Atmos. Meas. Tech. Discuss.,
doi:10.5194/amt-2016-389-RC2, 2017

Atmospheric Measurement Techniques Discussions

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

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 tot 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.
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 re-

gions or it should be recomputed to adjust to the local climatology?

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)...

4 Specific comments

1. Title: is microphysical information really derived from macrophysical information? Figure 9 shows that there is a link between a given neuron and (N_w, D_m) but we do not know how the events “attached" to a given neuron are spread in the (N_w, D_m) space.
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.
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.
4. P.2, l.13: there are more than a few disdrometers worldwide! Please rephrase.
5. P.2, l.20: some disdrometers allow the estimation of rain rate (and other variables) at higher temporal resolution than 1 min.
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.

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.
8. P.4, l.7: the term “descriptors" could also be used here.
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?
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?
11. P.4, l.33 - p.5, l.2: is this paragraph about PCA really necessary?
12. P.5, l.30: vector should be denoted in bold font.
13. P.5, l.32: why 60 chromosomes?
14. P.5, l.35: “training data set": same as above for learning set, it is not defined...
15. P.5, l.36-37: “Once training each ... variables": I do not understand this sentence, it seems there is a syntax issue.
16. P.6, l.6: could you provide the functional form of $te(x^k)$?
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?
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.

19. P.7, l.21: why 8×8 neurons?
20. P.8, l.21-27: is this valid for all climatologies or just for the one studied here (temperate mid-latitudes)?
21. P.10, l.1: the delineation could be illustrated in Figure 6.
22. P.10, l.19: maybe you could add the coordinates (2,0.7) in the (R_m,β_{L3}) space to help the reader.
23. P.10, l.32: "body rain disturbances"?
24. P.12, l.11: given the definition provided in Eq.6, D_m is the mass-weighted diameter.
25. The quality of figures 2, 3 and 6 should be improved.