Atmospheric
Measurement
Techniques

Open Access

Discussions

EGU

AMTD

Interactive
comment

Interactive comment on "Machine learning for improved data analysis of biological aerosol using the WIBS" *by* Simon Ruske et al.

D. Baumgardner (Referee)

darrel.baumgardner@gmail.com

Received and published: 9 July 2018

Developing techniques for extracting information from UV-LIF measurements is an important endeavor and the more that can be tested, the better.

The current study follows in the footsteps of an earlier exercise that tested a number of methodologies in the search for optimum approaches that can discriminate PBAP types.

Although the current study appears to have made some additional progress I am of the opinion that there are a number of aspects to the data and the PBAP properties that are missing and that should be included before this study is published.

1) Table I and II are both labeled 2008 2) If Tables I and II are actually 2008 and 2014,

then there needs to be a third table that summarizes the properties that are shown in Figs. 4&5, not only the averages but their variances, as well. These need to be listed for both years for the same test particles because from an examination of the figures, it certainly appears that the properties are quite different for the same biotypes. If this is indeed the case, then it is no surprise that there are different results using the various clustering methods for the two data sets. 3) Why are just FL1,2 and 3 used. In Hernandez et al (not cited here, unfortunately), we found that FL1&2 and FL 1&2&3 are important markers. Leaving them out seems like a loss of useful information. 4) Bioaerosols are by their nature irregular in shape and in their fluorescing. Why isn't the variance also used as a parameter in the clustering?

I think that additional work remains to further separate by general categories within bacteria, fungi and pollen. In our analysis of the lab results we were able to quite clearly separate the three bio types just by fluorescence type and size without any sophisticated machine learning. I would have to assume that this can be improved upon using more sophisticated approaches like the current study.