D. Dunkerley (Referee)
david.dunkerley@monash.edu
Received and published: 29 December 2016

- This paper presents an analysis of a set of disdrometer data (from 2008-2014) processed to yield a 1 minute rainfall series. This was then grouped into 545 simple rainfall events using an inter-event time of 30 minutes (page 3 line 31). Half of the events were used for 'learning' using some grouping methods, and the other half for testing the grouping methods.
From the literature, 17 characteristics of rainfall events (such a duration, depth, or mean rainfall rate) were explored as criteria for characterizing events. Since some of the characteristics are correlated, the authors sought to identify a smaller set of more parsimonious variables that might be sufficient to characterise rainfall events.
In terms of method and approach, the paper seems to be generally sound and to employ useful methods that are well applied.

An aspect that I found to be missing was a discussion of the purpose for which rainfall events might need to be characterized. After all, it seems reasonable to think that this should guide the selection of parameters that might be meaningful for particular applications. For instance, in hydrology, it is well known that rainfall events cannot be characterized adequately through the use of simple descriptors such as average rainfall rate or peak intensity. Rather, measures of the time-distribution of rain and no-rain periods within an event are important, as is the position of the intensity peak(s) within the event – either early or late, for instance.
These characteristics in turn affect the partitioning of water when rain arrives at the soil surface (e.g., far more of the rain tends to become overland flow if the largest intensity peak occurs late in the rainfall event, and much less becomes overland flow if the intensity peak is early).

We agree on the fact that, given an application, some parameters clearly carry more information than others. This article mainly aims at validating a methodology able to select the most parsimonious subset of variables representative of the whole. The 23 parameters of this study are clearly not exhaustive. They encompass the parameters usually used by the authors themselves, to which are added those derived from a bibliographic study. They nevertheless allow to validate the methodology. In further studies it will be possible to improve the set of parameters to test their ability to better characterize the rain events. It is also possible to adapt the parameter set with a given application in mind.

Often parameters are the result of the knowledge of various given physical/environmental contexts. When dealing with a data set, the researcher often has to relate his case to the numerous situations encountered in the literature. It can be puzzling to choose the most appropriate parameters. The methodology allows to automatically bring out a reduced set of pertinent parameters. This article illustrates a way to discriminate among numerous parameters.

***The SOM algorithm implies that each neuron of the map gather similar data, rain events in our case. By selecting the best map according to the topological error we make sure that the events gathered by a neuron are close to the events of its neighbors. By doing so we make sure that the map is well spread in the data space.
The map can be seen as a classification of the rain events. The five parameters we hold on to are the one best minimal subspace preserving the topology and thus the similarity in the original data space.
The underlying hypothesis is that the most valuable information is the one which best represent the similarity/dissimarity in the original data space.
This assumption is based on the fact that the parameters are redundant and sufficiently coherent to learn a meaningful map. This is validated by the validation process.***

The following sentence is added p.4 L7
Of course this set of 17 features is not exhaustive and some other features could be added depending the application. For example is hydrology the positions of the intensity peaks inside the event could be a relevant feature.

The following sentence p1 L36 is modified

The first goal of this study is to select the most relevant features to characterize the events through a data-driven approach without taking into account the application context and thus to characterize a rain event in the most parsimonious and efficient possible way.

New formulation

The first goal of this study is to select the most relevant features to characterize the events through a data-driven approach without taking into account the a priori knowledges of the field of application and thus to characterize a rain event in the most parsimonious and efficient possible way.

- A useful typology of 5 classes of event emerges from the numerical analyses presented in the paper, and I thought that this formed a significant and useful contribution of the paper. The link to convective and stratiform precipitation is well explored. However, here again, I felt that material was missing that might have been included in the paper. For instance, orographic rainfall can exhibit very different characteristics from other forms of precipitation, such as convective rain. This can include very long event durations and very prolonged rain of consistent or rising intensity. The authors give the impression that they only recognize convective and stratiform rainfall as end members.
Likewise, they don't consider rainfall over the oceans, which has some temporal characteristics that distinguish it from terrestrial rainfall. As a result, I felt that the authors should offer some caveats about the extent to which they argue that their approach and conclusions might, or might not, be more widely applicable than to the single geographical location (and rainfall climate) represented by their disdrometer data. These caveats should be reflected in the title of the paper, which should be less sweeping, and perhaps refer to stratiform and convective rainfall, and/or to the particular study region (France) and its particular rainfall event characteristics. Of course, the authors also neglect seasonal and inter-annual changes in rainfall event character, and this also warrants mention and possibly some discussion. The authors seem to consider that their results are in some way definitive and universally applicable. I doubt this, and would like them to consider how their results might have changed had they used a larger data set, including data from different rainfall climates.

We agree, we were not clear enough on the purpose of this work. The conclusions drawn in this article follow from the data set used in the article which is representative of a region (Ile de France in France). There is no orographic rainfall ***events***
Given the size of the data set we could not take seasonal and inter-annual changes in rainfall event.
When parting the complete rain event set in two. To insure a correct generalization we made sure to take different time periods. When looking closely to the data, we notice that the 2 periods are noticeably different which ensure a better generalization. With a larger data set, it would be interesting to study the seasonal and the inter-annual information encompassed in the map.

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.

Even if it is far from exhaustive the data set gather a large variety of events. If the map was learned on a more exhaustive data set first the number of data would increase implying the possible use of a bigger map. It would also imply a richer panel of data most certainly with a wealth of behavior not present in the current data set. In this case, new parameters better characterizing these new rain events ought to be added to highlight their specific properties.

For example, concerning a very long and consistent Orographic rain event, the current five parameters stressed out in the article may be enough to characterize such an event. If we were to take into account something as specific as a rising intensity a dedicated parameter ought to be added.
Adding new data should modify the data density allowing the rise of a neuron specific of a new class of events. Stressing the preservation of the topology in the algorithm is done in this sense.

Once the methodology validated and vetted, we are going to work on larger data sets potentially exhaustive. It is a work in itself.
The following sentences were added in the conclusion : see below

Some caveats were added
p11 L36 : …. But sufficiently heterogeneous between them. Of course this classification is obtained for middle latitude climate. The data set used in this study is only representative of a particular region and topography (Ile de France in France, temperate climate). Data driven analysis cannot 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. If the SOM was learned on a more exhaustive data set the number of data would increase allowing the use of a bigger map allowing to represent possible new behaviors which are not present in the current data set. This point will have to be addressed in future works

- I recommend some minor revision to address the above points.

The written English could be improved in places. For instance, ‘criteria’ is the plural form; when referring to a single parameter the word to use is ‘criterion’ (singular).
‘Dysfunction’ (page 3 line 36) should be ‘malfunction’.

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