Review of the paper

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

by N. Besic et al.

No.: amt-2016-105

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.

- At C-band the sampling of measurements are restricted in range between 3 and 40 km (at X-band 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?
- 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 co-occurrence 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 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…
- 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 post-processing that include semi-supervised classification ?
- 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?
- 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).
- 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.

Minor Points

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.

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?

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.

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.

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?

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.

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. Pleas clarify this point.

Pag 11 line 22 How is calculated the normalize matching matrix?

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

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° is not included, using the scheme proposed in the text is not possible to correctly identify the convective core.

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.

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.

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.

Table 2 What is the coefficient “C” ?

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

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?

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

Fig 10 What is the ΔHSS?

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, doi:10.5194/amt-9-535-2016, 2016.