**Anonymous Referee #2**

Received and published: 5 January 2017

1 Summary

This manuscript proposes a data-driven approach to analyze rain-rate time series at the event time scale in order to link micro- and macro-physical properties of rainfall. After defining what is a rain event, a genetic algorithm is combined with a self-organizing map (SOM) method to identify the most informative descriptors of a rain-rate time series for a parsimonious approach. The obtained 5 descriptors are then used with the corresponding self-organizing map in order to "project" the initial 23-dimensional space into a 2-dimensional (map of neurons). An unsupervised clustering technique (hierarchical ascending clustering) is then applied to identify clusters in the neurons of the SOM, corresponding to clusters of rainfall events. Using 2 clusters, the usual convective/stratiform dichotomy is retrieved, and the authors proposed to use up to 5 clusters.
Taking advantage of the fact that the employed time series come from a disdrometer, the links between the rain descriptors or the neurons of the SOM and two important parameters of the drop size distribution (DSD) are investigated. In this way some relationships between rainfall macro and micro-physical properties are highlighted.

2 Recommendation

I enjoyed reading this manuscript (despite the quality of the English that must be improved) because the proposed approach is original and promising. Such characterization of rainfall events and the possible links between the macro and micro properties of rainfall are highly relevant to AMT readership and to the community in general. I have some relatively minor comments/suggestions listed below, I hence recommend to send the manuscript back to the authors for minor revisions.

3 General comments

1. The dimensionality reduction is well explained, but I did not find a quantification of the amount of information lost in the process. The obtained 5 descriptors and the corresponding SOM are optimal with respect to the criterion defined (topological error), but this optimum could be bad in absolute term (i.e., a significant amount of information is lost overall even if the selected descriptors/SOM are better than other combinations of descriptors/SOMs). I missed such discussion in Sec.3.1.

***The SOM algorithm aims at giving a low quantization error. First the map, is learned with a coarse training allowing the map to spread well (with no folding) in the data space, keeping the topology of the original data space. The second step, fine tuning, insures that the neurons become local averages of the data. This step insures a good quantization of the data.***
***The final map is a compromise between the topological error and the number of parameters. Since the topological error is computed on the total data space, it insures that the parameters used to learn the map also preserved the topology for most of the other parameters. This insures that data close with respect to the five final parameters are also close in the full/original data space. For parameters used to characterize natural events, like rain, ruled by the law of physics this implies that the map was able to learn the relationship between the parameters (even for the parameters not used in the mapping of the data space), otherwise the topology would not be preserved.***

***Unlike linear approaches, like principal component analysis, that uses an orthogonal transformation, the non-linear dimensionality reduction method uses a mapping quantification of the amount of information. A quantification of the amount of information lost in the process is quite complex to evaluate due to nonlinear relationships between variables. For this reason we have provided in table 4 the R2 determination coefficient for the whole variables. As it can be seen in this table the determination coefficient are rather good even for the unlearned variables. This guarantees the good behavior of the map and consequently only few information is lost.***

2. How transferable to other climatic regions are the results obtained from the presented analyses? Can interested reader use the exact same SOM in other regions or it should be recomputed to adjust to the local climatology?

***Of course novel data such as orographic and oceanic rain events would probability not be as well taken into account by the map. Nevertheless, the map was thoroughly validated. By making sure that the topology was preserved we made sure that the behavior of the map would be as good as possible. This point will have to be addressed in future work***

**The following sentences were added :**
**p11 L36 : …. But sufficiently heterogeneous between them. Of course this classification is obtained for mid latitude climate. The data set used in study is only representative of a particular region and topography (Ile de**

**France in France, mid latitude climate). Data driven analysis can not lead information on process that are not sampled in the data set, there are no orographic** rainfall **events nor oceanic observations which could present particular features and lead to additional specific clusters of events.. The last step …..**

**P15 L18 (conclusion) : The present study was conducted with observations from middle latitude area in plain region. The relevance of this classification needs to be confirmed with other data set collected in different climatic areas and for different meteorological situations encountered for example in mountain or coastal areas. This point will have to be addressed in future work.**

3. There are many grammar and vocabulary mistakes throughout the manuscript.
The authors must have the manuscript edited by a professional or a native speaker at least. I cannot list all of them but here are a few examples: precipitation without s, clusterS (p.2, l.34), “In a second time" (p.2, l.36), punctual should be point (p.3, l.2), “is more able for detecting" (p.5, l.10), “the variables those the components" (p.5, l.36)...

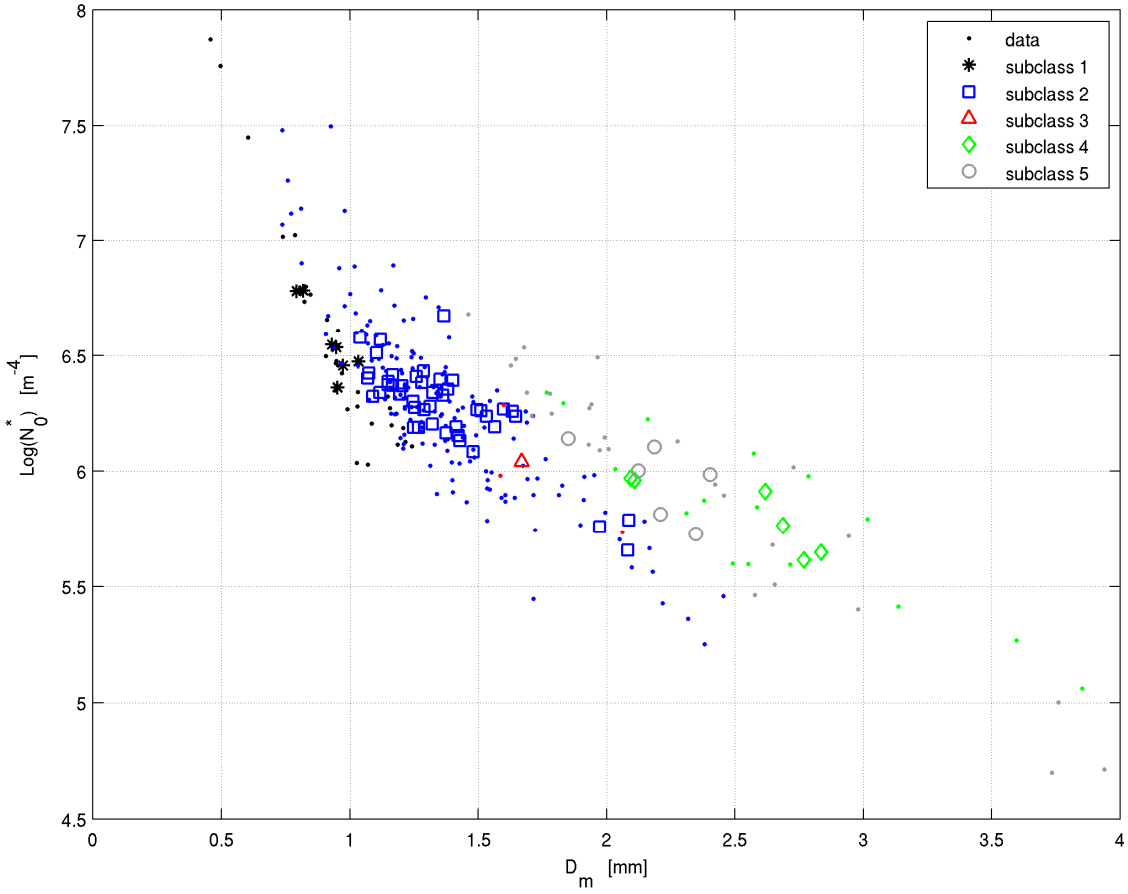
***We agree. A professional translator will edit the manuscript.***

4 Specific comments

1. Title: is microphysical information really derived from macrophysical information?
***It is mean that some microphysical properties can be inferred thank to the knowledge of the macrophysics subclass. Hence an event belonging to subclass 4 has high Dm (2.7 mm) and low N0* (5.7)***

Figure 9 shows that there is a link between a given neuron and (Nw , Dm ) but we do not know how the events “attached" to a given neuron are spread in the (Nw , Dm ) space.

***Here is Figure 9 with added data points. The points are colored with the color of the class they belongs to. We can see that for each class the corresponding neurons are quite well spread among the data belonging to the same class.***



2. P.1, l.38-39: dimensionality reduction implies more or less information loss. What can be discarded exactly may differ from one application to the other... Hence the intended application may be important.

***The dimension reduction indeed implies potential information loss. Nonetheless, the large number of parameters implies redundancy. Dimensionality reduction is intended to reduce this redundancy thus part of the information loss can be considered as denoising.***

***P1, l.38, we replaced "to select the most relevant features" by "to select a reduced set of features"***

***Concerning what can be discarded or not, let's consider the unlearned IETp variable because we knew that it would not bring pertinent information on a rain event.. The resulting map is not structured with respect to the IET which is, as it should be expected, considered as noise. Thus any loss of information associated to an impertinent parameter cannot be considered in the end as a loss of information. An impertinent parameter even if learned is inconsistent with the rest of the information used to describe an event. It is thus discarded during the learning process. Hence the resulting map is not structured according to this variable.***

***We intended a general approach aimed toward the retrieval of microphysical information. We believe that this approach is generic and that the resulting map, due to the specificity of the algorithm, should also be useful for other applications. Either way, for a specific application, if a dedicated data set is provided, it should be used to apply the algorithm.***

3. P.2, l.8: microphysics does not reduce to the DSD (which corresponds more to the microstructure of rainfall). The use of microphysics in this context is a bit ambiguous and confusing.

We agree. These sentences :
**"Usually the microphysics is characterized by the raindrop size distribution, noted N(D), which is defined by the number of raindrops per unit of volume and per unit of raindrop diameter (D). Actually, information on rain microphysics is often displayed through proxies of N(D) as it will be explained further. At the present time, the rain microphysics features are not accessible by rain gauges which only provide macrophysical information."**
***are replaced by:***
**"One of the key information for remote sensing is the raindrop size distribution, noted N(D), which is defined by the number of raindrops per unit of volume and per unit of raindrop diameter (D). Actually, information on raindrop size distribution is often displayed through its proxies as it will be explained further. At the present time, these kinds of features are not accessible by rain gauges which only provide macrophysical information."**

4. P.2, l.13: there are more than a few disdrometers worldwide! Please rephrase.

***We agree.***
**"Around the world, there are few disdrometers where there are tens of thousands of rain gauges."**
***is replaced by***
**"Around the world, there are tens of thousands of rain gauges where there are a lot less disdrometers."**

5. P.2, l.20: some disdrometers allow the estimation of rain rate (and other variables) at higher temporal resolution than 1 min.

**"Disdrometers provide drop size distribution and consequently they allow estimating one-minute rain rates which are the measurements used to get hydrological information."**

***is replaced by:***

**"Disdrometers provide drop size distribution and consequently they allow estimating one-minute (or less) rain rates which are the measurements used in this study to get hydrological information."**

6. P.3, l.10: a rain event will also strongly depend on the considered spatial and temporal resolution. You work at the point scale, but a rain event could also be defined over a given area (using model grids for instance). This should be mentioned I think.

***"Indeed a rain event will depend on the sensor characteristics (specific surface caption, detection threshold, instrumental noise)."***
***is replaced by:***
**"Indeed a rain event will depend on the sensor characteristics (specific surface caption, detection threshold, instrumental noise) as well as on the spatial or temporal resolution chosen for the study."**

7. P.3, l.30: how sensitive are the results to this MIT value of 30 min? As mentioned above, the spatial scale probably has an influence on the relevant MIT value.

***We have run some quick test for various values of the MIT, but we have not really investigated this point and consequently we are not really in position to answer to this question. A greater MIT value will aggregate rain periods together and consequently will modify some parameters like the event duration 'De' or the Rain event Depth 'Rd'. Some others parameters can remain unchanged or not, this is the case for example for parameters Pci (which is sensitive to the variability of the rain rate). We expect that a higher MIT may aggregate events of a different type, whereas a shorter MIT tends to increase the number of events while they belong to the same group of rain cells***

8. P.4, l.7: the term "descriptors" could also be used here.

***Changed***

9. P.4, l.27: could you provide some quantitative information about the goodness-of-fit to the normal distribution of the different transformed variables? Are they close enough to Gaussian distribution?

***The transformations are empirical. Our algorithm or the PCA do not rely on the gaussianity of the data. Properly speaking, there is no statistical or probabilistic framework in this article.***
***(No estimator, no confidence interval or no whatsoever are involved.)***
***As it is written P.4, l.25 and P.4, l.27 the final distributions are "quasi-normal" or "the closest to a normal distribution". As it is stated P.4, l.26 the transformations were chosen "empirically" on the basis of the kurtosis and skewness estimation. Before using any learning algorithm, the data must be checked and transformed to be fitted for the algorithm. Here the transformations are simply used to contract/dilate the data space to guarantee a better spreading of the map among the data.***
***Having transformed distributions, close to normal, also helps for the interpretation of the PCA.***

10. P.4, l.33: "learning data set": it is not defined... I guess it is a subset of the total data sets, but how was it obtained?

***The learning and test sets are defined at the end of section 2.1 where a reference is made to the Table Tab.1..***

***"on the learning data set."***
***is replaced by***
***"on the learning data set (cf. end of section 2.1 and Tab.1)."***

11. P.4, l.33 - p.5, l.2: is this paragraph about PCA really necessary?

***The PCA is a well-established technique known from most. Many readers may wonder what could have been done with a PCA. Since it could be counterproductive the PCA is done. It is also a way to display the redundancy in the parameters. It also allows to introduce the fact that some parameters (such as IETp) are potentially non-informative.***

12. P.5, l.30: vector should be denoted in bold font.
ok

13. P.5, l.32: why 60 chromosomes?
***This value is not critical and any even value can be used. A smaller number of chromosomes involves more generations to converge. The GA theory shows that the algorithm converges to the same solution whatever the number of chromosomes.***

14. P.5, l.35: "training data set": same as above for learning set, it is not defined...
Training data set is replaced by learning data set

15. P.5, l.36-37: "Once training each ... variables": I do not understand this sentence, it seems there is a syntax issue.

***There is indeed a syntax issue.***
***"Once training each of the 60 Maps"***
***is replaced by:***
***"Once trained each of the 60 Maps"***

16. P.6, l.6: could you provide the functional form of te(xk )?
***Additional information is now provided in sections 3 and 4.***
*the topological error* $te(x^k)$ *is computed according to eq. 2 in Uriarte and Martín (2008)*

17. P.6, l.30: “describe quite well the original space": could you provide quantitative information on this aspect? Based on what can you state this?

***The discussion concerning the representativeness of the 5 selected variables is discussed in sections 4.1 & 4.2. Consequently this sentence has nothing to do here. The sentence is replaced by the following one :***

**At the 187th generation we get a subspace formed by 5 variables namely : Event duration D_e (#1), Standard deviation σ_R (#9), rain rate peak in event R_max (#11), Rain event depth R_d (#13), Absolute rain rate variation P_c1 (#14).**

18. P.6, l.28-32: if I am correct, the 5 selected variables are transformed ones, so their physical interpretation may be slightly less straightforward than suggested in the text.

***As told previously the transformations are simply used to contract/dilate the data space to guarantee a better spreading of the map among the data. It has an impact on the distances between the data. It does not affect the ordering of the data. It will not impact the interpretation of the map.***

19. P.7, l.21: why 8 × 8 neurons?

***64 neurons is a compromise. When we use the topological error we aim at a fine mapping of the data space. A smaller map would imply a poorer description of the data space. We have 234 rain events. 64 neurons imply more or less 3 to 4 data points to compute a neuron referent. We could not have more neurons. Choosing a square map implied not making hypotheses on the shape of the map that could have been representative only of our non-exhaustive data set.***

20. P.8, l.21-27: is this valid for all climatologies or just for the one studied here (temperate mid-latitudes)?

***It is not valid for all climatologies. The text has been clarified.***
***“These authors have noticed that successive rain and no rain periods are found to be uncorrelated, thus a rain time series can be considered by an alternation of rain event and no rain independently drawn periods. That is similar to say that inter-event time (IET) doesn’t characterize the rain events. Brown et al. (1983) also investigated a possible dependence between IETp and the intra-event characteristics and they conclude that the assessment of their data gave no indication that such dependency exists.”***
***is replaced by:***
**“These authors, working in temperate mid-latitudes for relatively short periods, have noticed that successive rain and no rain periods are found to be uncorrelated, thus a rain time series can be considered by an alternation of rain event and no rain independently drawn periods. That is similar to say that inter-event time (IET) doesn’t characterize the rain events. For other places and other climatologies it is less clear, Brown et al. (1983) also investigated a possible dependence between IETp and the intra-event characteristics and they conclude that the assessment of their data gave no indication that such dependency exists.”**

21. P.10, l.1: the delineation could be illustrated in Figure 6.

***Added***

22. P.10, l.19: maybe you could add the coordinates (2,0.7) in the (Rm ,βL3 ) space to help the reader.

***It has been modified.***

23. P.10, l.32: “body rain disturbances"?

***It has been corrected.***
“the synoptic precipitations caused by the mid-latitude depressions are an example of stratiform precipitations. ***They manifest themselves in the body rainy disturbances associated with warm and cold fronts.”***
***is replaced by:***

"the synoptic precipitations caused by the mid-latitude depressions are an example of stratiform precipitations **they manifest themselves in disturbances due to warm and cold fronts."**

24. P.12, l.11: given the definition provided in Eq.6, Dm is the mass-weighted diameter.

***You are right.***
***"This variability is represented by two microphysics parameters namely the mean volume diameter (Dm) and the parameter N*."***
***is replaced by:***
**"This variability is represented by two microphysics parameters namely the mass-weighted diameter (Dm) and the parameter N*."**

25. The quality of figures 2, 3 and 6 should be improved.

***We agree. The resolution has been improved and the size of the font changed.***