

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

**Anonymous Referee #3**

Received and published: 9 May 2018

This paper describes methods and results which should help improve the interpretation and use of data obtained with UV-LIF instruments such as the WBS. The WBS measures light scattering, a light-scattering asymmetry factor, and fluorescence in three channels. Fielded instruments with data rates that can exceed hundreds of particles per minute are available. This paper uses a large set of WBS data measured for individual materials (Savage et al. 2017) to evaluate different preprocessing procedures for analysis of such data. Mathematical simulations of externally mixed particles of

C1

known composition are studied. The findings should be useful not only for understanding WBS data, but more broadly in applying Hierarchical Agglomerative Clustering to some other problems in aerosol analytical chemistry. I recommend publication. However, I request that several confusing items be made less confusing.

The use of the term “synthetic mixtures” (L31-32, L424, 707, L734) is confusing. Chamber studies with synthetic mixtures of real aerosols and real gases are not uncommon in aerosol science. A google search of “synthetic mixture” provides discussions of various real “synthetic mixtures.” I only looked at the first 8 or so items in that search, but I saw none with the meaning used in this paper. The online dictionaries I saw do not indicate this use of “synthetic” (which as far as I can tell indicates something about numerical or computational). Synthetic organic chemists make real chemicals. If “synthetic mixtures” is used for the simulated data investigated here, what terminology is left for researchers to use when they make real synthetic mixtures of aerosols in a chamber and investigate changes in clusters as time passes and as particles agglomerate? I do not see how a reader can see from the abstract or even well into this paper that “synthetic” is being used in this highly non-standard way, and that Savage et al., 2017 did not measure mixtures of particles. The “synthetic mixtures” are actually numerical (or mathematical) simulations of the WBS the data that should be obtained for dilute mixtures of particles. Real mixtures of particles can form agglomerates, and some may agglomerate quickly unless they are sufficiently dilute.

L 20-22 (Abstract). “Here we show for the first time a systematic application of HAC to a comprehensive set of laboratory data collected using the wideband integrated bioaerosol sensor (WBS-4A) (Savage et al., 2017).” Suggest change to: “Here we show for the first time a systematic application of HAC to a comprehensive set of laboratory data collected for individual particle types using the wideband integrated bioaerosol sensor (WBS-4A) (Savage et al., 2017). Here the WBS data for single-composition aerosols is combined numerically to generate data to simulate WBS values for mixtures of aerosol.”

C2

L31-32 (Abstract): “Lastly, six synthetic mixtures of four to seven components were analyzed.” Might be changed to: “Numerical simulations of mixtures of four to seven components were HAC analyzed.”

L424: “Investigating cluster ability to separate complex synthetic mixtures” Might be changed to: Investigating the capability to separate particles in simulations of complex synthetic mixtures

L426-429: “To better simulate real-world scenarios, we analytically synthesized six mixtures of particles by pooling existing data from selected particle types in prescribed ratios. Each mixture was synthesized to roughly represent a different hypothetical mixture of particles that might be expected.” “Analytically” suggests equations or functions were used in obtaining the data for the mixtures. Isn’t “numerically” or “computationally” what is meant?

L426-429 might be changed to: “To better simulate real-world scenarios, we numerically simulated six mixtures of particles by pooling existing WIBS data from selected particle types in prescribed ratios. Each simulated mixture was assembled to roughly represent a different hypothetical mixture of particles that might be expected. Also, the particles in each simulated mixture are assumed to be so dilute that any agglomeration is negligible.” Also, a significant fraction of readers read the abstract and then look at the figures to see what the results will be. Adding clarifying words to the figure captions and tables would be useful.

I don’t know what “normalized to particle size” means here. Please clarify, possibly with an equation. Please also give the ranges of error in particle sizes expected. Why is scenario D worse than B? I think it is because D adds noise to the FL signals, making them less informative by decreasing the S/N. This added noise occurs in the elastic scattering measurements, and also results from the approximations used in estimating solutions to the inverse problem for size (with unknown shape, orientation and refractive index). If the scattering measurement and the solution to the inverse problem were

C3

perfect, then D and B should give very similar results, at least for spherical particles and some methods of normalizing to particle size and shape. It may be useful to cite a paper or data with WIBS measurements of size and fluorescence for uniformly-sized fluorescent PSL. For a single size of PSL, do plots of the WIBS-measured scattering and fluorescence fall on a line or are they spread more randomly? Even for a spherical PSL particle, with known refractive index, would you suspect that the noise is large enough to make D less useful than B?

Can the authors say anything about the length of times bacteria or fungal spores might last in an urban environment before a significant fraction of the bioparticles combine with soot, and how that might affect the usefulness of the WIBS? I’ll be very interested to see the results when (sometime in the future) the authors inject bacteria or fungal spores into a chamber, add soot particles, use the WIBS to sample with time, and then repeat the some of the analyses in this paper with the results given as a function of time.

L23: In abstract: “ratio” of what? In the text, “ratio” first appears in “distance ratio.” Suggest change first use of “ratio” in abstract to “ratio of particle concentrations.”

L117: please add wavelength ranges of FL1 to FL3. Aim for a little broader set of readers.

L171: replace “will be” with “were”.

L199: Suggest change to: Ambient particle number vs size distributions can often be well approximated by lognormal distributions (citation), although specific subsets of particles, such as bacteria, pollens or fungal spores, may not exhibit lognormal distributions.

L245: “placed into a conceptual pool”? How about, “A subset of the particles were selected randomly for analysis”?

L258-259: “diesel soot particles . . . commonly observed . . .” Is this referring to WIBS

C4

measurements? Please provide a citation(s).

L299-300: Do you mean: "In each case the input particles are a random subset . . ."

Please also note the supplement to this comment:

<https://www.atmos-meas-tech-discuss.net/amt-2018-109/amt-2018-109-RC1-supplement.pdf>

---

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