Anonymous Referee #3

Received and published: 11 January 2017

Paper revision. Title: Data driven clustering of rain events: microphysics information derived from macro scale observations. Author(s): M. D. Dilmi et al. MS No.: amt-2016-389

In this paper, the authors present a data-driven approach to analyze the rain events in order to characterize them both from a micro- and macro-physical point of view.
The authors use a disdrometer dataset of 545 events (according to their definition of event), which is divided in two uneven groups, one for learning and on for testing.
A Genetic Algorithm (GA) is combined with a self-organizing map (SOM) method to identify a number of indicators (from a list of 23 micro- and macro-variables) able to full describe each event. A previous step led to identify the independent indicators though a Principal Component Analysis (PCA). They found that the best performance is obtained by using 5 macro indicators. A hierarchical cluster analysis is then applied to the 5 indicators subset to cluster all the events in the two "commons" convective/stratiform groups and in five finer groups.

I found the manuscript interesting because the authors use a new technique (moreover derived from a different science field) and even because they aimed to link micro- and macro-physical variables to describe a rain event. Generally, the quality of figures and tables is good, but I suggest to box all the figures and to increase the quality of figure 1a. However, some questions and doubts arise reading the manuscript. I suggest the publication of the manuscript after the authors address the following points.

We partially agree. The figure 1 was re-drawn with a better resolution and boxed. We chose not to box the other figures.

- Generally, the English has to be improved. Several errors are found by reading the text and I suggest you to review the manuscript by a native speaker. Some errors are very basic and could be just writing error (i.e. "is justify" Pag.1 line 16, the lacking of "s" in some third person verbs, etc.). Please revise the expression like "one can see/consider/note. . .." and check the correct use of the singular/plural words.

We agree. A thorough revision of the manuscript will be done by a professional translator.

.

- What I mainly miss within the manuscript are two aspects: what does it happen if the Minimum Inter-event Time (MIT) is chosen different from 30 minutes? What I ask here is if the authors have been conducted a sensitivity study to show the influence of the MIT on the results. This could be very useful if the technique is applied to a different definition of event.

Our choice was made based on the literature. Even if it cannot be called a sensitivity study, we also run some test to see if the conclusion drawn in the lite nevertheless literature were coherent with our data set.
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

The second point is related to the measurement instrument. All their analysis are based on disdrometer data that are not so widespread with respect to the rain gauges. Moreover, they found that the best performance is obtained by five macrophysical indicators, which are also a rain gauge outputs. So, do the authors think that they get the same results if the rain gauges data are analyzed? It could be interesting, if they have some rain gauges data available, to apply their technique to this type of data. It is well known that rain gauges have some problems in measuring very light and heavy precipitation. How this can impact on the results?

It is one of the perspectives of this work. It would give access to more exhaustive (as well in space and in time) data sets. We are working on this aspect. It has to be noticed that the nature of the measurements is different. (For example, a tipping bucket rain gauge is measuring the time to fill a given volume. It has some important implications for stratiform rain events.)
Since the resolution is less good it will definitely impact the results. It will strongly impact some parameters such as the Rmax and the PC parameters. Since Rmax and PC_1 are two of the five parameters selected by the algorithm it will impact the results. This is nevertheless a study work in itself.

- In the section 3.1, the authors describe the methodology used. Even if it is quite well explained, I suggest to slightly improve so that every reader can be able to correctly reproduce it (i.e. they should better explain what is the topological error).

We agree. Section 3 has been partially re-written.

Minor comments:

- Page 2, line 13: “Around the world, there are few disdrometers. . .”.
The expressions is incorrect even if I understand the will of the authors to highlight the high ratio between the number of disdrometers.

The sentence :
“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.”

- Figure 3: I do not understand the colorbar values. Can you better explain them?
The colorbar represents the average distance between a particular neuron and its neighbors
The caption of Fig. 3 is modified:
The color of each neuron represents the average distance between the neuron and its neighbors.