Atmos. Meas. Tech. Discuss.,
doi:10.5194/amt-2018-109-RC3, 2018

Atmospheric Measurement Techniques Discussions

Interactive comment on "Evaluation of a Hierarchical Agglomerative Clustering Method Applied to WIBS Laboratory Data for Improved Discrimination of Biological Particles by Comparing Data Preparation Techniques" *by* Nicole Savage and J. Alex Huffman

Anonymous Referee #2

Received and published: 3 June 2018

This manuscript discusses application of Hierarchical Agglomerative Clustering (HAC) to analysis of data collected using the Wideband Integrated Bioaerosol Sensor (WIBS-4A). While real-time detection of bioaerosols has been quite well controlled, the analysis and classification is still challenging and vital problem. Therefore, investigation and improvements in this area are very important and crucial for understanding the abilities and limitations of LIF aerosol detectors. The manuscript is well written and in detail reveals important problems of fluorescence data analysis of bioaerosols. I rec-

ommend presented manuscript to publication, however some corrections and further explanations to the following remarks will be appreciated:

1. The techniques of single particle detection using LIF devices, like WIBS, reached relatively high reliability and perfection. The device collects data in real time, on the other hand the presented results are offline. The data analysis takes a long time. Finally, the standard methods like particle collection on tape is still competitive with LIF. My question is: Did the authors try or are going to apply real-time aerosol data analysis? 2. L67 - principle or principal component analysis? 3. L116 – "The WIBS collects 3 channels of fluorescence intensity. . ..." – collect channels or collects fluorescence intensity in 3 channels? 4. L170 – ". . .both saturating and non-fluorescent particles were retained. . ." – Did authors collect the particles? 5. L370 – ". . .gains. . ." or grains? 6. L494 - ..fluorescence and non-fluorescent particles.. - The phenomenon should not be compared with the property. 7. L 424 and further – I think that term "synthetic mixtures" for recorded numerical data is confusing and should be corrected. Firstly, it sounds like a chemical synthesis process. Secondly, the final result of clustering should be the same and independent whether the particle data are sorted or not. Otherwise, the order (sequence) of detected particles would change final result. I think that actual meaning of used data is well described in L298-300 ("...subset taken from the pool of particles..". 8. L 426 – "analytically synthesized" – analysis has opposite meaning to synthesis should be corrected 9. L 428, 431, 434, 436, – ". . .mixture synthesized. . ." – see point 7. 10. The authors compared clustering ability using selected small groups of substances. It would be interesting to see the clustering output for all 14 types together. Why it was not presented?

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2018-109, 2018.