Hydrometeor classification based on disdrometric spectral data

Author: Marc Lladó Duran
*Facultat de Física, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain.**

Advisor: Joan Bech, joan.bech@ub.edu

Abstract: Two classification methods for spectral data of a Parsivel2 disdrometer are proposed using empirical relations between fall velocities and diameters of the different types of hydrometeors. To this end, data obtained by a Parsivel2 in Das (La Cerdanya) during 2018-2020 is used to model and compare both classification methods with the one given by Parsivel2.

I. INTRODUCTION

The monitoring of meteorological data has always been of special interest, but it was from the Ancient Greece and Aristotle that data started to be collected systematically. For many centuries the activity of the meteorological observer, who only had some basic instruments such as thermometers and barometers, was essential for the science of meteorology. However, it is from the technological revolution up to our days that systems have been developed in order to monitor in an automatic and more objective way, with accuracies and amounts of information that ancient meteorological observers could not afford.

A particular variable is the precipitation type, which due to its qualitative characteristics has been difficult to classify in an objective, systematic and non-manned way. Although the Code 4677 by WMO [1] is a systematic classification code of precipitation types that was developed for the reporting of weather from manned weather stations, the invention of disdrometers has permitted to perform the classification of precipitation types automatically, and therefore to use the code 4677 to this aim.

Parsivel2 [2], which is an evolution of the original Parsivel, is a widely used disdrometer. It is a laser-based optical system for the measurement of all kinds of precipitation. To do so, it measures the fall speed v and diameter d of every single hydrometeor falling along the detection zone. Many variables can be derived from these data, such as the precipitation type using Code 4677, but also the kinetic energy, the intensity of precipitation or the radar reflectivity.

Although Parsivel2 provides a relatively simple and not too expensive technology, it is of special interest for scientific usage the understanding of the internal Parsivel2 algorithm used to determine the precipitation type according to the v-d characteristic. However, because of understandable commercial and 'know-how' reasons, Parsivel2 manufacturers prefer keeping this knowledge to themselves.

Therefore, data collected by a Parsivel2 disdrometer in Das, la Cerdanya (1097m a.s.l.) during the period 2018-2020 has been used to develop a classification method which tries to get the precipitation type from the v-d characteristics as Parsivel2 algorithm does.

II. METHODOLOGY

A. Raw data

Raw data used is contained in three netcdf4 files, one per year. Parsivel2 takes measurements of speed and diamater of every falling particle during a one-minute series. Since data information of speed and diameter is organized into 32 classes for each variable, 32×32 v-d classes are given. For each class, Parsivel2 returns the number of counts in the minute series. Other computed variables as the precipitation rate R, the kinetic energy or the precipitacion type according to Code 4677 are also given. We have therefore one row per minute, from which we can retrieve the matrix of class counts, the precipitation type and the equivalent rain rate, which will be the three variables that will be used. It is worth noting that in the raw data we already have the precipitation type of every minute series computed by Parsivel2 according to its internal algorithm.

From now on we will refer to the precipitation code given by Parsivel2 as *PSV2*, while we will refer to the equivalent rain rate as *R*. We discard one-minute series containing non-valid values (NaN), to select only those for which R is positive then, i.e. the minutes in which there is some kind of precipitation, which implies less than 4% of the total records available (see Table I).

TABLE I: Statistics of raw data showing the number of minutes and percentage over total data of all data, valid (not NaN) records, and precipitation ($R > 0$) records.

	Minutes	%
Total	1576800	100.00
not NaN	1255923	79.65
$R > 0$	48166	3.84

*Electronic address: marclladoduran@gmail.com

B. Preprocessing of data

From 18 possible precipitation types output by Parsivel2, 17 types are found in our data set (Table II). However, it is observed that type 62 does not have a clear sense, since it does not follow the classification method of intensity as it is reported by [2]. Moreover, it is an *intermittent* type of precipitation, making no sense for a minute series. Since 62 is a Rain type, we preprocess these raw data changing them into $61, 63$ or 65 (which are all Rain types), depending on the rain rate R, following strictly what [2] explains. Thus, we have worked with not 17, but 16 types of precipitation.

Since Code 4677 might be too detailed in order to make a first classification, it is added to every minute series an alternative code to classify precipitation in a more simple way, that we call *Code M1*. In fact, we will develop two classification methods, which will be called *M1* and *M2*. We will use Code M1 and Code 4677, respectively, for each classification method. A table of equivalences between codes M1 and 4677 is given in the Appendix (Table VI). It will be seen that M2 is actually obtained adding further detail to M1.

TABLE II: Minutes and percentage over total precipitation of each precipitation type according to PSV2 in Code M1.

Type	Minutes	%	Type	Minutes	%
Rain	33510	69.57	Soft hail	603	1.25
Mixed	10567	21.94	Drizzle	510	1.06
Snow	2816	5.85	Hail	160	0.33

III. CLASSIFICATION METHOD DEVELOPMENT

A. Pure precipitation types

In order to develop both classification methods we need relations between the particle speed, which is assumed to be terminal, and its diameter for each pure kind of precipitation. These are essentially empirical relations [3, 4]. Table III summarizes them. It should be noted that they are pure types, with no relation at the moment with Code M1. Precisely, Hail and Soft hail refer to pure Hail and Soft hail types, while in Code M1 mean Hail or Soft hail with or without other precipitation types.

Of course data does not follow exactly these relations, but with some degree of dispersion. Therefore we consider that classes whose v differ $30-50\%$ from the ideal relation of a particular type are as well within that precipitation type. Precisely, a 30% margin is given to types Rain and Soft Hail, while a 50% to Hail and Snow. By doing so, we obtain a way to get the counts of each pure precipitation type. The v-d classes whose centers fall into the margin of each precipitation type have been selected

TABLE III: Empirical relations used for the classification. The diameter range in which they are applied is given. Precipitaton types are pure, they have no relation with Code M1.

Prec. type	Fall speed [m/s]	d [mm]
Drizzle	$v = 9.65 - 10.3e^{-0.6d}$	$d \leq 0.5$
Rain	$v = 9.65 - 10.3e^{-0.6d}$	$0.5 < d \leq 8$
Snow	$v = 0.67d^{0.25}$ $v = 0.87d^{0.23}$ $v = 0.55d^{0.23}$	$0.5 \leq d \leq 14$
Hail	$v = 8.445(0.1d)^{0.553}$ $v = 10.58(0.1d)^{0.267}$ $v = 12.43(0.1d)^{0.5}$	$5 \leq d$
Soft hail	$v = 1.3d^{0.66}$ $v = 1.5d^{0.37}$ $v = 1.2d^{0.65}$ $v = 1.1d^{0.57}$	$2 \leq d \leq 5$

to create a mask assigned to that precipitation type. Applying these masks to the matrix of counts gives us then the counts of every pure precipitation type.

Some v-d classes have been assigned not a single pure type but two, in particular the combinations of Rain and Hail, and Snow and Soft Hail. We decide making Rain and Snow prevail in these cases, thus setting these classes as Hail and Soft hail negative. Figure 1 shows the classes corresponding to each pure precipitation type.

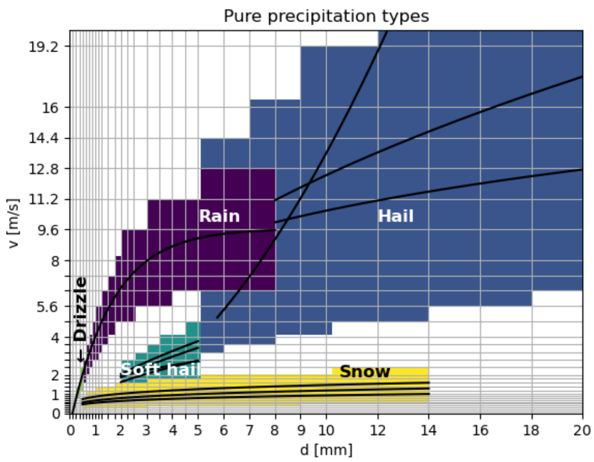

FIG. 1: Correspondence between $v-d$ classes and pure precipitation types: Hail (dark blue), Rain (purple), Snow (yellow), Soft hail (turquoise blue). Drizzle (green) is hardly visible since it follows the rain curve on the interval $0-0.5$ mm.

Since we have already decided which v-d classes correspond to each precipitation type, now the process is, for every single minute series, to apply each mask and to sum up the elements of the resulting matrices, getting hence the number of particles of each precipitation type for every minute series.

B. Quality control

A double quality control procedure is developed, one of them before the application of pure type masks and the other right after. Following [3, 5], masks for strong winds, splashing and margin fallers are prepared to detect and clean these cases.

The presence of strong winds makes the hydrometeors fall non-vertically, increasing the time the laser detects that particle and hence, computing a speed much lower than the normal. A first quality control is hence applied to detect large particles ($d > 5$ mm) falling unrealistically slowly ($v < 1$ m/s). A Boolean variable is created for every time series, being True when at least a particle with these characteristics is detected. Time series with potential strong winds are not considered for the classification.

The presence of splashing is caused by the impact of raindrops with fix parts of the Parsivel2, breaking into multiple tiny drops and being detected by the sensor. Classes with $d < 2$ mm and fall speeds below 60% of the v-d relation for rain are put to zero, considering that the presence of positive counts in those classes are likely to be caused by splashing. Finally, the opposite effect is produced by margin fallers, in which raindrops fall through the edges of the sample area and get accelerated. Therefore, classes with $d < 5$ mm and speeds above 60% of the v-d relation for rain are also excluded in the particle counting.

Right after the application of the first quality control and the counting of each pure precipitation type, a second quality control consisting on the presence of an isolate pure kind of precipitation. To this aim, we ask that for a determined time series with a particular pure type positive, at least that type should be positive as well within the interval of two minutes before and after. If this is not true, that counter is set to zero, considering it as a false alarm.

C. Assigning precipitation types

Once we have the pure type counts for every minute series we are ready to assign a precipitation code for both classifications M1 and M2. Since the main schema of the algorithm is shown in Figure 2, it is only explained what the model does inside the green ovals of Figure 2.

1. When pure rain or drizzle are positive, we assign Rain (Code M1) if rain is more frequent; Drizzle (M1) if drizzle is more than three times frequent than rain; and Mixed (M1) if none of the conditions is positive. M2 codes are assigned based on M1 classification and following [2] to select codes depending on the intensity rain rate R. In the case of Mixed, codes 58 and 59 (4677) are assigned depending as well on R.

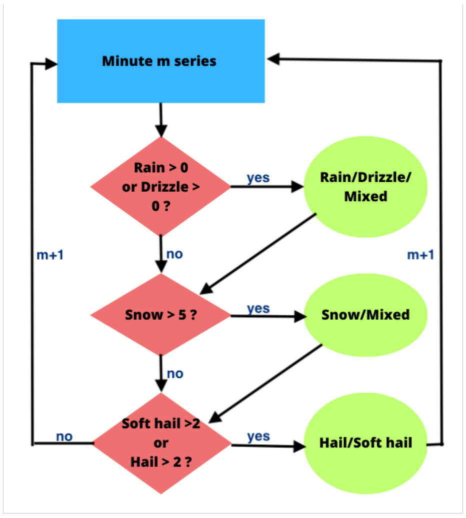

FIG. 2: Schematic decision tree for the assignment of precipitation types valid for both M1 and M2 classifications.

2. When snow is positive, we assign Mixed (M1) when the counts of liquid precipitation are 80% or higher than the counts of liquid plus snow. If not, Snow (M1) is assigned. As for Rain and Drizzle [2] is followed for M2 classification, being 68 and 69 the codes 4677 corresponding to mixed in this case.

3. When hail or soft hail are positive, Hail (M1) is assigned if it is positive and 10% of soft hail is not greater than hail. If it is not satisfied, Soft hail (M1) is assigned. Intensities and hence M2 classification is applied according [2]. It is worth noting that the presence of any other type of precipitation does not change the assignment of Hail or Soft hail, as they are not pure precipitation types, neither in M1 nor in M2.

IV. RESULTS

A. M1 classification

Verification scores are computed in order to see the reliability of the classification of each precipitation type. Table IV summarizes these coefficients for the classification M1. It can be seen that the type Snow is the one with best scores, but Rain, Drizzle and Soft hail work well enough. We should note that Hail has a low FAR, although the POD does not reach 10%. A reason might be that there is not enough data to perform the model to this type. However, the main reason seems to be related to the fact that many events of Hail are highly unpure, being the number of particles of hail much small than other types.

To illustrate the performance of M1, a case of study is

TABLE IV: Verification scores of M1 against PSV2 for each precipitation type in Code M1.

P. type	POD	FAR	ORSS
Drizzle	0.40	0.00	0.99
Rain	0.85	0.47	0.74
Snow	0.94	0.01	1.00
Soft hail	0.40	0.00	0.98
Hail	0.09	0.00	0.99
Mixed	0.33	0.13	0.52

given in Figure 3, where M1 and PSV2 are plotted. The data of this case is also used in [6], so it can be compared with that too. The behaviour of M1 is clearly pretty similar to the classification given by Parsivel2.

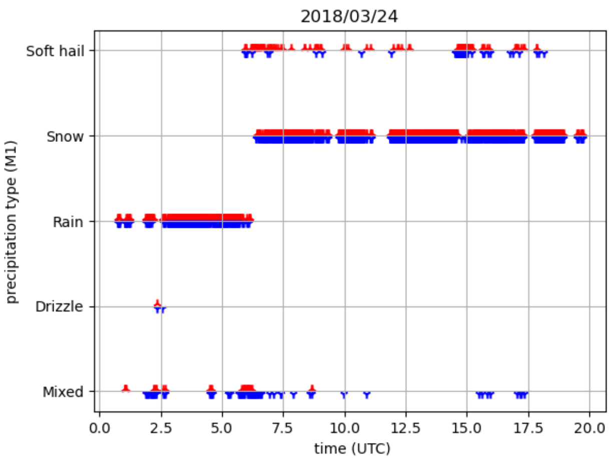

FIG. 3: Precipitation type given by PSV2 (red) and M1 (blue) during 2018/03/24.

B. M2 classification

Verification scores in M2 classification (see Table V) get worse compared to M1, since they are correlated but with more variables and types in M2. It is given the case of study (Figure 4), now according to Code 4677 and comparing M2 and PSV2 classifications.

We can observe that, while in M1 the Snow events are predicted with excellent verification scores and visually they fit well, in Figure 4 snow events, although detected, are not well classified according to the intensity of the event (snow codes are 71, 73, 75, with intensity increasing with the code number). Since the detection of the Snow type implies directly the detection of 71, 73 or 75 in the M2 classification, and given that to make this classification the variable R is used, there exists some kind of error in the data given by Parsivel2.

It is also worth noting that the verification scores for Hail and type 89 coincide, since Hail contains types 89 and 90, but there is not any type 90 in our data.

TABLE V: Verification scores of M2 against PSV2 for each precipitation type in Code 4677 which at least is detected once.

P. type	POD	FAR	ORSS	P. type	POD	FAR	ORSS
51	0.36	0.00	1.00	68	0.32	0.07	0.74
53	0.34	0.00	0.99	69	0.30	0.05	0.79
55	0.20	0.00	1.00	71	0.09	0.01	0.89
58	0.22	0.02	0.86	73	0.04	0.01	0.47
59	0.21	0.03	0.85	75	0.81	0.04	0.98
61	0.89	0.32	0.89	87	0.43	0.01	0.98
63	0.71	0.01	0.99	88	0.37	0.00	0.98
65	0.34	0.00	0.99	89	0.09	0.00	0.99

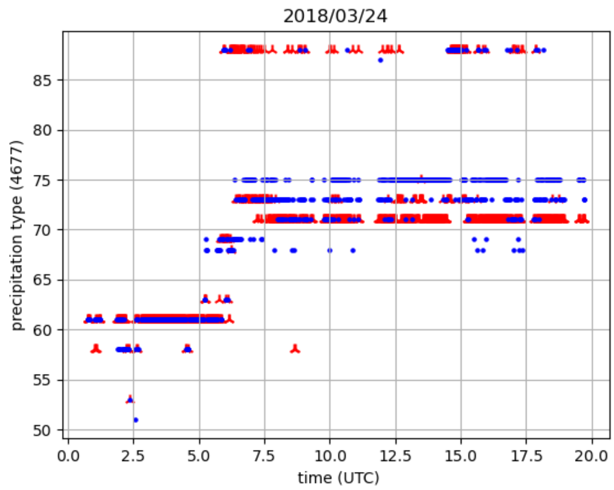

FIG. 4: Precipitation type given by PSV2 (red) and M2 (blue) during 2018/03/24.

V. CONCLUSIONS

- The presence of type 62 (4677) in PSV2 suggests that either Parsivel2 uses more data than the matrix of counts and R to perform the classification, or there is some kind of error in that data. Since 62 is an intermittent type, if it was right, it would mean that Parsivel2 knows the evolution of precipitation within a series of data, i.e. within a minute. Anyhow, it is contradictory with what [2] explains.

- There are many snow events with very high equivalent rain rates ($R > 200$ mmh^{-1}). Moreover, although they have high verification scores in M1, when one considers R to output M2 classification, verification scores fall down. Therefore, Parsivel2 uses some equivalent rain rate which is not exactly the one that gives us in its output.

- The sensibilities to Soft hail and specially to Hail are low. However, it is possible to increase these sensibilities by asking the mere presence of a single particle of these types. However, the payback is getting the other precipitation types scores worse.

- No potentially strong wind positive events have been found. In fact, it makes sense since the mask, which was inspired in [3, 5], was used for hurricane events. However, since in Das there is a complete meteorological station, it should be studied the dependence between the matrix of counts and the wind, because it might exist some kind of deviation.

- Probabilistic information associated to PSV2 classification seems to be nonexistent. It might be of interest since there exist many unpure types which are not clear, getting hence the comparison with our models more difficult.

VI. APPENDIX

A. Verification scores

A contingency table has been used to compare our models with the classification given by Parsivel2. 'Hits' represent the number of events positive according in either our model and PSV2. 'Misses' are the number of events negative in our models which are positive in PSV2. 'False alarms' are positive according our models but negative in PSV2. Finally, 'Correct negatives' are negatives in both classifications.

The Probability of Detection (POD) estimates how likely is a certain precipitation type to be detected. The False Alarm Rate (FAR), also named Probability of False Detection, tells us how likely is our model to give as positive an event which didn't occur. Finally, the Odds Ratio Skills Score indicates the forecast skills of the model compared to random chance. Best (worst) scores are 1, 0 and 1 (0, 1 and 0) respectively [6, 7]. POD, FAR and ORSS are computed with the following expressions:

$$POD = \frac{\text{hits}}{\text{hits } + \text{ missses}}$$

$$FAR = \frac{\text{false alarms}}{\text{false alarms } + \text{ correct negatives}}$$

$$ORSS = \frac{\text{hits} \cdot \text{correct negatives } - \text{ misses} \cdot \text{false alarms}}{\text{hits} \cdot \text{correct negatives } + \text{ misses} \cdot \text{false alarms}}$$

B. M1 and M2 equivalences

TABLE VI: Equivalence between Code 4677 and Code M1.

Code M1	Code 4677	Code M1	Code 4677
Drizzle	51, 53, 55	Soft hail	87, 88
Rain	61, 63, 65	Hail	89, 90
Snow	71, 73, 75	Mixed	58, 59, 68, 69

Acknowledgments

Firstly I want to thank my advisor, Dr. Joan Bech, for giving me the chance to work on such an amazing field. And last but not least, to my close family and friends. Thank you all.

[1] WMO Code Table 4677. (n.d.). National Centers for Environmental Information (NCEI). https://www.nodc.noaa.gov/archive/arc0021/0002199 /1.1/data/0-data/HTML/WMO-CODE/WMO4677.HTM.

[2] OTT Hydromet. (2016). Operating instructions. Present Weather Sensor OTT Parsivel2. *https://www.ott.com/es-la/productos/meteorologia-80/ott-parsivel2-272/*.

[3] Friedrich, K., Kalina, E. A., Masters, F. J., & Lopez, C. R. (2013). Drop-Size Distributions in Thunderstorms Measured by Optical Disdrometers during VORTEX2, *Monthly Weather Review*, 141(4), 1182-1203.

[4] Brandes, Ikeda, K., Thompson, G., & Schnhuber, M. (2008). Aggregate Terminal Velocity/Temperature Relations. *Journal of Applied Meteorology and Climatology*, 47(10), 27292736.

[5] Friedrich, K., Higgins, S., Masters, F. J., & Lopez, C. R. (2013). Articulating and Stationary PARSIVEL Disdrometer Measurements in Conditions with Strong Winds and Heavy Rainfall, *Journal of Atmospheric and Oceanic Technology*, 30(9), 2063-2080.

[6] Garcia-Benadi, A., Bech, J., Gonzalez, S., Udina, M., Codina, B., & Georgis, J.-F. (2020). Precipitation Type Classification of Micro Rain Radar Data Using an Improved Doppler Spectral Processing Methodology. *Remote Sensing*, 12(24), 4113.

[7] *Forecast Verification methods Across Time and Space Scales.* (s.f.). https://www.cawcr.gov.au/projects/verification/.