

Interactive comment on “Evaluation of Machine Learning Algorithms for Classification of Primary Biological Aerosol using a new UV-LIF spectrometer” by Simon Ruske et al.

D. Baumgardner (Referee)

darrel.baumgardner@gmail.com

Received and published: 1 September 2016

This study compares a broad variety of techniques for pattern recognition of bioaerosols. The description of these technique is not rigorous but probably sufficient for the reader to get a general idea of how they work and how they differ from one another. There are adequate references for the interested reader to pursue more information on how to apply the techniques.

If the purpose of this paper is to educate those readers interested in selecting the optimum technique for processing bioaerosol measurements, in my opinion it falls short of the mark because too little analysis accompanies the results. My suggestions on how to improve the paper:

C1

1) Explain why one of the most common machine learning techniques was not tested, i.e. neural networks.

2) Evaluate which of the particle metrics provides the most information on the different types of particles by each of the techniques, i.e. is it the size, a particular wavelength, a combination of wavelengths? For example, in Hernandez et al., Fig. 3 shows that bacteria and fungi have similar fluorescence signatures but differ by size, whereas fungi and pollen have similar sizes but differ in fluorescence signatures. Given that the MBS has 8 fluorescence channels, and the purpose of more channels than the WBS is that it provide more information, then this needs to be validated with these processing techniques.

3) Which of the lab bio types was the most efficiently identified, bacteria, fungus or pollen?

4) Could any of the techniques separate between the individual bacteria or pollen species?

5) The real surprise is that removing the shape information not only didn't reduce the ability to separate bio types but in some of the techniques removing this improved their efficiency. So the question is what would happen if you reduced the number of wavelengths down to the three of the WBS? Given the number of WBS users, that would be of great interest. The Hernandez study suggests that maybe three wave lengths are sufficient to resolve the three general bio-types but probably not to differentiate within a species. As it is, the reader will conclude that the shape detector is not needed. Would that be a valid conclusion?

6) The paper concludes with a brief mention of processing time but I think that this is a critical topic that needs to be included since real-time identification of bio-types is an important application of this technology so there needs to be an evaluation of the detection efficiency versus processing time.

C2

