Anonymous Referee #3

Received and published: 9 May 2018

Note regarding document formatting: black text shows original referee comment, blue text shows author response, and red text shows quoted manuscript text. Changes to manuscript text are shown as *italicized and underlined*. Bracketed comment numbers (e.g. [R1.1]) were added for clarity. All line numbers refer to discussion/review manuscript.

[R3.0] This paper describes methods and results which should help improve the interpretation and use of data obtained with UV-LIF instruments such as the WIBS. The WIBS 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 WIBS 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 known composition are studied. The findings should be useful not only for understanding WIBS 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.

[A3.0] Author response: We thank the referee for her/his positive summary of the manuscript and recommendation to publish after comments are addressed.

[R3.1] 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 WIBS 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.

[A3.1] This is a good point that we had not previously considered. The same point was raised by Referee #2 [R2.7, R.2.8, and R2.9]. We removed all use of the term “synthetic mixtures” and changed most instances of the term to “simulated mixtures.” Note that this comment also impacts comments [R3.3] and [R3.6].

[R3.2] 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 (WIBS-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 (WIBS-4A) (Savage et al., 2017). Here the WIBS data for single-composition aerosols is combined numerically to generate data to simulate WIBS values for mixtures of aerosol.”

[A3.2] The text of the abstract was modified as suggested.

[R3.3] 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.”

[A3.3] The text of the abstract was changed as requested to:
“Lastly, six *numerical simulations of* ~~synthetic~~ mixtures of four to seven components were analyzed *using HAC*.”

[R3.4] 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

[A3.4] The sub-title was changed along the suggested lines to:
“Investigating *the capability* ~~cluster ability~~ to separate *particles in simulations of* complex ~~synthetic~~ mixtures”

[R3.5] 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?

[A3.5] The word “analytically” was changed to “computationally.”

[R3.6] 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.

[A3.6] These are good suggestions that add clarity to the text. The section was re-written with the suggested text. Words “computational” or “numerical” added to captions of several figures and tables to increase clarity, as suggested.

[R3.7] [a] 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. [b] 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 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?

[A3.7] To clarify the first question [a], additional text was added to L207:
“…fluorescence intensity was normalized to particle size *(by dividing fluorescence intensity value by light scattering signal when a particle interacts with the diode laser beam)* in order to …”

With respect to the second question [b], the referee is likely correct that results for Scenario D (fluorescence normalized) are worse than for Scenario B (fluorescence not normalized), because for Scenario D additional uncertainty with respect to size is propagated into the intensity value. Normalizing in this way would also propagate uncertainty for field measurements, and so given the poorer results of the tests analyses represented here we chose not to further explore parameters represented by Scenario D.

[R3.8] 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.

[A3.8] This an interesting question, but we do not have a good answer to the hypothetical thought about atmospheric lifetimes of these particles at this point. It would be great to explore external mixing of different particles types in the future in order to see how these mixtures could further influence fluorescence and particle size properties observed by instruments like the WIBS. This is beyond the scope of the experimental process for now.

[R3.9] 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."

[A3.9] Text edited as requested.

[R3.10] L117: please add wavelength ranges of FL1 to FL3. Aim for a little broader set of readers.

[A3.10] This was also requested by Referee #1. Additional text was added, as shown here:
"The WIBS collects 3 channels of fluorescence intensity information (FL1, FL2, and FL3), particle size, and particle asymmetry for each interrogated particle. *The bands of excitation and fluorescence emission are: FL1 (λ_{ex} = 280 nm, λ_{em} = 310 – 400 nm), FL2 (λ_{ex} = 280 nm, λ_{em} = 420 – 650 nm), and FL3 (λ_{ex} = 370 nm, λ_{em} = 420 – 650 nm)."*

[R3.11] L171: replace "will be" with "were".

[A3.11] The phrase "will be" changed to "*is*" to match correct tense.

[R3.12] 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.

[A3.12] Text revised as suggested.

[R3.13] L245: "placed into a conceptual pool"? How about, "A subset of the particles were selected randomly for analysis"?

[A3.13] Text was changed, as suggested, to:
"For each trial, a *subset* ~~given number~~ of particles from each material type was *selected randomly for HAC analysis* ~~placed into a conceptual pool before running through the algorithm to organize clusters~~."

[R3.14] L258-259: “diesel soot particles . . . commonly observed . . .” Is this referring to WIBS measurements? Please provide a citation(s).

[A3.14] The text as originally written was indeed over-stated and confusing. The text has been revised to the following:
“The first two trials include diesel soot particles, because *light-absorbing carbon aerosol* ~~they~~ are commonly observed in ~~almost all~~ *aerosol* ~~atmospheric~~ samples with ~~even minimal~~ anthropogenic influence *(Bond et al., 2013)* …”

[R3.15] L299-300: Do you mean: “In each case the input particles are a random subset . . .”

[A3.15] Yes, the words “number of” was inserted incorrectly here and the typo was corrected as suggested by the referee.