employed, while the specific differential phase for the DX50 radar has been derived using a routine provided by the manufacturer, based on the FIR filtering of differential propagation phase (Hubbert and Bringi, 1995) and a linear regression in deriving K_{dp}.

S6 *P.7, L.19: you should explain better the inverse sampling method. This is a fundamental part of the method, for the comparison of the two sample distributions, but it is actually not clear how the reference observations are determined in practice.*

Inverse sampling method generates the reference values by using the inverse version of the normalized cumulative sum of the membership function. We supply it with a sequence of random uniformly distributed numbers (between 0 and 1) and obtain at the output the set of observations whose distribution is determined by the parameters of the membership function (Fig. 5).

As soon as we have a simulator whose inputs are representative enough, the reference values are going to be taken directly from the non-parametric output of the simulator, which will emphasize further more the importance of using Kolmogorov-Smirnov test.

Page 8: **Inverse sampling method generates the reference values by using the inverse version of the normalized cumulative sum of the reference parametric distribution. We supply it with a sequence of random uniformly distributed numbers (between 0 and 1)**

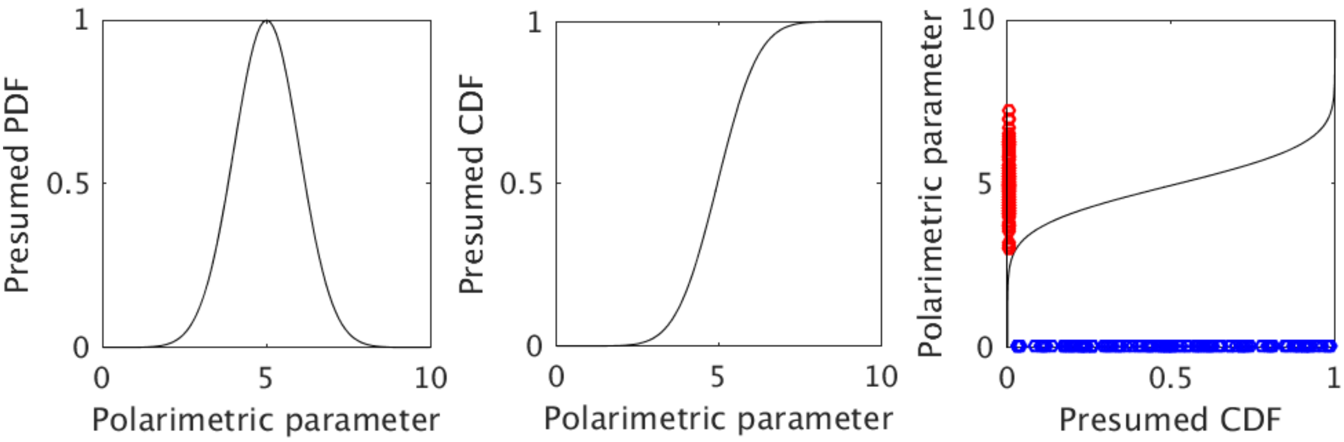

Figure 5: Illustration of inverse sampling method: Gaussian membership function.

and obtain at the output the set of observations whose distribution is determined by the parameters of the membership function.

S7 *P.7, L.30-P.8, L.7: this part (cluster splitting) is one of the original contributions of this work, but it is also not clearly illustrated. I suggest improving this description with additional figure/practical example. In addition, it may be useful to provide some statistics on the percentage of observations which needs to go through the internal loop with cluster splitting.*

Cluster splitting is basically the unconstrained clustering of a non-identified cluster on two smaller clusters. We illustrate it here (Fig. 6), but prefer not to include the figure in the article, due to the obvious space limitations.

After the initial step of unconstrained clustering on nine classes, basically all "big" clusters go into the loop which contains cluster splitting. These "big" clusters are actually too big to be able to pass the pretty rigorous identification test (correspondence with the reference values).

S8 *P.8, L. 10: correct "focussed" with "focused"*

Acknowledged and done.

S9 *P.9, L. 10-11: where is this shown?*

In can be seen in Figure 6.

S10 *P.9, L.19-26: Pixels assignment: what is the purpose of using two different normalization procedures, one for the centroids (eq. 2), using the standard deviation, and another one for the pixel assignments, using arbitrarily defined ranges for each variable?*

Please refer to #4.M1.

S11 *P.10, L. 11-14: please provide evidence of this comparative analysis. In particular I'm wondering about the impact of the correlations inherent in the radar observations (e.g. Zdr vs. Zh or Kdp vs. Zh in rain,. . .). The Mahalanobis distance provides a way to account for the correlations in the dataset and therefore should give a better distance measure. How have you evaluated this and eventually preferred the simple normalization using arbitrary ranges?*

Indeed, Mahalanobis distance should be a more plausible criterion for pixel assignment. However, as you can see in the following Fig. 7, due to a reason we still do not quite understand, it neglects too much the information provided by the phase indicator (the fifth parameter), which significantly deteriorates the output of the classification.

Besides it very SIGNIFICANTLY deteriorates the computational efficiency, one of the most important benefits of the proposed method. Obtaining a result in Fig. 7b take **93** times more time than obtaining one in Fig. 7b.

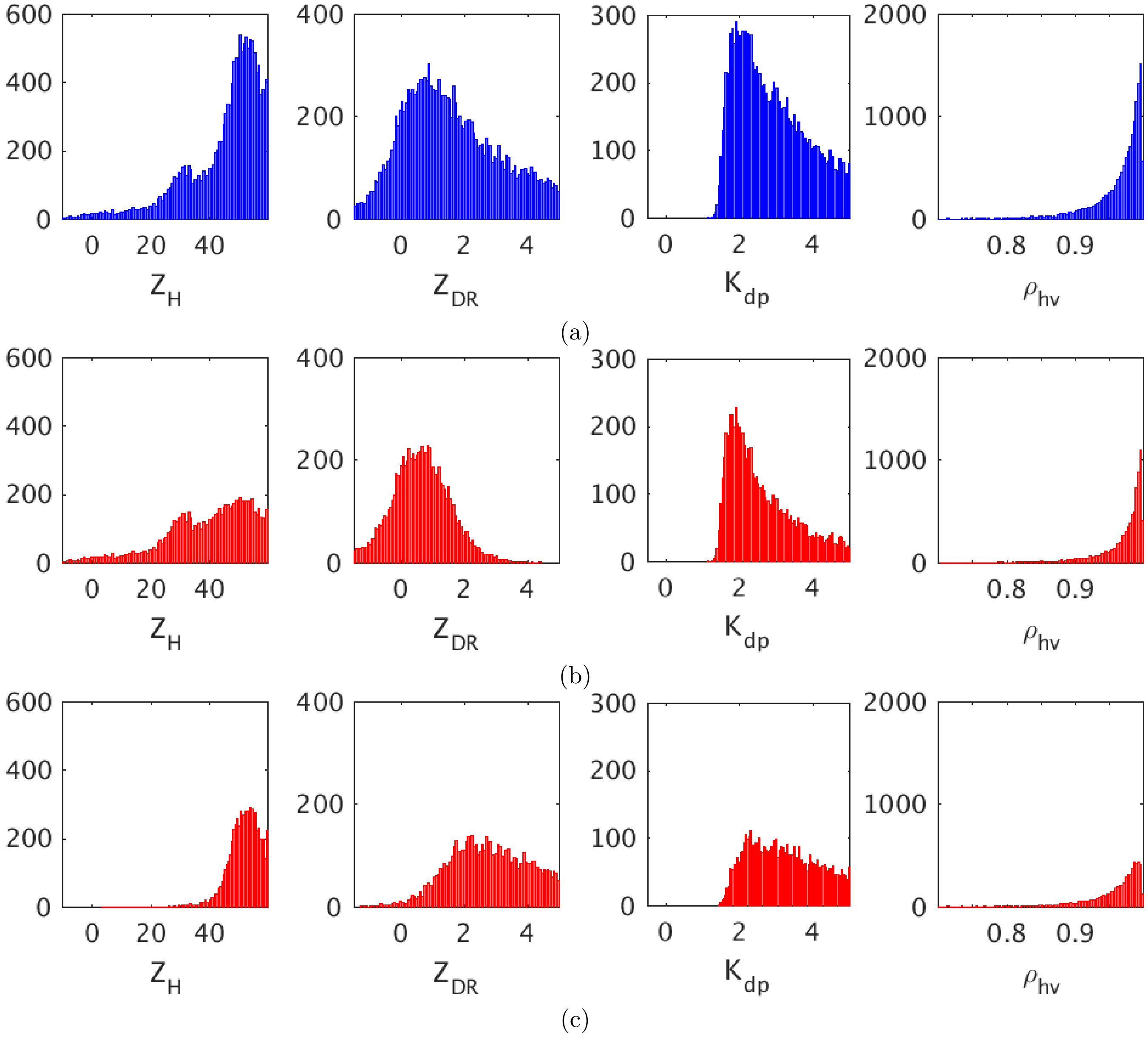

Figure 6: Illustration of cluster splitting: (a) “big” cluster, (b,c) “baby” clusters.

Page 11: The choice of the employed distance was investigated by comparative analysis with the standardized Euclidean distance (including standard deviation) and the Mahalanobis distance (including covariance estimate, Mahalanobis 1936). By taking into the account the simplicity and the computational efficiency required for the operational purpose, we have adopted the simplest option.

S12 *P.11, L.4: “they are derived using computationally expensive, fairly sophisticated clustering method”. The use of a computationally expensive procedure is not per se a guarantee of better quality. I suggest to drop this statement. The following (human expertise, complementary data) is enough to justify why you take this as reference.*

Acknowledged, the qualification “computationally expensive” is removed in th revised version.

Page 12: Namely, the centroids derived from the unsupervised method could in a way be taken as reference due to the following reasons: they are derived using a fairly sophisticated clustering method (AHC); the information about the texture is explicitly introduced; the identification is performed through human expertise, using complemen-

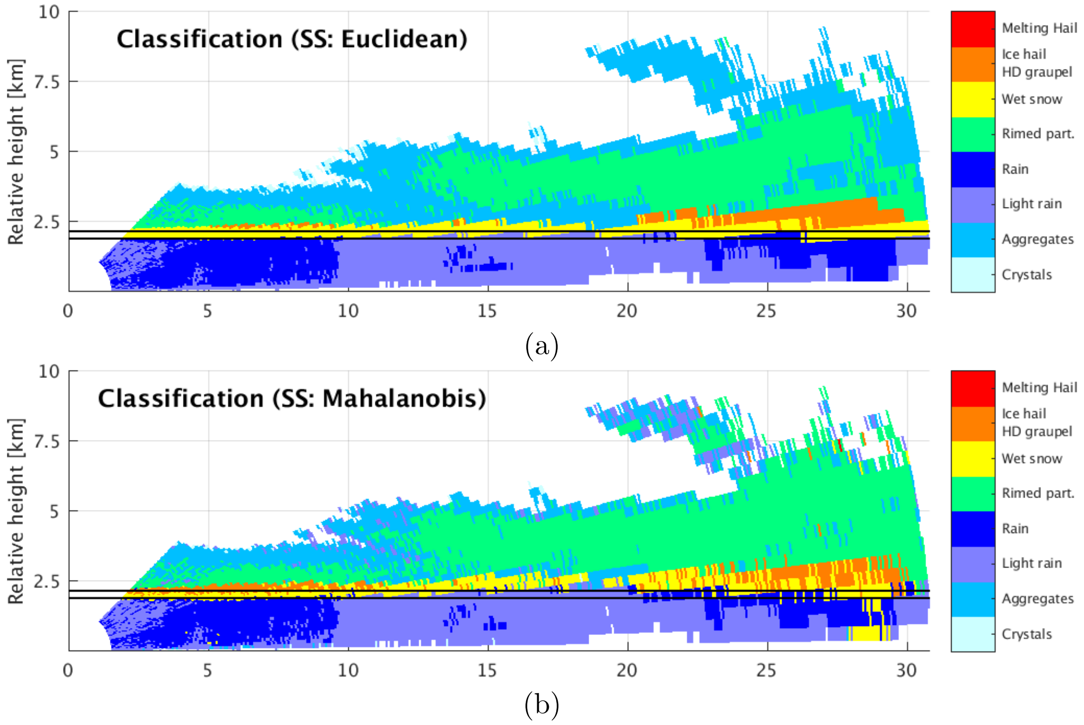

Figure 7: Classification presented in Fig. 9c in the article: (a) original example (Euclidean distance), (b) alternative example (Mahalanobis distance).

tary data when possible.

S13 *Fig. 8: you should add the RHI plots for all the dual-pol radar moments. Otherwise the interpretation of the classification and the evaluation of differences between the different techniques is not possible.*

Acknowledged and done.

S14 *Appendix A: "As well, a number of other parameters from the original membership functions has been altered to fit the specific purpose these clustering constraints have in the framework of our approach." This is quite obscure; please explain more in detail what has been altered, with proper justification.*

Please refer to #3.M5.

References

Baldauf, M., Seifert, A., Forstner, J., Majewski, D., Raschendorfer, M., and Reinhardt, T.: Operational Convective-Scale Numerical Weather Prediction with the COSMO Model: Description and Sensitivities, Mon. Wea. Rev., 139, 3887–3905, 2011.

Bechini, R. and Chandrasekar, V.: A Semisupervised Robust Hydrometeor Classification Method for Dual-Polarization Radar Applications, J. Atmos. Oceanic Technol., 32, 22–47, 2015.

Bringi, V. N., Thurai, R., and Hannesen, R.: Dual-Polarization Weather Radar Handbook, AMS-Gematronik GmbH, 2007.

Cheng, L., English, M., and Wong, R.: Hailstone size distributions and their relationship to storm thermodynamics, J. Climate Appl. Meteor., 24, 1059–1067, 1985.

Dolan, B. and Rutledge, S. A.: A Theory-Based Hydrometeor Identification Algorithm for X-Band Polarimetric Radars, J. Atmos. Oceanic Technol., 26, 2071–2088, 2009.

Dolan, B., Rutledge, S. A., Lim, S., Chandrasekar, V., and Thurai, M.: A Robust C-Band Hydrometeor Identification Algorithm and Application to a Long-Term Polarimetric Radar Dataset, J. Atmos. Oceanic Technol., 52, 2162–2186, 2013.

Gabella, M., Boscacci, M., Sartori, M., and Germann, U.: Calibration accuracy of the dual-polarization receivers of the C-band Swiss weather radar network, Atmosphere, 2016.

Germann, U., Boscacci, M., Gabella, M., and Sartori, M.: Peak performance: Radar design for prediction in the Swiss Alps, Meteorological Technology International, pp. 42–45, 2015.

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

Grazioli, J., Tuia, D., and Berne, A.: Hydrometeor classification from polarimetric radar measurements: a clustering approach, Atmos. Meas. Tech., 8, 149–170, 2015.

Haralick, R. M., Shanmugam, K., and Dinstein, I.: Textural Features for Image Classification, IEEE Transactions on Systems, Man, and Cybernetics, 6, 610–621, 1973.

Heymsfield, A. J., Schmitt, C., and Bansemer, A.: Ice Cloud Particle Size Distributions and Pressure-Dependent Terminal Velocities from In Situ Observations at Temperatures from 0° to -86°C, Journal of the Atmospheric Sciences, 70, 4123–4154, 2013.

Hubbert, J. and Bringi, V. N.: An Iterative Filtering Technique for the Analysis of Copolar Differential Phase and Dual-Frequency Radar Measurements, J. Atmos. Ocean Tech., 12, 643–648, 1995.

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

Lim, S., Chandrasekar, V., and Bringi, V.: Hydrometeor classification system using dual-polarization radar measurements: model improvements and in situ verification, IEEE T. Geosci. Remote Sens., 43, 792–801, 2005.

Liu, H. and Chandrasekar, V.: Classification of hydrometeors based on polarimetric radar measurements: development of fuzzy logic and neurofuzzy systems, and in situ verification, J. Atmos. Oceanic Technol., 17, 140–164, 2000.

Mahalanobis, P. C.: On the generalised distance in statistics, Proceedings of the National Institute of Sciences of India, 2, 49–55, 1936.

Marzano, F., Scaranari, D., and Vulpiani, G.: Supervised Fuzzy-Logic Classification of Hydrometeors Using C-Band Weather Radars, IEEE T. Geosci. Remote Sens., 45, 3784–3799, 2007.

Mishchenko, M., Travis, L., and Mackowski, D.: T-matrix computations of light scattering by nonspherical particles: a review, J. Quant. Spectrosc. Radiat. Transfer, 55, 535–575, 1996.

Nisi, L., Martius, O., Hering, A., Kunz, M., and Germann, U.: Spatial and temporal distribution of hailstorms in the Alpine region: a long-term, high resolution, radar-based analysis, Q. J. R. Meteor. Soc., doi:10.1002/qj.2771, 2016.

Pearson, E. and Hartley, H.: Biometrika tables for statisticians vol. 2, Cambridge University Press, Cambridge, UK, 1972.

Rényi, A.: On measures of information and entropy, in: Proceedings of the fourth Berkeley Symposium on Mathematics, Statistics and Probability, pp. 547–561, 1960.

Roberto, N., Adirosi, E., Baldini, L., Casella, D., Dietrich, S., Gatlin, P., Panegrossi, G., Petracca, M., Sanò, P., , and Tokay, A.: Multi-sensor analysis of convective activity in central Italy during the HyMeX SOP 1.1, Atmos. Meas. Tech., 9, 535–552, 2016.

Ryzhkov, A., Pinsky, M., Pokrovsky, A., and Khain, A.: Polarimetric Radar Observations Operator for a Cloud Model with Spectral Microphysics, J. Appl. Meteor. Climatol., 50, 873–894, 2011.

Ryzhkov, A. V., Kumjian, M. R., Ganson, S. M., and Khain, A. P.: Polarimetric Radar Characteristics of Melting Hail. Part I: Theoretical Simulations Using Spectral Microphysical Modeling, J. Appl. Meteor. Climatol., 52, 2849–2870, 2013.

Schneebeli, M., Grazioli, J., and Berne, A.: Improved Estimation of the Specific Differential Phase Shift Using a Compilation of Kalman Filter Ensembles, IEEE T. Geosci. Remote Sens., 52, 5137–5149, 2014.

Straka, J. M., Zrnic, D. S., and Ryzhkov, A. V.: Bulk Hydrometeor Classification and Quantification Using Polarimetric Radar Data: Synthesis of Relations, J. Appl. Meteor., 39, 1341–1372, 2000.

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

Vulpiani, G., Montopoli, M., Delli Passeri, L., Gioia, A. G., Giordano, P., and Marzano, F. S.: On the Use of Dual-Polarized C-Band Radar for Operational Rainfall Retrieval in Mountainous Areas, J. Appl. Meteor. Climatol., 51, 405–425, 2012.

Zrnic, D. S., Ryzhkov, A., Straka, J., Liu, Y., and Vivekanandan, J.: Testing a Procedure for Automatic Classification of Hydrometeor Types, J. Atmos. Oceanic Technol., 18, 892–913, 2001.