Atmospheric
Measurement
Techniques
Discussions

Open Access

Interactive comment on "Data driven clustering of rain events: microphysics information derived from macro scale observations" *by* M. D. Dilmi et al.

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.

âĂć 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.

âĂć 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. 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?

âĂć 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). Minor comments: âĂć 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.

âĂć Figure 3: I do not understand the colorbar values. Can you better explain them?