Hydrometeor classification from 2 dimensional video disdrometer data
amt-2013-333

Jacopo Grazioli, Devis Tuia, Samuel Monhart, Marc Schneebeli, Tim Raupach and Alexis Berne

June 3, 2014

Responses to the reviewers

We would like to take this opportunity to thank the reviewers for their constructive comments about our work. We have addressed all their concerns. Detailed replies to the reviewer's comments are provided in the next sections of this document.

As a general point, the manuscript has been edited by a native speaker and we have corrected the typos and incoherences that were present in the first version. We hope that the manuscript now meets the quality standards of AMT.

In the response below, the comments of the reviewers are reported in *italic*. Quotations from the manuscript in its revised or in its original form are reported in blue.

Anonymous Referee 1

General comments

1. *This is an interesting and nice presented work on classification of hydrometeors with 2DVD, which is a disdrometer that can measure the shape of hydrometeors.*

   We thank the Referee for his/her positive comment.

2. *Have the authors compared the classification results with the results from other commercial systems like Parsivel? This would be a useful information for Parsivel users.*

   The classification provided by Parsivel systems and the classification output presented in the manuscript are not directly comparable mainly because the available hydrometeor classes are different.

   Parsivel systems provide a SYNOP-4680 precipitation code, developed according to the guidelines of the World Meteorological Organization. The codes are:

   - 51,52,53: drizzle
   - 57, 58: drizzle with rain
   - 61, 62, 63: rain
   - 67, 68: rain, drizzle with snow

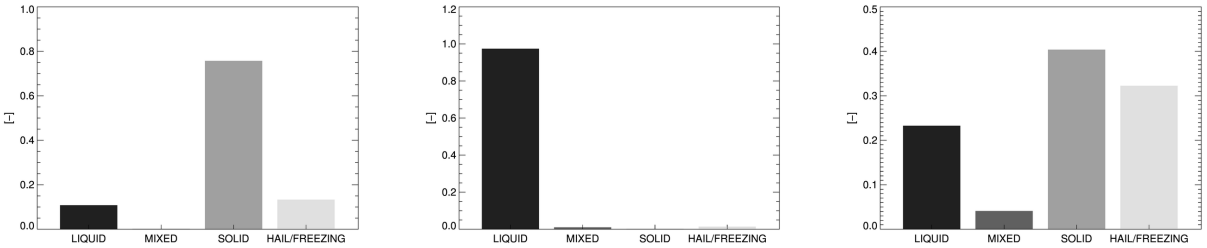
- 71, 72, 73: snow
- 77: snow grains
- 87, 88: freezing rain
- 89: hail

Even though it is difficult to perform a one-by-one comparison between these hydrometeor classes and the output of the proposed 2DVD algorithm, we can perform a global comparison to find whether the predictions of the methods are on average in agreement to each other.

To do so, we merge some of the 2DVD hydrometeor classes of the present work to obtain broader classes. AG, D, C, G and RIM together can be fused into a class called "Solid phase", representing solid precipitation; R indicates "Liquid precipitation", and MS "Mixed phase precipitation".

In an analogous way, the Parsivel classes can be grouped in "Liquid precipitation" (precipitation code ≤ 63), "Mixed phase" ($63 <$ precipitation code ≤ 68), "Ice phase" ($68 <$ precipitation code ≤ 77) and "Hail/Freezing" ($77 <$ precipitation code). With the exception of this last class, the two classification products have become comparable.

To perform the comparison suggested by Referee 1, we make use of collocated measurements taken by a 2DVD and a Parsivel (first generation) disdrometer between 2009 and 2011 in Davos (CH), in the location described in Section 2.1 of the manuscript. The results of this comparison are shown in Figure 1, where we can see that, overall, the classification of the Parsivel disdrometer agrees with the one provided by the 2DVD classification, especially in the cases of solid and liquid precipitation. In the case of "Mixed phase", the agreement is less satisfactory. Please note that, in absolute terms, the occurrence of "Mixed phase" precipitation is about one order of magnitude lower than solid or liquid precipitation, thus showing that the proposed classification system provides results that are consistent with those obtained with the Parsivel system, for these wide classes.



(2DVD prediction: SOLID) (2DVD prediction: LIQUID) (2DVD prediction: MIXED)

Figure 1: Classification of precipitation type given by a Parsivel disdrometer. The three graphs show the distribution of the precipitation types identified by a Parsivel disdrometer, when a collocated 2DVD classifies Solid, Liquid or Mixed-phase precipitation.

We agree with the Referee that this is a valuable information for Parsivel users. However, we prefer not to introduce this discussion in the manuscript, for the following reasons:

(a) We think that the manuscript should remain focussed on the proposed classification based on 2DVD data.

(b) The classification method implemented in the Parsivel disdrometers is not clearly described, either in the user manuals of the instrument or, to our knowledge, in available literature (e.g. Löffler-Mang and Joss, 2000)

3. *The classification method is quite complex and not applicable to hydrometeor mixture or separate particles in each step. In addition, quantitative information on each particle is needed to calculate useful products e.g. radar reflectivity. Future work in this direction is going to be based on the method proposed in this work or it is not practical computationally?*

   We believe that the information on the hydrometeor type is definitely relevant for the purpose of scattering simulations, but this work should be considered only as a starting point towards that objective.

   To achieve this goal, future work should at first extend the present classification method to single particles. Then, different techniques should be employed (or developed) to get an estimate of the hydrometeor density ρ [gm^{-3}], as well as an optimal (for the purpose of scattering simulations) description of the global geometry of the hydrometeors.

Specific Comments

1. *page 13, line 1-5: A short description of the basic idea of the "one-against-one" rule would be helpful, even though the reader has a hint from its name.*

The following sentences have been added at the end of Section 3.2.1, in order to provide a brief explanation of the "one-against-one" rule:

One-against-one builds as many binary classifiers as there are pairs of classes. Each classifier is therefore used to assign the time step to one of two possible classes. The time step is eventually classified into the class that received the most assignments.

Anonymous Referee 2

General comments

1. *This study is unique in that it is the first attempt to quantitatively perform an automated hydrometeor classification of 2DVD measurements. Although their method does not classify individual hydrometeors, it does provide information that can be useful to several areas of research (e.g., validation of dualpolarimetric radar hydrometeor identification; numerical weather prediction, including cloud resolving models) and perhaps also has operational use as well. So this study is most definitely a worthwhile contribution and suitable for publication in AMT. The authors provide a thorough discussion of their scientific method, and there are only a few minor concerns that need to be addressed prior to publication.*

   We thank the Referee for his/her positive comment.

Specific questions/issues

1. *There are numerous grammatical errors that must be addressed prior to publication. Although the grammatical mistakes are not major, they do somewhat detract the reader, and thus it is suggested that the authors enlist a person proficient in English grammar to help correct those mistakes. Many of these have been pointed out in the technical corrections section at the end of this review, but a more thorough review of the grammar would be better.*

   The authors agree with the Referee that the grammatical quality of the manuscript needed improvements, and we are thankful to him/her for all the help provided in order to correct many mistakes and typos. The manuscript is now revised according to the suggestion of the Referee. Additionally, it has been read and checked by a native English speaker.

2. *There is no mention of the wind effects induced by the 2DVD itself. This can be a concern for the representativeness of small rain particles sampled by the first generation taller profile version of the 2DVD Nespor et al. (2000). The wind effect is undoubtedly present and at lower wind speeds in snowfall. Although the proposed method may correctly represent the bulk type of hydrometeor within the 2DVD measurement, it may not be representative of the bulk ice crystal habit falling to the ground. It could mislead those who use this hydrometeor classification as a validation tool. Thus the authors should mention the 2DVD induced wind effect and resultant possible under-sampling of the small and less dense type hydrometeors. In doing so, the authors need to specify which version(s) of the 2DVD is being used.*

   We would like to thank the Referee, especially for pointing to us the work of Nespor et al. (2000). In order to mention the potential issue of under-sampling smaller and less dense particles due to the instrument-induced wind, we rephrased the last paragraph of Section 2.2 as follows:

   Two additional potential sources of uncertainty (whose magnitude is currently not known in snowfall) are the image distortion that can occur when the horizontal component of the falling velocity of the particles is significant, and the local winds generated by the geometry of the instrument. To date, image distortion can be corrected only in rain, and in particular only for raindrops that possess an axis of rotational symmetry (Schönhuber et al., 2008). On the contrary, the winds induced by the instrument itself can lead to an under-estimation of particles having lower density and dimension. Further research, which is beyond the scope of this paper,

is needed to develop correction schemes for snowfall measurements in order to compensate for these two potential issues.

Additionally, the 2DVD versions employed in the present study are now mentioned in a foot note, inserted after the words "...lowest density and dimension":

This issue is more severe for the first generation of the 2DVD instrument (Nespor et al., 2000). All the data employed in the present study were collected with second and third generation 2DVDs.

3. *Some discussion about the variability of hydrometeors within the 400 time steps that were examined to construct the training data set should be included. It seems like this information would provide a measure of uncertainty in the classification method and perhaps a starting point for future efforts to classify individual hydrometeors from the 2DVD measurements.*

The Referee has raised a good question here. The training data set consists of 400 time steps, each one labelled by two operators who look at all the particles recorded. Despite the use of two independent operators, a certain degree of subjectivity persists. Following this comment by the Referee, we clarified in the manuscript that the interpretation is performed independently by two operators. This is done by rephrasing a sentence in the first paragraph of Section 3.1, as follows:

Two operators interpreted independently the images by visualizing particle shapes, velocities...

We agree that some information about the variability of hydrometeors within the 400 time steps should be given. In any case we think that it is very difficult, and maybe confusing, to discuss this in detail in the manuscript. Therefore we provided a series of additional material related to this work that can be found at the following address:

`http://lte.epfl.ch/page-109172-en.html`

There, the interested reader will find part of the codes employed in the classification algorithm and some visual examples (A - B images) of particles belonging to the time steps of the training set.

4. *There is some uncertainty in how the authors calculate equivalent spherical volumetric diameter (D_e). They should either provide an equation for D_e or clarify if they are using the Huang et al. (2010) definition of particle diameter*

D_e is indeed calculated according to Huang et al. (2010). This is clarified, by rephrasing some sentences of Section 2.3.1 as follows:

This descriptor was originally developed for raindrops, for which volumes could be calculated accurately from the 2-D views. It can be extended to particles of any shape as a reference measure of particle size. In the present work D_e is calculated according to the formulation of Huang et al. (2010).

We decided to provide the appropriate reference and not the equation in order to avoid misunderstandings between the terms of the equation provided in Huang et al. (2010), and similar notations used in the present manuscript.

Technical corrections

1. *Title... "videodisdrometer" should be two words, "video disdrometer"*

   The new title is: Hydrometeor classification from two-dimensional video disdrometer data.

2. *Abstract... line 7 ... change to "... trained over 60 second precipitation time steps and labeled by visual inspection."*

   The manuscript has been modified according to the suggestion of the Referee. The sentence now reads:

   ... trained over 60 second precipitation time steps labeled by visual inspection.

3. *Abstract... line 8... "... algorithm achieves* **nearly** *accurate classification, with..."*

   The sentence has been rephrased as:

   The algorithm achieved high classification performances, with median overall accuracies...

4. *Page 2... line 13... change "... improved the use of ground observations..." to "... improves the ability..."*

   The sentence has been rephrased as:

   ... significantly improves the capability of ground observations to describe the microphysics...

5. *Page 2... line 18... suggest replace "... compare rainfall observations to ground truth with weather radar measurements..." with "... validate weather radar rainfall estimates..."*

   The manuscript has been modified according to the suggestion of the Referee.

6. *Page 2... line 21-22... replace "... conversion of weather radar observations to..." with "radar retrieval of..."*

   The sentence has been rephrased according to the suggestion of the Referee.

7. *Page 2... line 23... replace "... in agreement with..." with "... from..."*

   The replacement has been made.

8. *Page 3... line 5... should mention that the use in validation of radar hydrometeor identiciation.*

   We respectfully disagree with the Referee on this point. In this part of the introduction we introduce the use of hydrometeor classification schemes in general terms (i.e., without referring specifically to the possible applications of the 2DVD classification algorithm presented here). Such potential application is made later in the text, in the conclusions (Section 6, middle):

   This work has the potential to be a starting point for ground-based quantitative evaluation of products derived from polarimetric weather radars.

9. *Page 3... line 11... "sensors* **enable the sampling of** *large domains..." AND "...time lapse..." should be "...time* **scale**..."

   Both changes have been implemented according to the suggestions of the Referee.

10. *Page 3... line 14... "...aircraft* **flight** *paths..."*

    The typo has been corrected.

11. *Page 3... line 15... replace "...specific measurement..." with "...intensive measurement..."*

    The change has been implemented, according to the suggestions of the Referee.

12. *Page 4... line 3... "...algorithms,* **such** *as the support vector..."*

    The change has been implemented, according to the suggestions of the Referee.

13. *Page 4... line 4... "nowadays* **are** *used to face* **similar kinds of tasks**. **For** *example,* **such techniques have been used in** *land cover..."*

    The sentences have been rephrased as:

    Supervised classification algorithms, such as the support vector machine, (SVM, Boser et al., 1992) are nowadays used to perform similar kinds of tasks. For example, such techniques have been used in land cover classification...

14. *Page 4... line 12-13... replace "...to be top ranked for weather prediction classification tasks..." with ... "to perform relatively well at the prediction of weather type..."*

    The sentence has been rephrased as:

    ...has been shown to perform relatively well on the prediction of weather types...

15. *Page 4... line 17... remove "...into..." AND line 18... "...classes* **as** *the dominate..."*

    Both changes have been implemented.

16. *Page 4... line 24... "...the* **southern part of Ontario, Canada**..."

    The change has been implemented according to the suggestion of the Referee.

17. *Page 6... line 1... replace "...in the range of..." with "...around..."*

    The replacement has been made.

18. *Page 6... lines 3-4... sentence would be better stated as follows, "...at 34 kHz, and the vertical distance between the measurement areas of both cameras A and B enables the measurement of fall velocity."*

    We agree with the suggestion of the Referee, the sentence has been rephrased.

19. *Page 6... line 9... replace "... workflows..." with "... methods..."*

    The replacement has been made.

20. *Page 6... line 10... sentence would be better as follows, "... those studies, which were interested in snowfall only, restricted the maximum fall velocity to 4 m/s and 6 m/s, respectively.*

    The sentence has been rephrased in agreement with the suggestion of the Referee.

21. *Page 6... line 13... "... variation* **found in** *rain..."*

    The typo has been corrected.

22. *Page 6... line 16... the authors should mention what they believe the large UFOs (unidentified falling objects) present in some of their 2DVD measurements might be. They are likely either icicles or human limbs entering the measurement area during calibration.*

    The Referee asked a very interesting question. We added the following sentence to mention our interpretation of the large UFOs (Section 2.2, towards the end):

    The exact nature of these artifacts is unknown, but their vertical orientation and dimension suggest that they may be associated with small-scale wind effects, melting, or dripping, causing some particles to reside for an anomalous amount of time in the measurement areas of the two cameras.

    In fact, we do not believe that the UFOs are actually icicles or human limbs for the following reasons: (i) they also appear when the surrounding temperatures are steadily positive, (ii) they appear even when nobody was on-site, for calibration or maintenance of the system. We rather believe that the large vertical dimension of the UFOs has to be associated with long crossing times, given the measurement principle of the 2DVD.

23. *Page 6... line 21-22... the statement that the distortion due to the horizontal motion can be corrected for rain is too assertive. This effect can only be corrected for raindrops that possess an axis of rotational symmetry (e.g., those undergoing axisymmetric oscillation at the time of measurement; Schönhuber et al. (2008);* **?**.

    We thank the Referee for the valuable references that he/she is providing about this topic. We rephrased the sentence as:

    To date, image distortion can be corrected only in rain, and in particular only for raindrops that possess an axis of rotational symmetry (Schönhuber et al., 2008).

24. *Page 6... line 25... replace "Couples..." with "Pairs..."*

    The replacement has been made.

25. *Page 6... section 2.3.1... D_e needs to be better defined (see fourth specific questions/issue above).*

    Please see the answer to comment # 4 of this response.

26. *Page 6... section 2.3.2... It would be useful to the savvy reader if the authors indicated which level of 2DVD data (i.e., AB or HYD) they are using to calculate the particle characteristics, especially since only some of the data is readily available for most data users.*

    The descriptors are calculated using SNO data. To add this information in the manuscript we added a footnote in Section 2.3, after the words "relevant descriptors". The text of the footnote is the following:

    The particle descriptors are calculated in the present work from ".SNO" 2DVD data.

27. *Page 8... line 21... "... 16* **derived** *from..."*

    The replacement has been made.

28. *Page 10... line 14... "... riming processes* **smooth** *the shape..."*

    The typo has been corrected.

29. *Page 17... line 1... "... in total as* **a** *training set..."*

    The modification has been made.

30. *Page 17... line 18... "... now on 200 realization***s** *of..."*

    The typo has been corrected.

31. *Page 18... line 18... "... outperforms LDA by more than* **20%** *and NN by more than* **10%** *in terms of..." (i.e., assuming SVM performs 20% and 10% better)*

    With this sentence we meant that the overall $\mathcal{K}$ was 0.2 [-], and 0.1 [-] higher in absolute units, and not in percentage. To give a relative appreciation, we rephrased the sentence as:

    Comparison with these 2 methods showed that the proposed SVM scheme outperforms LDA by more than 25% and NN by more than 15% in terms of $\mathcal{K}$.

32. *Page 18... line 22... "... 60s), account***ed** *for..."*

    The replacement has been made.

33. *Page 19... line 14... replace "In a second time..." with "During the next relatively stable phase..."*

    The sentences have been rephrased according to the suggestion of the Referee.

34. *Page 19... lines 17-20... "***The median** *PF is around 0.7 during the* **entire event**..." *"***The median** *D_e is initially below 1 mm (SP phase)...* **in the latter part of the event characterized mostly by G and RIM classes.***"*

    The sentences have been rephrased according to the suggestion of the Referee.

35. *Page 19... line 21... graupel is misspelled*

    The typo has been corrected (grauple changed to graupel).

36. *Page 19... final 2 lines... remove "in median value." Place the median descriptor prior to the mention of D_e... "... with median D_e ranging... "*

    In the first part of the sentence we were referring to the full range of variation of D_e, while in the second part we were referring only to median D_e. We understand that this might be misleading. The main message of the sentence was about the wider range of variation of D_e, and therefore we rephrased as follows:

    ... we observe a wider range of variation of particle sizes, with D_e ranging between 0.5 mm and 8 mm (AG).

37. *Page 20... line 5... "... with* **median** *PF below 0.7 through***out** *the event,* **and** *slightly... "*

    The sentence has been rephrased following the suggestions of the Referee.

38. *Page 20... line 6-7... "... and dendrites* **relative to** *small particles... "*

    The modification has been made.

39. *Page 20... line 16-17... "***The** *rain is characterized by* **small** D_e **and** $2 \leq v \leq 5$ *m/s (***i.e.***, light rain),* **which is** *larger than the... and very high compactness with* **a median** *PF around 0.9."*

    The sentence has been rephrased in agreement with the suggestion of the Referee.

40. *Page 20... line 20-21... "... around 1 m* s^{-1}*, the* **spread** *of* D_e *increases ... median PF drops* **to** *0.6 in... "*

    The sentence has been rephrased in agreement with the suggestion of the Referee.

41. *Page 20... line 23... "***Generally***, the transition... "*

    The modification has been made.

42. *Page 20... lines 24... "Figure 12 shows the* **relative number of classifications** *for each of these three types... ".*

    The modification has been made.

43. *Page 20... line 25... The classification type as a function of temperature is somewhat confusing. Was temperature used as an input for the classification? If not then perhaps state that it was not used by the classification algorithm (even if stated earlier)... that is an interesting result. This also raises the question of freezing rain and how would that be classified by the algorithm. Perhaps it is worth mentioning something about freezing rain in the summary/conclusions.*

    Temperature was not used as an input to the system. All the necessary inputs of the method are listed in Section 2.3, where we describe the particle descriptors, and their statistical description over time intervals of 60 seconds. In order to clarify that the algorithm does not employ temperature as input, we rephrased a sentence at the end of Section 5.3 as follows:

    Figure 12 shows the relative number of classifications for each of these three types of hydrometeors as a function of the temperature. Please note that temperature has not been used as an input in the proposed system (Table 1).

Also, in the beginning of section 5.1, we rephrased the second sentence as:

The air temperature recorded in the close vicinity (less than 50 m away)...

The second point of the Referee here is about freezing rain, and how would it be classified by the algorithm. Unfortunately, no samples were classified as freezing rain in our database, or at least not with sufficient confidence to justify a specific hydrometeor class. Therefore, strictly speaking, we cannot state a-priori the behavior of the algorithm in this case.

If the Referee defines "freezing rain" as the rain that freezes on contact with the ground (i.e., `http://www.weather.com/encyclopedia/winter/precip.html`), we are convinced that it would be classified as rain because its characteristics are those of rain, until the drops reach a surface (but it is difficult to imagine how the instrument itself would be affected by such conditions).

As a last remark we would like to stress the fact that if samples of a new hydrometeor class (i.e., hail, plates...) become available, our approach is flexible enough to incorporate them in the training set and add the class to the classification output.

44. *Page 21... line 8... "... is provided,* **which is** *of particular importance..."*

The modification has been made.

45. *Page 21... lines 14-15... "***Each of** *the classes are identified with* **a median overall** *accuracy exceeding 84%."*

The sentence has been rephrased in agreement with the suggestion of the Referee.

46. *Page 21... lines 20-21... "... classification of hydrometeors* **from the 2DVD measurements** *provides additional information that can also..."*

The sentence has been rephrased in agreement with the suggestion of the Referee.

47. *Page 21... lines 22-23... " This work also* **has** *the potential...".*

The modification has been made (and the word "also" has been removed).

48. *Page 21... line 24... "... products* **derived** *from polarimetric weather radars.* **I***t could* **also** *be adapted and implemented in* **other** *particle imaging systems,* **both** *ground-based..."*

The modifications and the corrections suggested by the Referee have been performed.

49. *Can it be also applied to one dimensional particle imagers (e.g., Snow Video Imager, optical spectrometer, etc.)? If so, then it should be mentioned too.*

We are confident that this approach, with the appropriate modifications, can be applied to all the particle imagers that allow computation of geometrical descriptors (not necessarily the same ones used in this manuscript) and that allow human interpretation, and therefore training. Therefore, we think that the method can be applied at least to the SVI.

We modified the conclusion section as follows:

It (*the method*) can also be adapted and implemented to receive inputs from other particle imaging systems (one or two dimensional), both ground-based or airborne, provided that human interpretation can be carried out for the particle in the training set and that geometrical descriptors can be computed from the particle images.

Anonymous Referee 3

General comments

Article provides new unique methods for identifying hydrometeor types from disdrometer data. Has potential to be useful for many applications.

1. *While the article is well thought out, the grammatical errors make reading difficult. There are too many to list, so I would recommend having someone who specializes in English grammar help rework sentence structures. Also, carefully check spelling. There were a few major spelling issues, for example: hydrometeor in the title is spelled incorrectly.*

   We renew our apologies for the poor quality of the English grammar, and for the numerous typos. The manuscript is now revised according to the suggestion of the referees. Additionally, it has been read and checked by a native English speaker.

Overall I would recommend this article for acceptance with minor revisions, a couple of specific comments/questions are listed below.

Specific Comments

1. *Title: Hydrometeor is incorrectly spelled and videodisdrometer is two words.*

   The title has been changed according to the suggestion of the Referee, and now it reads:

   Hydrometeor classification from two-dimensional video disdrometer data

2. *Section 2: At some point in this section it would be useful to readers if you included some quantitative percent of rejected/accepted particles. This will give an idea of how inclusive your methods are.*

   In Section 2.2, second half, we added the following sentence to address this relevant comment of the Referee:

   The proportion of rejected particles is on average 3%, and it ranges between 0.5% and 13% per day. A few precipitation events required higher rejection rates. They were excluded from the analysis presented in this work.

3. *Section 2.3.3: At the end of this section you mention statistical calculations of mean, median, and percentiles. From a statistical point of view, can you comment on whether your distributions were normal and what you hope to gain from these statistical calculations?*

   Let us begin with the second part of the question of the Referee, i.e., "what you hope to gain from these statistical calculations? " In the end of Section 2.3.3, we define the features that will be the input of the SVM classifier. These features are indeed some statistical indicators, calculated for each particle descriptor. The goal of these statistics is mainly to summarize over time intervals of 60 seconds the large amount of information that it is available when many particles are recorded by the instrument (each particle is associated with the full set of descriptors). This goal is mentioned at the beginning of Section 2.3, with the following sentence:

Pairs of 2DVD A-B images are available for each particle falling in the measurement area. For the purpose of the present work, it is useful to summarize this large amount of information by choosing a set of relevant descriptors. Then, the statistical distributions of these descriptors in a time step Δt are used as input information for the hydrometeor classification.

Let us now focus on the first part of the question of the Referee: " can you comment on whether your distributions were normal. . . "

The descriptors used to characterize single particles (i.e., the ones listed in Table 1 of the manuscript), are generally not normally distributed during the 60 s Δt. This is the reason why we decided to describe their distribution in Δt by means of (Section 2.3, middle):

. . . mean, median, some quantiles (10 %, 25 %, 75 %, 90 %) and interquantiles (Q75–25, Q90–10) of each descriptor over the N particles available . . .

instead us simply using mean and standard deviation.

To summarize, two aspects must be kept in mind:

- The descriptors of Table 1 of the manuscript do not need to be normally distributed over each time Δt, as they are summarized by a set of statistical moments that do not assume normality.
- Furthermore, these statistical moments (which are in turn the inputs of the SVM) do not need to be normally distributed over a population of time intervals Δt, because the classifier is non-parametric and does not require it to function correctly.

Anonymous Referee 4

The paper by Grazioli et al. presents several issues that make me difficult to recommend publication.

1. *The English should be greatly improved before submitting the paper to a reputable journal such as AMT. Starting from the first word of the paper (it reads Hydrometor instead of Hydrometeor), there are a large number of typos and incorrect and careless sentences that do not build confidence on the work.*

   We apologize for the many typos and grammatical mistakes of the AMTD manuscript. The manuscript is now revised according to the many constructive suggestions received, and it has been proofread by a native English speaker.

2. *There are several unsubstantiated statements in the paper, thus for instance the comment "Some commercial disdrometers (i.e., PARSIVEL), orginally [sic, typo] designed for rainfall studies, provide an estimation of the precipitation type associated with each measurent [sic, typo] by making assumptions on fall velocity and and [sic,typo] equivalent rainfall intensity. For this reason, they are prone to unreliable estimates in complex site conditions."*

   About the grammatical mistakes, the sentence now is corrected as:

   Some commercial disdrometers (e.g. PARSIVEL), originally designed for rainfall studies, provide a basic estimation of the precipitation type associated with each measurement by making assumptions on fall velocity and equivalent rainfall intensity.

   One issue is the precipitation type classification and a different one is the spatial variability of the estimates. The sentence mixes both problems and does not provides any support for any of the two statements. Besides, there are 3 typos in just one sentence.

   We believe that the the Referee here was considering as non-acceptable our sentence "For this reason, they are prone to unreliable estimates in complex site conditions". This sentence has been removed from the revised manuscript.

   Our goal in this section is to acknowledge the fact that some commercial disdrometers do already provide an estimate of the precipitation type. We would like to stress that the classification provided by PARSIVEL, to our knowledge, (1) is not well described/documented in any available piece of literature, (2) has never been evaluated extensively, (3) is more focused on the simple separation between liquid and ice phase precipitation. For these reasons we were defining it as "prone to unreliable estimates", because it is based only on fall velocity and equivalent rainfall intensity.

   About the *spatial variability of the estimates*, we would like to underline that in the sentence quoted by the Referee, and more generally in the whole manuscript, we actually do not talk about spatial variability, since in the manuscript we consider only point measurements.

   Moreover, the same sort of PARSIVEL pre-processing and filtering is also required for 2DVDs (p6, sentence starting "The raw images need to be processed before being employed.")

   We do not understand this comment by the Referee. Pre-processing is a normal step in the treatment of the data coming from any measurement sensor, and in this specific case it has nothing to do with the hydrometeor classification itself. Additionally, PARSIVEL is a laser disdrometer (e.g., Löffler-Mang and Joss, 2000; Jaffrain et al., 2009), while the 2DVD is a video disdrometer (Kruger and Krajewski, 2002). The measurement principle of these two instruments is extremely different, as is their data pre-processing.

3. *Disdrometers are usually deployed in pairs to check for consistency. The paper does not provide enough information on the actual setup used in the three campaigns to evaluate the consistency of the estimates. A comparison of the results of two backto- back 2DVDs would make the case for the method, but if that is not the actual setup then the whole exercise falls apart.*

   We respectfully disagree with this comment, for the following reasons:

   - The goal of the manuscript is to describe how to use the 2DVD for hydrometeor classification purposes, by proposing an original approach. The evaluation of the accuracy of the instrument (derived by " *A comparison of the results of two backto-back 2DVDs* ") is beyond the scope of the present manuscript, and it should be the object of future studies.
   - The 2DVD is an expensive instrument, one order of magnitude more expensive than (for example) PARSIVEL disdrometers.
   - A sufficiently large database of observations from collocated pairs of 2DVDs, to our knowledge, does not exist.

4. *The wind effect needs to be addressed in relation to the representativeness of the estimates.*

   In order to clarify our assumptions, and the unknown uncertainties connected with the wind, the last paragraph of Section 2.2 is modified as follows:

   Two additional potential sources of uncertainty (whose magnitude is currently not known in snowfall) are the image distortion that can occur when the horizontal component of the falling velocity of the particles is significant, and the local winds generated by the geometry of the instrument. To date, image distortion can be corrected only in rain, and in particular only for raindrops that possess an axis of rotational symmetry (Schönhuber et al., 2008). On the contrary, the winds induced by the instrument itself can lead to an under-estimation of particles having lower density and dimension. Further research, which is beyond the scope of this paper, is needed to develop correction schemes for snowfall measurements in order to compensate for these two potential issues.

   Additionally we included a foot note to clarify that we do not use the first generation of the instrument. The text of the foot note, that can be found after the words "...lowest density and dimension", is:

   This issue is more severe for the first generation of the 2DVD instrument (Nespor et al., 2000). All the data employed in the present study were collected with second and third generation 2DVDs.

   We believe that in this way the reader will become aware of the potential issues related to wind.

5. *Section 3.2 is unnecessary, as it only report information already published elsewhere which is not original from the authors, and of a general character, i.e. not specific to the case. Stating that "We used the SimpleMKL algorithm (Rakotomamonjy et al., 2008)" and then provide the parameterization would suffice. The same applies to figure 4, which is well-known and unnecessary for this paper.*

   We do not agree with this observation of the Referee. The concepts detailed in Section 3.2 are definitely not "well-known and unnecessary" for the public that this manuscript is intended to reach. Our main "target" is in fact the precipitation measurement community. We believe that: (i) a reader from this community might have some difficulties without a proper introduction

of the SVM, and (ii) by providing these details a reader who is not aware of the concepts of supervised classification might find a source of inspiration to solve similar problems.

6. *More detail is required on the manual classification method.*

   To improve the description of the manual classification method, we rephrased Section 3.1 as follows:

   The 2DVD offers the possibility to visualize the actual hydrometeor images, and the supervision was therefore conducted manually, according to the judgement of trained operators. Two operators independently interpreted the images by visualizing particle shapes, velocities, and taking into account the on-site environmental conditions (time of the year, temperature). Additionally, for the data collected in Davos (CH), X-band radar observations over the region were available (e.g. Schneebeli et al., 2013), thus providing contextual information about the structure of the precipitation, and in stratiform cases, about the altitude of the melting layer.

   Then, in the fourth paragraph:

   The creation of the training set involved the inspection of all the particles within each time step, in order to retrieve the dominant particle type and to provide the appropriate label. Particular attention was paid to select time steps that were as pure as possible, for the subsequent training of the classifier. The training set employed in the present work includes $N_{\text{train}} = 400$ time steps...

7. *The temperature interval in the top of figure 9 (and 10 and 11) is too wide, and therein useless.*

   The temperature interval of Figure 9, 10, 11 were set from -20 to 10 °C, in order to cover most of the cases encountered over all the available data. We agree with the Referee that in this particular case the interval was too wide. In the revised manuscript, we narrowed it between -6 and 6°C, ensuring both better readability and comparability of the three figures.

8. *Fig. 12. "... and temperature data are given by closely-located weather stations". Given the spatial variability of the estimates, one would expect a in-site estimate of the temperature, not a 'close' one.*

   The sentence is clarified as follows:

   ... by a closely located (distance ≤ 50 m) weather station...

   About the "*spatial variability of the estimates*", we could interpret this comment as:

   (a) Spatial variability of the temperature estimates. In this case we believe that the variation of temperatures for distances between the temperature sensor and the 2DVD smaller than 50 m and averaging times of 60 s is negligible.

   (b) Spatial variability of the classification estimates. In this case we would like to stress that we never treat this aspect in the manuscript (we consider only point measurements).

9. *Concerns about computational costs are not helpful without the actual details (how long does it takes, etc.).*

   We agree with the Referee and the following sentence in the conclusions (Section 6):

   The choice of the SVM as classifier makes the method well balanced in terms of accuracy and computational cost, and adaptable to real time applications.

   has been removed from the revised manuscript. Anyway we believe that it is important to underline that the classification can be used in future for real time applications, and we added the following sentence in the conclusions:

Each of the classes were identified with a median accuracy exceeding 84 %. Additionally, once trained, the classifier is fast enough to be potentially implemented in real time.

10. *My main concern is about the validation step (section 4.3). The authors use 300 samples for the training step and 100 for validation (actually, testing in the literature)*

    We decided to use the term "validation", because it is more commonly used in the precipitation measurement scientific community (even though we agree with the reviewer that in the data mining community the term 'testing' is more common to define this specific task).

    . . . but it seems that they do not use independent data from another episodes for a true validation of the generalization abilities of the algorithm. This would be a common pitfall meaning that you are capable of successfully discriminate different instances on your current case, but not necessarily on a different one.

    We would like to provide some clarification to the Referee about this aspect. At first, the 300/100 split is repeated 200 times and randomly (as specified in the beginning of Section 4.3). Most importantly, the 400 time steps used for this purpose come from different 2DVD instruments, and different field campaigns, as mentioned in Section 2.1:

    The 2DVD data employed in the experiments were collected during three distinct field campaigns, between September 2009 and March 2013. The first campaign. . .

    Therefore the method is actually tested on data issued from a wide range of environmental conditions and not, for example, from a single instrument and a single day of measurements (i.e., a single case).

    As an additional test, we trained the algorithm only on labelled data coming from the first two field campaigns (Davos and Remoray), and tested it on labelled data coming from the third field campaign[1], obtaining an Overall Accuracy of 87.9%, and $\mathcal{K}$ of 0.84. These values are in the range of variation of the values obtained in validation (Table 2 of the manuscript).

    Also, since a 2DVD will always collect just a sample of the whole, far larger population, this is a critical point even for the same episode. As currently described in the paper, the work cannot say anything about the ability of the algorithm to generalize and that makes the method not really useful for the intended main application (ground-based quantitative evaluation of products coming from polarimetric weather radars; section 6).

    Considering what is stated above, the algorithm actually generalizes, over the 2DVD labelled data available from different field campaigns. We do not think that the small sampling area of the instrument is an issue for the classification algorithm, because it is based on point measurements and does not include any spatial (or spatial variability) information.

    We agree with the Referee that the differences in sampling volumes (and sampling times), between the 2DVD and polarimetric weather radars will have to be taken into account, and the data will need spatial and temporal averaging, as done in similar comparisons between different measurement devices and radar data (e.g., Bringi et al., 1998; Zhang et al., 2001; Anagnostou et al., 2008).

    Finally, we also explicitly mention that (Section 6, end):

    This work has the potential to be a starting point for ground-based quantitative evaluation of products derived from polarimetric weather radars.

[1] In this way all the hydrometeor classes were available in training, and 4 hydrometeor classes were represented in validation.

without claiming that this has already been done, or that this comparison will be without assumptions. Details about a comparison strategy should in any case be part of any future work.

11. *Moreover, even if 100 cases were actually used for validation, that would mean little since the collecting area of the instrument is quite small. Training with 300 samples and validating with 3000 would make some sense and would help to make the case for the proposed method (always providing that a twin instrument 2 meters apart would yield the same results).*

    We do not understand why the Referee refers to the small collection area. In the present work we do not deal with spatial data, but with point measurements taken at different times.

    We agree with the Referee that a larger validation set would improve the validation procedure. We would like to underline that:

    - Figure 6 in the manuscript suggests that the training set size is sufficient for the proposed application.
    - 3000 validation samples were not available. The manual and independent classification of time steps Δt by two operators, detailed in Section 3.1 is extremely time consuming, and the generation of 3000 properly labelled samples would require an excessive amount of time and effort. The method is flexible enough to include more observations in validation and training, as soon as they become available.

References

Anagnostou, M. N., E. N. Anagnostou, G. Vulpiani, M. Montopoli, F. S. Marzano, and J. Vivekanandan, 2008: Evaluation of X-band polarimetric-radar estimates of drop-size distributions from coincident S-band polarimetric estimates and measured raindrop spectra. *IEEE T. Geosci. Remote Sens.*, **46 (10)**, 3067–3075, doi:10.1109/TGRS.2008.2000757.

Boser, B., I. Guyon, and V. Vapnik, 1992: A training algorithm for optimal margin classifiers. *5th ACM Workshop on Computational Learning Theory*, Pittsburgh, USA, 144–152.

Bringi, V. N., V. Chandrasekar, and R. Xiao, 1998: Raindrop axis ratios and size distributions in Florida rainshafts: an assessment of multiparameter radar algorithms. *IEEE T. Geosci. Remote Sens.*, **36 (3)**, 703–715, doi:10.1109/36.673663.

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

Jaffrain, J., A. Berne, A. Studzinski, and F. Pantillon, 2009: A network of disdrometers to quantify the small-scale variability of the raindrop size distribution. *Proc. 34th AMS Conf. Radar Meteorol.*, Williamsburg, USA, AMS.

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

Löffler-Mang, M. and J. Joss, 2000: An optical disdrometer for measuring size and velocity of hydrometeors. *J. Atmos. Oceanic Technol.*, **17 (2)**, 130–139.

Nespor, V., W. F. Krajewski, and A. Kruger, 2000: Wind-induced error of raindrop distribution measurement using a two-dimensional video disdrometer. *J. Atmos. Oceanic Technol.*, **17 (11)**, 1483–1492.

Rakotomamonjy, A., F. Bach, S. Canu, and Y. Grandvalet, 2008: SimpleMKL. *J. Mach. Learn. Res.*, **9**, 2491–2521.

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

Schönhuber, M., G. Lammer, and W. L. Randeu, 2008: *The 2D video disdrometer. Chapter 1 in 'Precipitation: Advances in Measurement, Estimation and Prediction'*. S. Michaelides, Ed.

Zhang, G., J. Vivekanandan, and E. Brandes, 2001: A method for estimating rain rate and drop size distribution from polarimetric radar measurements. *IEEE T. Geosci. Remote Sens.*, **39 (4)**, 830–841.