August 13, 2018

Dear Editor,

Below is a composite file with point-by-point responses to all three anonymous referees and an additional community member. Following that is the revised manuscript document with all changes tracked from the originally submitted version. We are confident that you will find the revised document to be well improved after systematic and comprehensive revisions following all the review period. Please let us know if you have additional comments or questions.

Best regards,

Alex Huffman

Point-by-point responses:

Anonymous Referee #1
Received and published: 1 June 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.

[R1.0] This paper builds on existing literature examining unsupervised learning techniques to improve the interpretation and classification of data obtained with WIBS UV-LIF spectrometers. As shown in previous publications, Hierarchical Agglomerative Clustering (HAC) can serve as a robust data analysis method for classification/interpretation of bioaerosol data but the accuracy of technique is highly sensitive to the choice of clustering linkage and data pre-treatment (e.g., Crawford et al., 2015); this is further explored in this paper which elucidates how data pre-treatment choices such as choice of fluorescent threshold and log normalising data may influence clustering accuracy using laboratory samples of known particle types (Savage et al., 2017) in various synthetic mixtures, and thus the authors present tentative recommendations of data pretreatment regimes depending on the analysis goals. Overall the paper is well written and the computational experiments well thought out. The findings here are useful and further validate the usefulness of Hierarchical Agglomerative Clustering for interpretation of WIBS data. The results also provide a useful framework for testing Hierarchical Agglomerative Clustering data pre-treatment regimes for other atmospheric science data problems and neatly demonstrate the potential pitfalls of not rigorously performing such tests. I recommend publication after the following comments have been addressed.

[A1.0] Author response: We thank the referee for her/his positive summary of the manuscript and recommendation to publish after comments are addressed.

Specific comments
[R1.1] L73-77: The authors have conflated some of the terminology relating to unsupervised and supervised leaning methods. I'm uncomfortable with the use of the term clustering when discussing supervised methods as clustering specifically relates to cluster analysis. I suggest replacing "clustering techniques" with "classification algorithms" and "(trains) the clustering algorithm" with "(trains) the classification algorithm".

[A1.1] The referee raises a good point. We changed terminology on page 2 according the referee suggestions, as listed below:

- L68: "*Classification algorithms, including several* clustering techniques in particular, have shown successful results …"
- L73: "~~Clustering techniques~~ *Classification algorithms* can be divided …"
- L76: "This type of method enhances (trains) the ~~clustering~~ *classification* algorithm in that the output ~~cluster classes~~ *groups* are predetermined …"

[R1.2] L120: Please state the bands and what they relate to.

[A1.2] Additional text was added, as shown below:
"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). The excitation and emission wavelengths chosen for each of the 3 fluorescence channels were designed to maximize the information gained about key biological fluorophores present in a broad range of bioparticles (Kaye et al., 2005; Pöhlker et al., 2012). *Early generations of UV-LIF bioaerosol spectrometers were often interpreted to be able to detect proteins via channels similar to FL1 and products of active cellular metabolism (i.e. riboflavin and NAD(P)H) via channels similar to FL3, but these approximations are gross simplifications that confound more detailed investigation of particle types.*"

[R1.3] L198: Can the authors please clarify why they have used log spaced bins. Do you mean that you have taken a log of the data and it is binned naturally by the discrete nature of the detector resolution (i.e., fine bins) or have you binned the data into specific (coarse) log bins? If it is the latter can you please state what the bins are and can you comment on how forcing the data to in bins may influence the clustering? My concern here is that too coarsely binning the data may create artificial hotspots due to reduced resolution and bias the clustering, reducing the capacity to differentiate between particles with similar properties. Can the authors comment on this and demonstrate the effect this may have by providing an example for comparison where the data is converted to log space and not binned. I also wonder if the bins should be normalised by the bin width to further complicate matters.

[A1.3] Aspects of this discussion are presented in L209-212. To summarize in different words, the data values from a given channel were either used as recorded (i.e. "value") or as logarithmically transformed (i.e. "log(value)"), depending on the Scenario. The values were not forced into specific bins, but rather input into the cluster algorithm using the exact value in either of these forms. The reason that logged values can provide different results by HAC is that the distance between points is different in linear space or log space, because the cluster process does not independently take into account whether a value is as recorded or as log(value). Because many real-world particle variables can present normal distributions only in log space (i.e. lognormal size distributions), we explored inputting values in both raw and log forms.

The following sentence was added to the manuscript at L211 for clarity:
"*By this process, data values were input into the algorithm as log(value), but without additional binning*."

[R1.4] L254: Can the authors comment on the environmental applicability of the chosen ratios. I would suspect that in an urban environment you may expect something closer to a ratio of 1:99 fungal to diesel particles with the converse being true in a forest environment. How does the clustering perform under such extreme mismatches?

[A1.4] We originally explored three different ratios of particle concentrations (80:20, 50:50, and 20:80) for each of the three match-ups discussed in Figure 3 in order to show that input ratio can be important to how the algorithm responds. This was certainly not intended to be exhaustive, and one could additionally explore more extreme ratios. So to limit the scope of the analysis here, we chose to present evidence only that the ratio matters, without trying in all cases to predict ratios that could be relevant to a wider range of ambient environments.

The question the referee brings up is interesting, however, and so we explored 1:99 ratios of each of the three particle type combinations presented in Figure 3, where the bioparticle is the minority concentration in each experiment. The results are shown below in a plot/table form identical to how they are presented in Figure 5. The Bacteria:Diesel and Fungi:Dust separations still performed quite well (6.6% and 13.5% misclassification, respectively), even with the extreme mismatch of input concentrations. The Fungi:Diesel separation was poor, however, in a 2-factor

solution, because the Diesel particles split into both clusters, and the Fungi particles were likely too low in concentration to influence the cluster properties. We added text including a summary of these new experiments to the manuscript at L304:

"To extend the investigation of particle input ratio, the three match-ups presented in Figure 3 were investigated using Scenario B with 1% bioparticles and 99% non-bioparticles in each respective case. In these experiments the Bacteria:Diesel and Fungi:Dust particles separated relatively well (6.6% and 13.5% misclassification, respectively). The Fungi:Diesel separation was poor, however, because the Diesel particles were nearly evenly split into both clusters, and the Fungi particles were too low in concentration to influence the cluster properties. More investigation is needed to explore how extreme disparities in particle ratio could negatively influence cluster quality in real-world settings."

Part A: Individual Clusters (Particle Number)

	Cluster	B3	F2	S4	D12
Fungi : Diesel	1	-	37	2588	-
	2	-	0	1111	-
Bacteria : Diesel	1	57	-	4	-
	2	0	-	5653	-
Fungi : Dust	1	-	45	-	7
	2	-	12	-	5650

Part B: Grouped Clusters (Particle Number)

	Cluster	Bio	Non-bio
Fungi : Diesel	1	37	2588
	2	0	1111
Bacteria : Diesel	1	57	4
	2	0	5653
Fungi : Dust	1	45	7
	2	12	5650

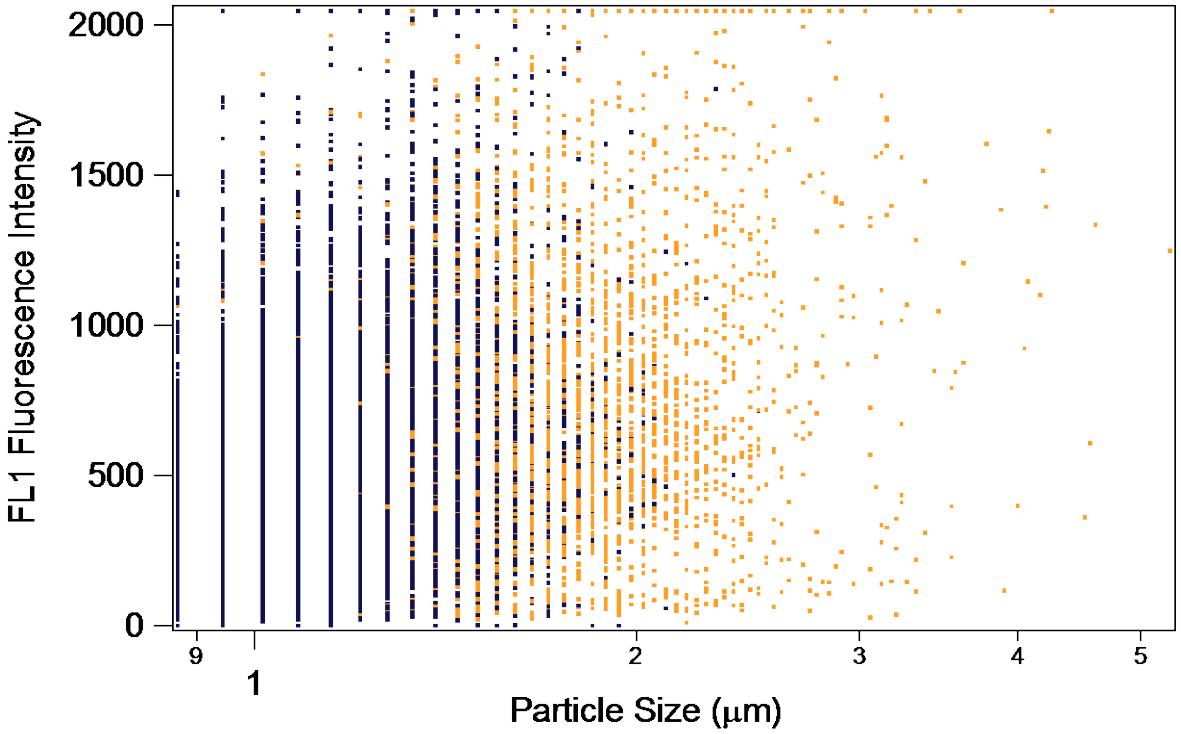
Part C: Summary (Cluster Quality)

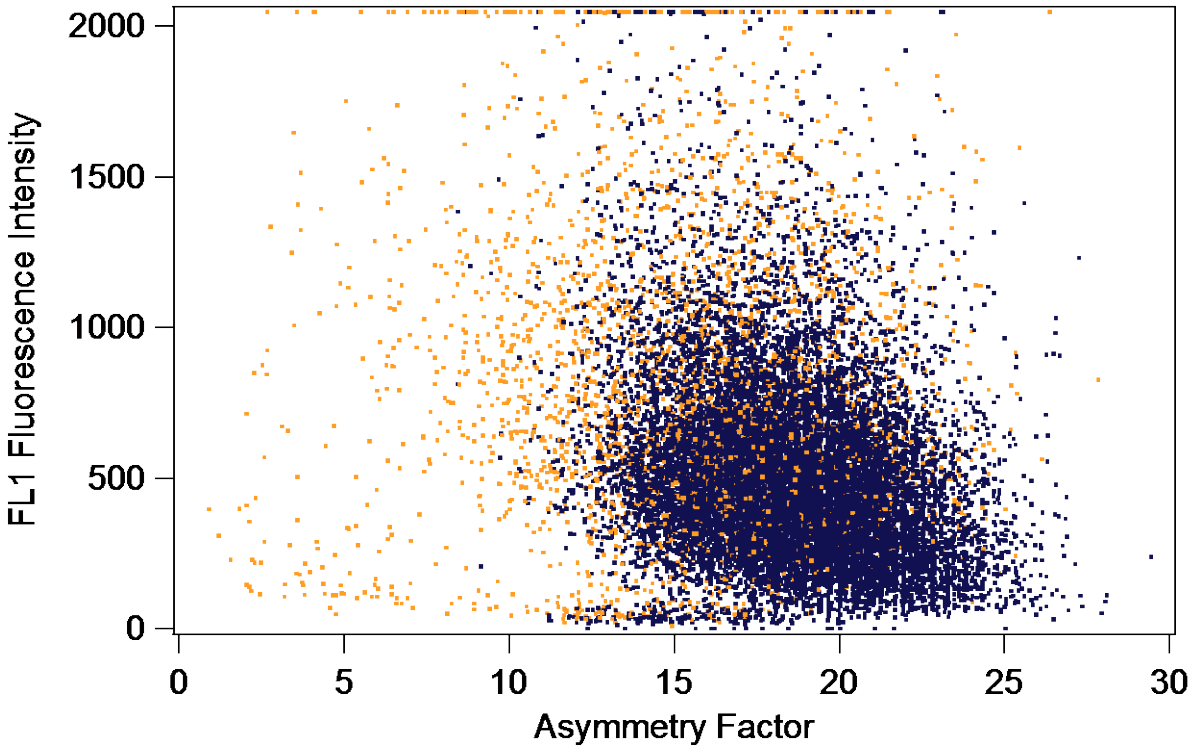
	Total P.	Miscl.	Cat.
Fungi : Diesel	2625	98.6%	Fungi
	1111	0.0%	Diesel
Bacteria : Diesel	61	6.6%	Bacteria
	5653	0.0%	Diesel
Fungi : Dust	52	13.5%	Fungi
	5662	0.2%	Dust

[R1.5] L238: Would it be possible to show examples of the cluster centroids for a case where there is significant misclassification? This may illuminate why the algorithm is failing to correctly attribute particles. It may also be useful to examine the fluorescence/AF characteristics of each cluster as a function of size here. A 2D histogram or color density plot could show distinct hot spots that haven't been separated correctly and could provide a basis for manual separation based on sensible thresholds.

[A1.5] To address the referee's suggestion, we included an additional set of plots here as suggested. The results below correspond to the match-up between Bacteria 1 and Bacteria 3 using Scenario B and the 3-sigma threshold, which corresponds to Experiment 22 from Table 2 (65% misclassification). The two colors of dots in the plots represent clusters 1 and 2. In this case it is still unclear how to utilize a single threshold to separate between the two particle types here.

In the process of analyzing results of this study we produced countless plots and tables, each of which showed slightly different angles of the same story. We chose to simplify the results in many cases to make the manuscript shorter and more manageably readable. We find that the table of fluorescence intensity and AF median values (Table 2 from original data published in Savage et al., 2017) often summarizes the differences in the particle types rather well and so were rarely able to separate using 2D histograms as the referee suggests. One example of these two additional plots is included here for reference, however.





[R1.6] L312-315: Can you describe the method for producing the soot as they seem rather large as compared to that in the study of Toprak and Schnaiter (2013) which were also coincidently found to be weakly fluorescent in FL1. Perhaps the soot used in this study is larger and more fluorescent than we may

expect of ambient/urban soot which may cause some of the difficulty in correctly attributing in in some cases?

[A1.6] The method for aerosolization of particle types discussed was presented in Section 3.2 of the associated Savage et al., 2017. Specifically, the aerosolization details related to soot are copied here:

> From Page 4284, Section 3.2.3 of Savage et al., 2017: "Dry powders were aerosolized by mechanically agitating material by one of several methods mentioned below and passing filtered air across a vial containing the powder. For each method, approximately 2.5–5.0 g of sample was placed in a 10 mL glass vial. For most samples (method P1), a stir bar was added, and the vial was placed on a magnetic stir plate. Two tubes were connected through the lid of the vial. The first tube connected a filter, allowing particle-free air to enter the vessel. The second tube connected the vial through approximately 33 cm of conductive tubing (0.25 in. inner diam.) to the WIBS for sample collection."

The referee is correct that the method of producing/aerosolizing particles, including soot, will bear heavily on the fluorescent properties observed. In particular, different aerosolization methods are likely to produce very different size distributions, which then will dictate the overall fluorescence properties. For this reason, we included the following statements in the Savage et al., 2017 paper:

> From Page 4292, Section 4.3: "It is important to note, however, that the method chosen for particle generation in the laboratory strongly impacts the size distribution of aerosolized particles. For example, higher concentrations of an aqueous suspension of particle material generally produce larger particles, and the mechanical force used to agitate powders or aerosolize bacteria can have strong influences on particle viability and physical agglomeration or fragmentation of the aerosol (Mainelis et al., 2005). So, while the absolute size of particles shown here is not a key message, the relative fluorescence at a given size can be informative."

The referee points out that the work by Toprak and Schnaiter (2013) presented small soot particles that also exhibited relatively weaker fluorescence in FL1. This is consistent with the expectation that fluorescence intensity will scale strongly with particle size. Differences in particle size could also impact clustering separation properties somewhat, and so further investigation of clustering using multiple narrow size ranges of different types of particles could further explore this process. This exhaustive process was beyond the scope of this work, however.

To make sure these points are clear in the revised manuscript we have added the following text at L327:
"It is also important to note here that the method of aerosolization for each particle type plays an important role in the observed size distribution and so results involving laboratory particles should be interpreted with this in mind. Observed fluorescence properties, in contrast, are expected to be conserved at a given particle size and intrinsically related to particle composition."

[R1.7] L384: Would we expect to be able to differentiate between 2 different particles of the same type with such coarse spectral resolution?

[A1.7] The referee's implied point is correct. No, we would not expect to be able to separate between very similar types of particles using such coarse resolution as is available in the WIBS.

Frankly, the fact that HAC paired with WIBS data was able to separate as well as it did was somewhat remarkable and surprising. To make the point clearer, we added text at the end of that paragraph as follows at L390:
"…separating more finely to quantify differences between types of individual biological particles is ~~likely to be~~ significantly more challenging *and not likely to be possible in most situations.*"

[R1.8] L415: Again I wonder if the use of too coarsely separated bins may compromise the 9-sigma thresholding and cause misclassification?

[A1.8] This question also loops back to [R1.3] and stems from a miscommunication. Values of the five WIBS data parameters were not separately binned (either during the logging process or when used as recorded), but are input into the cluster algorithm in the same spacing provided in the raw output of the instrument. The bin resolution is therefore limited by the WIBS optics and PMT settings.

Further, fluorescence intensity is relayed by a integer units between 0 and 2047, and resolution is not a limiting factor. For example, see Figure 5 of the Savage et al. 2017 paper. Biological particles typically exhibit median fluorescence intensity much higher than non-biological particles, thus using different threshold strategies can help separate particle classes from one another by this strategy.

[R1.9] L514: Can the authors comment on the applicability of their findings to new high resolution UV-LIF instruments that are beginning to become commercially available. Some of these new instruments have significantly more channels/greater fluorescent resolution than the WIBS.

[A1.9] This is a helpful suggestion. To extend the applicability of results, the text was amended as follows:
"Results here are ~~only~~ generally extendable to other UV-LIF instruments, *however, whether they offer single or many channels of emission spectral resolution, in that the methods of particle pre-preparation and the impact of particle number ratio are likely to relay similar effects on clustering strategy."*

[R1.10] Technical corrections
L63: instruments, not instrument.
L370: grains, not gains.
L112: Suggest "Experimental and Computational Methods"
L131: "each of the three"
L181: "was the best"

[A1.10] All typos corrected.

Anonymous Referee #2

Received and published: 3 June 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.

[R2.0] This manuscript discusses application of Hierarchical Agglomerative Clustering (HAC) to analysis of data collected using the Wideband Integrated Bioaerosol Sensor (WIBS4A). 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 recommend presented manuscript to publication, however some corrections and further explanations to the following remarks will be appreciated:

[A2.0] Author response: We thank the referee for her/his positive summary of the manuscript and recommendation to publish after comments are addressed.

[R2.1] 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?

[A2.1] I think the statement that "LIF devices … reached relatively high reliability and perfection" is already an very optimistic statement, but I agree that when operated and analyzed properly the data can often be useful. The referee's suggestion about real-time data analysis is an interesting idea that has been discussed. We are working on this type of analysis from a different angle and with respect to a different class of instruments, but we have not had the ability to investigate real-time analysis strategies with respect to WIBS data. This would be a worthwhile project, but is outside the scope of what we were aiming to accomplish in this study and would likely require dedicated project funding.

[R2.2] 2. L67 - principle or principal component analysis?

[A2.2] In this case the word "principal" is the correct one. I often get this word confused with "principle" and have to look up the definitions to make sure I'm correct.

[R2.3] 3. L116 – "The WIBS collects
3 channels of fluorescence intensity. . ..." – collect channels or collects fluorescence intensity in 3 channels?

[A2.3] This was indeed poor grammatical construction. The sentence has been changed to: "The WIBS collects *information about* ~~3 channels of~~ fluorescence intensity ~~information~~ *in three channels* …"

[R2.4] 4. L170 – ". . .both saturating and non-fluorescent particles were retained. . ." – Did authors collect the particles?

[A2.4] We did not physically collect the particles. The wording here was unfortunately confusing. In this case we have "retained" the data in the analysis process by not removing particles based on certain attributes. To clarify, the word "retained" was changed to "analyzed" as shown here: "*… both saturating and non-fluorescent particles were analyzed* ~~retained~~ …"

[R2.5] 5. L370 – ". . .gains. . ." or grains?

[A2.5] This is a typo; "gains" was corrected to "*grains*".

[R2.6] 6. L494 - ..fluorescence and non-fluorescent particles.. - The phenomenon should not be compared with the property.

[A2.6] This typo was changed for the discussion version of the manuscript to be "fluorescent and non-fluorescent particles."

[R2.7] 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..".

[A2.7] The term "synthetic mixtures" is indeed confusing terminology, and this is a point raised also by Referee #3 (i.e. [R3.1], [R3.3], and [R3.6]). Referee #3 suggested the term "computational simulations" or "simulated mixtures" among several possibilities, and we have changed the text in a variety of places through-out the manuscript to reflect this new terminology.

[R2.8] 8. L 426 – "analytically synthesized" – analysis has opposite meaning to synthesis should be corrected

[A2.8] Here the term was changed to "computationally simulated."

[R2.9] 9. L 428, 431, 434, 436, – ". . .mixture synthesized. . ." – see point 7.

[A2.9] The word "synthesized" was changed to "simulated" in each of these cases and all others within the manuscript.

[R2.10] 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?

[A2.10] This additional experiment might be interesting, but it is unlikely to add anything to the general nature of the conclusions. The 14 types of particles assembled for these match-up experiments (i.e. Sections 4.1 – 4.3) were meant to be individually instructive, but not to represent the entirety of the types of particles one might see in a more complex, ambient environment. So collecting all 14 into one experiment would represent another experimental combination, but would in itself not be any more relevant than the individual simulations already discussed.

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.

Public Comment- Simon Ruske (simon.ruske@student.manchester.ac.uk)
Received and published: 3 June 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*. All line numbers refer to discussion/review manuscript.

[Public Comment] The study presented is an extremely well structured and written investigation into the use of Hierarchical Agglomerative Clustering for classification of biological aerosol using a UV-LIF sensor, and will make an excellent addition to the literature upon publication.

[Author Response] Simon, thanks for taking the time to read and comment on the manuscript. We appreciate the useful comments, which will help improve the quality of the manuscript. We respond to each comment in detail below.

However, the authors may have made a small error [L161-L162] where they state that the conclusions for Ruske et al. (2017) were for ambient data, whereas in the abstract they correctly state that the study was on standardised laboratory particles [L19-L20]. Please could you correct this prior to final publication.

I apologize for this mistake. I am not sure where this error came in our writing process, but I removed the incorrect statement, as requested: "Their conclusions, however, were based on ambient field data using unknown particle types and did not investigate laboratory generated particles of known origin."

In addition the authors may wish to consider the following comments prior to publication.
[L78-L79] Would it be possible to clarify the starting conditions for supervised learning you are referring to? Hyper-parameter selection is an extremely important consideration for neural networks, but other supervised techniques such as decision trees and ensemble methods do exist where low classification error can be attained without providing the algorithm with any initial conditions other than the training data.

This may have been a bit of a miscommunication. We do not deal with any supervised learning methods in this manuscript. We trust your team as the experts in this area. Nicole simply wanted to provide a few sentences of general contrast between supervised and unsupervised methods. That is also why we pointed to your 2017 paper in this section. We have also included citation of your manuscript currently being reviewed in AMT.

[L84-L85] Is it necessary to apply unsupervised techniques to assess the advantages of supervised methods? Do you mean that supervised techniques require laboratory data of known types to assess their advantages? A very important disadvantage of supervised techniques is that they rely on adequate training data, and it is not clear at this point how much training data will be required to adequately represent an ambient environment, which is the point I think you are alluding to here.

This is the way I understand some of the pros/cons of supervised and unsupervised. I agree that the community (probably you first) will continue to lean about how this all works together and how well lab-generated data can be useful to train supervised data algorithms. As you well know, the differences between nicely behaving lab particles and more complicated particles collected in the field confounds most areas of aerosol science to some degree. So these problems will not necessarily be trivial to solve, but I think collectively we are all learning little pieces that will help.

[L186 - 187] Does the z-score rely on the assumption of normality? The z-scores of a normal random variable will be normally distributed whereas the z-scores of a non-normal random variable will be non-normally distributed. **Applied to any data set, regardless of distribution, the resultant variables after z-scoring will have mean of 0 and standard deviation of 1.** Is the purpose of standardising the data to prevent one of the variables from dominating in the analysis or to produce normally distributed data?

> Thanks to your prompting, we looked into these details and learned a bit more, which has been helpful to us. You are right that the way we characterized the z-scoring process was not correct. Talking back and forth with the university statistician, we now understand that values can indeed be input scaled to a normal distribution or not. We chose to standardize our variables to a mean of 0 and a variance of 1 so that the output variables would be on comparable scales, but this is also not the same as rigorously normalizing them in the rigorous sense. As a result, we have removed the statement you correctly indicated was inaccurate and updated the sentence as follows:
>
> Original text: "Standardization using the z-score method compares results to a normal (Gaussian) population, ~~and therefore relies on the assumption that input data can be described by a normal distribution (Gordon, 2006).~~"
>
> Updated text: "Standardization using the z-score method compares results to a normal (Gaussian) population, and we have chosen to standardize our variables to a mean of 0 and a variance of 1 so that the output variables would be on comparable scales."

[L203] It would be worth noting that in Crawford et al., 2015, there are particles for which negative measurement of fluorescence was recorded. The option of logtransformations may have been overlooked, as the logarithm is undefined for negative values. This was not intended to imply an assumption of normality, although this assumption has been stated explicitly in Robinson et al., 2013. In these cases would you recommend translating the fluorescence measurements to a range bounded below by 1, or alternatively would it be more appropriate to reject measurements for which the fluorescence produced was negative? It is also important to note that even if the data is log transformed, the data will still have a finite range due to the saturation point on the detector, and hence the data will have a truncated normal distribution rather than a normal distribution, and depending on how often saturation occurred there may still be a peak to the right hand side of the distribution. It is however, perfectly acceptable to apply HAC when the assumptions for best performance are not met as stated in Norusis, 2011.

> My understanding is that negative fluorescence values can be observed after subtracting some threshold value from the fluorescence intensity data. Instead of subtracting the data and looking only at positive values, we did the same thing by filtering the data at several discreet thresholds. This gets around the problem of negative values. In any case, we looked at three thresholding scenarios (Table 3), i.e. no threshold, 3 sigma, and 9 sigma. The ultimate result is that we found the most consistently positive results to be as a result of 3 sigma filtering, but this could be different in other situations. You are correct about the fact that particles that exhibit saturation of the detector in any channel will truncate a normal distribution.

[L222] How often did the CH index conclude that there were 2 clusters? When the CH index concluded a number of clusters other than 2, how much of an impact did this have on the quality of the results? Were the two cluster solutions always the best solution?

> We did not explore solutions that had more than 2 solutions, simply as a matter of limited time. There are certainly many scenarios in which individual bioparticle types (i.e. pollen, in many instances) can split into two reasonable clusters by themselves, and so independently allowing 3

or more cluster solutions could significantly improve results in many cases. We just didn't have the time to do this systematically, and so we chose to limit analysis to only 2 clusters in all cases. To help clarify this point, we added text at:

L227: "In order to reduce the length and complexity of *discussion, analysis of results in Sections 4.1-4.3 was limited to using cluster products only from the 2-cluster solution. In some cases a 3-cluster solution may have produced higher quality results, but these cases were not investigated.*"

[L267-270 & Figure 3] The HAC algorithm may not necessarily output clusters in the same order that they were inputted as demonstrated in Figure 5. In Figure 3 for preparation strategy A for bacteria and diesel for the 80:20 ratio, is it possible to attain 80% misclassification for a two cluster solution? Perhaps I have misunderstood, but would this not mean that there were more diesel particles in the bacterial cluster and more bacterial particles in the diesel cluster, and hence a better classification error could be attained simply by swapping the labels on the clusters?

You are correct that the order of cluster numbering is unrelated to the order of particles input and so the source of individual particles must be known already, but it is not possible to improve the results by swapping labels in the way you suggest. We independently tracked the source of each particle assigned to each cluster so we can rigorously calculate which particles were incorrectly assigned. The numbering of the clusters is arbitrary and the naming was assigned simply as a function of which particle was assigned in the largest concentration.

[Figure 3 & Table 2] Could you extend the results presented in Figure 3 to include at least one biological versus biological matchup? I notice when considering matching ups which contained only biological material the classification error is much higher. I believe that by not standardising the data this would cause the fluorescence to dominate more in the analysis. In the case of attempting to discriminate between fluorescent and non-fluorescent particles, this may be advantageous. However, in the case of attempting to discriminate between two different types of biological particle, it may be advantageous to give the size and shape measurements more weight, and hence it would be better in these cases to standardise the data. In addition other instruments such as the WIBS-NEO will have fluorescence measurements over a much larger range and fluorescent measurements are recorded often above 10000. What would the implication then be when not standardising the data in this case?

This is another interesting idea, but it was beyond the scope of what we were able to accomplish in the relatively short time we had available for this project. We chose to focus on the ability to separate bio from non-bio particles. While we didn't explore all Scenarios (e.g. A-F) for biological particles, we chose to look at bio-bio separations using Scenario B (i.e. Tables 2 and 3).

Title: Evaluation of a Hierarchical Agglomerative Clustering Method Applied to WIBS Laboratory Data for Improved Discrimination of Biological Particles by Comparing Data Preparation Techniques

NICOLE SAVAGE[1#], J Alex Huffman[1]
[1] *University of Denver, Department of Chemistry and Biochemistry, Denver, USA*
[#] *Now at Aerosol Devices, Inc.*

Correspondence to: J. Alex Huffman (alex.huffman@du.edu)

Running Title: Evaluation of clustering applied to WIBS bioaerosol data

Keywords: Clustering, Thresholding, Ward's linkage, Bioaerosols, Fluorescence, Laboratory characterization

Abstract

Hierarchical agglomerative clustering (HAC) analysis has been successfully applied to several sets of ambient data (e.g. Crawford et al., 2015; Robinson et al., 2013) and with respect to standardized particles in the laboratory environment (Ruske et al., 2017; Ruske et al., 2018). Here we show for the first time a systematic application of HAC to a comprehensive set of laboratory data collected for many individual particle types using the ~~w~~Wideband ~~i~~Integrated ~~b~~Bioaerosol ~~s~~Sensor (WIBS-4A) (Savage et al., 2017). The impact of ~~particle~~ ratio of particle concentrations on HAC results was investigated, showing that clustering quality can vary dramatically as a function of ratio. Six strategies for particle pre-processing were also compared, concluding that using raw fluorescence intensity (without normalizing to particle size) and inputting all data in logarithmic bins consistently produced the highest quality results for the particle types analyzed. A total of 23 one-on-one matchups of individual particles types were investigated. Results showed cluster misclassification of <15% for 12 of 17 numerical~~analytical~~ experiments using one biological and one non-biological particle type each. Inputting fluorescence data using a baseline + 3σ threshold produced lower misclassification than when inputting either all particles (without fluorescence threshold) or a baseline + 9σ threshold. Lastly, six numerical simulations of~~synthetic~~ mixtures of four to seven components were analyzed using HAC. These results show that a range of 12-24% of fungal clusters were consistently misclassified by inclusion of a mixture of non-biological materials, whereas bacteria and diesel soot were each able to be separated with nearly 100% efficiency. The study gives significant support to the application of clustering analysis to data from commercial UV-LIF instruments being commonly used for bioaerosol research across the globe and provides practical tools that will improve clustering results within scientific studies as a part of diverse research disciplines.

1. Introduction

Particles of biological origin, or bioaerosols, make up a substantial fraction of atmospheric aerosol and have the potential to influence environmental processes and to negatively impact human health (Després et al., 2012; Douwes et al., 2003; Fröhlich-Nowoisky et al., 2016; Shiraiwa et al., 2017). In order to understand the impact bioaerosols, such as pollen, spores, and bacteria, play on various systems, it is important to be able to identify and characterize these biological particles in the atmosphere. One common method for the detection of bioaerosols is ultraviolet laser/light-induced fluorescence (UV-LIF), because it can provide particle detection in near real-time and at high particle size resolution (Fennelly et al., 2017; Huffman and Santarpia, 2017; Sodeau and O'Connor, 2016). Many commercial UV-LIF instruments have become available for bioaerosol detection, but all of these techniques are challenged with the need to differentiate between small differences in fluorescence properties in order to identify~~sort~~ and quantify biological aerosols from non-biological material. Recently commercialized instruments show improved ability to discriminate between particle types, for example by utilizing multiple excitation sources or other particle data (e.g. size and shape). UV-LIF techniques are inherently limited, however, by the broad nature of fluorescence spectra and so instruments face a ubiquitous problem of poor selectivity between particle types. By applying improved data thresholding and particle classification techniques, particle characterization can be further improved, but important limitations still remain (Hernandez et al., 2016; Huffman et al., 2012; Perring et al., 2015; Savage et al., 2017; Toprak and Schnaiter, 2013; Wright et al., 2014). One strategy to improving quality of differentiation between particles types has been to collect full, resolved emission spectra, each at multiple excitation wavelengths. This can lead~~s~~ to high instrumental purchase cost, and such instruments have not been widely applied or commercialized (Huffman et al., 2016; Kiselev et al., 2013; Pan et al., 2009b; Ruske et al., 2017; Swanson and Huffman, 2018). Most commercial UV-LIF instruments for bioaerosol detection utilize 1-2 excitation wavelengths and integrate fluorescence signals into a small number of emission bands. To extend the improvements in particle classification for these commercial UV-LIF instruments, a number of multivariate analysis techniques have been applied to ambient particle analysis. The most common of these techniques include principal component analysis, factor analysis, and cluster analysis strategies. Classification algorithms, including several c~~C~~lustering techniques~~,~~ in particular, have shown successful results in providing unbiased insights to the classification of bioaerosols (Crawford et al., 2015; Pinnick et al., 2013; Robinson et al., 2013; Swanson and Huffman, 2018).

Cluster analysis is a broad class of data mining methods in which data objects placed in the same group (or cluster) are more similar to one another than to those objects placed in other groups. Classification algorithms~~Clustering techniques~~ can be divided into two central models: (1) supervised and (2) unsupervised learning. Both models have associated advantages and disadvantages. Supervised learning methods allow the “training” of data and grouping to better reflect the data observations (Eick et al., 2004; Ruske et al., 2017; Ruske et al., 2018). This type of method enhances (trains) the classification~~clustering~~ algorithm in that the output ~~cluster classes~~groups are pre-determined rather than discovered, as is the case for unsupervised methods. Supervision requires the user to have appropriate starting conditions to put into the model, which are often difficult or impossible to determine. Supervised training methods are also much more time-efficient compared to unsupervised methods, which is important when analyzing ambient datasets where particle counts (individual objects) can be greater than 10^6 (Ruske et al., 2017). In contrast, unsupervised training methods present less bias and can adapt to unique situations,

because the resultant clusters are based on models that have not been previously trained. To access some of the advantages of supervised methods, however, it is important~~critical~~ to first apply unsupervised models to wide collections of laboratory data of known particle types in order to gain insight on how these models interpret data inputs and to learn how algorithms can best be trained (Ruske et al., 2017).

Hierarchical agglomerative clustering (HAC) is an unsupervised learning method that has been most commonly applied for bioaerosol related studies (e.g. Crawford et al., 2016; Crawford et al., 2015; Gosselin et al., 2016; Pan et al., 2009a; Pan et al., 2007; Pinnick et al., 2013; Pinnick et al., 2004; Robinson et al., 2013; Ruske et al., 2017; Ruske et al., 2018). Other unsupervised clustering techniques, such as the k-means clustering method, have shown poor results when applied to ambient data sets because the number of clusters used to represent the data are required a priori, and this information is usually unknown prior to analysis (Ruske et al., 2017). There are several different HAC methods or linkages including: Single, Complete, Average, Weighted, Ward's, Centroid, and Median (Crawford et al., 2015; Müllner, 2013). Ruske et al. (2017) compared a variety of HAC linkages and determined that Ward's linkage had a higher percentage of correctly classifying particles, in comparison to other HAC methods.

Recently, Savage et al. (2017) published a comprehensive laboratory study applying the ~~w~~Wideband ~~i~~Integrated ~~b~~Bioaerosol ~~s~~Sensor (WIBS-4A) to a large and diverse set of biological and non-biological aerosol types. Following on that work, the study presented here utilizes those data as inputs to evaluate and challenge the HAC strategy of particle differentiation using the Ward's linkage of unsupervised clustering. Previous HAC studies have focused primarily on (a) the analysis of simple particle standards (i.e. fluorescent microbeads) and (b) clustering of particles from ambient data sets. There have been relatively few published attempts to differentiate between biological particles and interfering particles by clustering methods using controlled laboratory UV-LIF data or to separate different kinds of biological particles from one another. Presented here are results of the HAC method applied to data from a comprehensive WIBS laboratory study showing that clustering can dramatically improve removal of non-biological particle types from data sets if operated under appropriate conditions.

2. Experimental and Computational~~ing~~ Methods

The WIBS-4A (Droplet Measurement Techniques, Longmont, CO) is a commonly used UV-LIF based instrument for the detection and characterization of biological particles. The instrument collects particles in the size range 0.8 – 20 µm and interrogates them in real-time as particles flow through the path between optical sources. The WIBS collects information about ~~3 channels of~~ fluorescence intensity ~~information~~ in three channels (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). The excitation and emission wavelengths chosen for each of the 3 fluorescence channels were designed to maximize the information gained about key biological fluorophores present in a broad range of bioparticles (Kaye et al., 2005; Pöhlker et al., 2012). Early generations of UV-LIF bioaerosol spectrometers were often interpreted to be able to detect proteins via channels similar to FL1 and products of active cellular metabolism (i.e. riboflavin and NAD(P)H) via channels similar to FL3, but these approximations are gross simplifications that confound more detailed investigation of particle types. For more information on the design, operation, and calibration of this instrument see e.g. the manuscripts listed here and references therein (Foot et al., 2008; Healy et al., 2012a; Healy et

al., 2012b; Hernandez et al., 2016; Kaye et al., 2005; Perring et al., 2015; Robinson et al., 2017; Savage et al., 2017; Stanley et al., 2011).

All aerosol materials utilized have been listed previously in Table 2 shown by Savage et al. (2017), where an overview of size and fluorescence properties of particles utilized for this study are also reported. No additional laboratory experiments were performed here beyond the results presented previously.

The fluorescence threshold applied to the differentiation of fluorescent from non-fluorescent particles is a key step in UV-LIF data analysis. Traditionally a fluorescence threshold has been determined as the average baseline fluorescence intensity measured in each of the three channels during the forced trigger (FT) mode when no particles are present, plus three times the standard deviation (σ) of that measurement (i.e. FT + 3σ) (Gabey et al., 2010). Savage et al. (2017) also reported that additional particle discrimination is possible by using FT + 9σ as the threshold. Both threshold definitions will be discussed here. After choosing a threshold of minimum fluorescence, the fluorescence characteristics of a particle can be classified into 7 different particle types introduced by Perring et al. (2015) and as summarized in Figure 1 shown by Savage et al. (2017).

3. Clustering Strategy

Hierarchical clustering methods work by grouping objects from the bottom up, meaning that each object (particle) starts as its own "cluster," and clusters are merged together based on similarities until a greatly reduced number of clusters are presented as a final solution. Ward's method for clustering is among the most popular approaches for HAC and is the only method based on a classical sum-of-squares criterion, minimizing the within-group sum of squares (or variance) (Müllner, 2013). The WIBS-4A used here for data collection provides 5 parameters of information for each individual particle detected (3 fluorescence channels, size and asymmetry factor:AF), resulting in 5 dimensions of data.

The clustering analysis was performed using the open-source software R package 'fastercluster' (Müllner, 2013) using a Dell Latitude E7450 laptop computer with an Intel® Core™ Processor (i7-5600U CPU @ 2.60 GHz, 16 GB RAM).

3.1 Data Preparation

Saturation of fluorescence intensity occurs at 2047 analog-to-digital counts (ADC) for each of the three FL channels in the WIBS-4A, at which point the photomultiplier tube (PMT) reaches its upper limit of detection. A study by Ruske et al. (2017) investigated whether non-fluorescent (in that case, particles below the FT + 3σ fluorescence threshold) and/or saturating data points included in the clustering analysis hindered the efficiency of the cluster output. The authors determined that removing both saturating and non-fluorescent particles before HAC analysis resulted in a better clustering performance in terms of correctly classifying ambient particles. ~~Their conclusions, however, were based on ambient field data using unknown particles types and did not investigate laboratory-generated particles of known origin.~~ The quality of the clustering results is~~are~~ likely to be impacted by types of particles involved and the assumptions placed on those. As shown by Savage et al. (2017), many biological particles present a large fraction that saturate one or more of the fluorescence detectors. Conversely, many non-biological particles present a large fraction of very weakly fluorescent particles with intensity below a given threshold and thus that are classified as non-fluorescent. To limit pre-modification of particle populations before clustering, the only filter applied before clustering was to remove particles

smaller than the lower particle size detection limit of the WIBS-4A (0.8 µm), similar to Ruske et al. (2017). In contrast, both saturating and non-fluorescent particles were analyzedretained and the clustering results will be evaluated. Figure 1 outlines the data preparation process, including the conceptual process of normalization, clustering, and validation of data, which iswill be explained in detail below.

3.2 Data Normalization

Normalization of the raw data is necessary before executing the clustering algorithm, because data parameters delivered from the instrument are measured on different respective scales. For example, fluorescent intensity values range from 0 to 2047 ADC (analog-to-digital counts), size from 0 to ~20 µm, and AF from 0 to 100 arbitrary units. Crawford et al. (2015) performed analysis on polystyrene latex spheres (PSLs) using several different normalization techniques, concluding that z-score normalization iwas the best technique when looking at cluster performance using Ward's linkage for the separation of PSLs. As a result, we utilize the z-score normalization of Ward's linkage HAC for the presented study. By this type of normalization, the mean value of all data points is subtracted from each individual data point, and then each data point is divided by the standard deviation of all points. Standardization using the z-score method compares results to a normal (Gaussian) population, and we have chosen to standardize our variables to a mean of 0 and a variance of 1 so that the output variables would be on comparable scales., and it therefore relies on the assumption that input data can be described by a normal distribution .

3.3 HAC Scenarios

Hierarchical agglomerative clustering performs optimally if all variables (1) are independent of one another and (2) can be described well by a normal (Gaussian) distribution (Norusis, 2011). To achieve meaningful results from the clustering analysis data values must, therefore, be input into the clustering algorithm with an careful understanding of how specific preparatory conditions can significantly impact results. To investigate optimal input conditions a total of 6 clustering scenarios were explored, with conditions summarized in Table 1. The impact of two separate variables were explored within these scenarios by varying: (i) whether fluorescence intensity were pre-normalized by particle size and (ii) whether the data values were input in after logarithmic transformationally spaced bins to produce a normal distribution.

Ambient particle number vs size distributions can often be are well approximated byknown to exhibit lognormal distributions, although specific groups of particles, including some bacteria, spores, and pollen, may not always exhibit lognormal distribution. Further, fluorescence intensity has been shown to scale with particle size (e.g. Hill et al., 2001; Sivaprakasam et al., 2011). Several previous studies attempted to utilize HAC for ambient lognormally-distributed particle size data (Crawford et al., 2014; Crawford et al., 2015; Robinson et al., 2013), but applied the assumption that particle fluorescence is normally distributed in a group of particles. If this assumption does not hold to be correct, however, weakly fluorescing particles are likely to be grouped into a single cluster based on the high abundance of these particles (Robinson et al., 2013). Scenarios C, D, and E (Table 1) utilize data input to the clustering algorithm after 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 explore whether the assumption that laboratory data should be treated like previously explored ambient data sets and not logged. Scenarios B and D take into account the logging of all parameters,

producing normal distributions of all variables (AF, particle size, 3 channels of fluorescence). By this process, data values were input into the algorithm as log(value) without separately binning the points. For comparison, scenarios E and F explore log-spaced distributions of size and AF, while retaining the assumption that the fluorescence output is normally distributed. Scenario A data is neither logged nor normalized. For comparison, Scenario F represents the input conditions that have been used frequently (e.g. Crawford et al., 2015; Ruske et al., 2017).

3.4 Cluster Validation

An important feature of HAC is that it provides clusters in an unsupervised manner, and the user must determine the number of clusters that makes physical sense. One useful tool to systematically determine the optimal number of final clusters is the Calinski-Harabasz (CH) index, which uses the interclass-intraclass distance ratio (Liu et al., 2010). For each clustering output the CH index was calculated for cluster solutions with one through ten clusters, and the solution with the highest CH value was generally determined to be the optimal number of clusters. Figure 2 shows an example CH versus cluster number plot for a mixture of *Aspergillus niger* fungal spores mixed with diesel soot particles. The curve suggests the optimal result to be a 2-cluster solution for this trial, as was generally the case for investigations where two particle types were mixed before clustering. In order to reduce the length and complexity of discussion~~analysis~~, analysis of results in Sections 4.1-4.3 was limited to using cluster products only from the 2-cluster solution~~all cases presented~~. In some cases a 3-cluster solution may have produced higher quality results, but these cases were not investigated. ~~in Sections 4.1-4.3 are products of a 2-cluster solution.~~

4 Results and Discussion

The analysis of clustering quality was performed systematically and with increasing complexity. Section 4.1 utilizes three pairs of particles types to explore the effect of particle ratio and normalization strategies on cluster performance. Using conclusions from this section, Section 4.2 then expands the exploration to 20 additional pairs of particle types. Section 4.3 explores the effect of three different fluorescence thresholding strategies on cluster output. Finally, Section 4.4 investigates the ability of HAC analysis to separate particle types from mixed populations of particle types.

4.1 Investigating pre-normalization scenarios and particle input ratio

To explore the ability to separate two distinct populations of particles from one another, three different clustering trials are presented in this section as one-on-one match-ups: (1) *Aspergillus niger* (fungal spores, F2) vs. NIST diesel soot (S4), (2) *Pseudomonas stutzeri* (bacteria, B3) vs. NIST diesel soot (S4), and (3) *Aspergillus niger* (fungal spores, F2) vs. California sand (mineral dust, D12). These four particle materials were chosen to represent key classes of coarse particles observed in ambient air. 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~~. The clustering process includes: (i) evaluation of cluster performance based on particle assignment and cluster composition, and (ii) visual representations of cluster outputs using particle type classification introduced by Perring et al. (2015). For each of these three trials, the clustering process was run separately using each of the six scenarios A-F described in Table 1. Additionally, while exploring the optimal data pre-processing scenario, the influence that different concentration ratios of particle types could play in the clustering output

was also explored. The cluster process for each trial was performed using ~~three~~ four different ratios of particles in each particle set including situations with an equal ratio ~~(50:50)~~ and ~~situations~~ where the concentration of each particle type was significantly mismatched ~~(80:20 and 20:80)~~. In total, this section represents 5~~7~~4 individual clustering experiments (3 trials x 6 scenarios x 3 particle ratios + 3 additional ratio trials) exploring three independent input variables. The results will be utilized to explore many more individual particle type match-ups in the following sections.

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), and because they can have fluorescence characteristics difficult to distinguish from small biological particles (e.g. Huffman et al., 2010; Pan et al., 2012; Savage et al., 2017; Yu et al., 2016). For example, when excited by photons with a wavelength of 280 nm, diesel soot can be misinterpreted as single bacterial cells using the WIBS, and so we explored here whether the two particle types could be clustered separately (Pöhlker et al., 2012). The three trials include two examples of biological particles, both exhibiting fluorescent properties, but with different excitation-emission characteristics and with different average particle size.

The output of the algorithm reports the particle type from which each particle was input in order to evaluate the accuracy of the clustering. The resulting output of each particle with an assigned cluster number is then compared to the originating particle type to determine classification accuracy. Figure 3 summarizes the relative accuracy of individual clustering experiments by representing the percent of particles misclassified with respect to known input identities (blue bar corresponding to correct classification, red bar and overlaid value corresponding to incorrect classification). The clustering process was generally effective for separating particles correctly when two particle types were considered, but results vary widely across the six scenarios. Several previous studies that used HAC to separate particles within an ambient data set assumed that particle fluorescence is already normally distributed (Crawford et al., 2014; Crawford et al., 2015; Robinson et al., 2013). As a result, these previous studies did not normalize fluorescence data and thus used data preparation scenario F in their clustering analysis. For comparison, scenarios B and D were explored to test whether the clustering efficiency would be improved or hindered by fluorescence normalization. Scenarios A and F produced inconsistent results, with some experiments (i.e. 50:50 ratio of fungal spores:diesel) producing misclassification <1.1%, whereas other experiments (i.e. 20:80 ratio of bacterial:diesel) producing misclassification up to~~></~~ 80%. In contrast, scenarios B and D produced consistently more accurate results. Scenario B, in particular, consistently exhibited the most accurate classification of particles for almost every individual experiment. No experiment involving scenario B produced greater than 9% misclassification of particles, regardless of particle input ratio, and most experiments produced results with 0.1 - 3% error. These observations taken together suggest that particle fluorescence properties may not be well described by normal distributions and that normalizing fluorescence data prior to analysis may be more effective.

The results of these experiments also highlight how important the ratio of input particles can be. While scenario B was relatively consistent, varying only between 0.1 and 3.8% error for different ratios of the fungal spore versus diesel match-up, other experiments depended strongly on particle ratio. It is clear that the input ratio of particle types cannot be controlled during an ambient study, and so these results suggest that it is important to keep the possibility of varying

concentration ratios in mind when interpreting time- or air mass-associated changes in cluster composition or when relaying the relative confidence in clustering results. For the remainder of the discussion, experiments will be limited to a 50:50 ratio following scenario B. In each case the ~~number of~~ input particles are~~represents~~ a random subset taken from the pool of particles in the experimental data. As a result, individual samples selected from the same experiments (i.e. Fig. 4a, Fig 4e) can show slightly different average properties. In some cases (i.e. ~~D~~diesel soot, Fig. 4d) the number of particles originally analyzed was small and so to keep the input particle ratio 50:50 the corresponding particle type was also limited to small numbers.

To extend the investigation of particle input ratio, the three match-ups presented in Figure 3 were investigated using Scenario B with 1% bioparticles and 99% non-bioparticles in each respective case. In these experiments the bacteria:diesel soot and fungal spores:dust particles separated relatively well (6.6% and 13.5% misclassification, respectively). The fungal spores:diesel soot separation was poor, however, because the diesel soot particles were nearly evenly split into both clusters, and the fungal spore particles were too low in concentration to influence the cluster properties. More investigation is needed to explore how extreme disparities in particle ratio could negatively influence cluster quality in real-world settings.

An important tool readily applied to analysis of ambient data is the categorization of particles into 8 fluorescent particle types (Perring et al., 2015). Thus, to further investigate the quality of cluster accuracy, Figure 4 shows inputs and cluster outputs from three clustering experiments stacked as a function of fluorescence particle type and particle size. The top row of Figure 4 shows the input data for *Aspergillus niger* and diesel soot (Fig. 4a-b) paired with the outputs of the 2-cluster solution (Fig. 4g-h). It can be seen that both particle materials have predominantly particle type-A characteristics, meaning that they are fluorescent only in channel FL1. The fungal material also presents roughly a third AB (green) and a small minority of non-fluorescent (gray) characteristics. The size distribution of the fungal spores peaks at ~3 µm, whereas diesel soot peaks at ~1 µm in size. While not shown in this plot style, the spores exhibit moderately higher FL1 channel fluorescence, with a median of 543 ADC, whereas diesel soot exhibits a median of 751 ADC in this channel (see Savage et al., 2017; Table 2). Both particle types show almost no fluorescent characteristics in either FL2 or FL3. In summary, the particle distributions are relatively similar in fluorescence particle type and their differences are largely related to particle size, so separation of these particles through Trial 1 was hypothesized to represent a relatively challenging initial exercise. The clustering outputs presented in Figures 4g-h, however, visually highlight the conclusion represented by Figure 3, which is that the particles in this trial separated very well. Cluster 1 was comprised predominantly of fungal particles and presented fluorescence and size traits qualitatively similar to the input fungal particles, whereas cluster 2 was comprised predominantly of diesel soot particles. Results from the 50:50 ratio of the scenario B experiments for the other two trials are also shown in the last two rows of Figure 4. In each case, the qualitative properties of the input particles are extremely well represented by the corresponding output cluster, corroborating the conclusion from Figure 3 that the scenario B cases accurately separated the particle groups investigated through these experiments. It is also important to note here that the method of aerosolization for each particle type plays an important role in the observed size distribution and so results involving laboratory particles should be interpreted with this in mind. Observed fluorescence properties, in contrast, are expected to be conserved at a given particle size and intrinsically related to particle composition.

4.2 Investigating cluster quality without fluorescence threshold

After concluding that scenario B exhibited the most consistently accurate clustering results using 2-cluster solutions from mixtures comprised of 2 particle type inputs, the analysis was expanded to include a broader range of particle types. Using 50:50 ratios of two types of input particles, prepared using scenario B (leaving fluorescence data un-normalized and forcing all five data parameters into logarithmically spaced bins), 20 new individual experiments were performed. The results of all 23 experiments (3 from Section 4.1 and 20 introduced in Section 4.2) are summarized in Table 2 as the percentage of particle misclassification. These trials were chosen to represent a broad range of individual match-ups that might be expected in ambient air. From the original 69 types of particles analyzed by Savage et al. (2017), 14 were used in experiments here: 8 types of non-biological particles and 6 types of biological particles (2 each of fungal spores, bacteria, and pollen species). Supplemental Figure S4 from Savage et al. (2017) shows size distributions stacked by fluorescence particle type for each of the particle species discussed.

Table 2a organizes clustering results into three rows, showing misclassification of F2 (*Aspergillus niger* fungal spore), B3 (*Pseudomonas stutzeri* bacteria), and P9 (*Phelum pratense* pollen) particles, respectively, with respect to a variety of other particle types represented by table column. Of the 15 cluster experiments between fungal spore or bacteria and non-biological material (top two table rows), only 3 showed misclassification greater than 7.5% (bold text), and 7 were less than 3%. The three outliers were: experiment (7) F2 vs BC3 (glyoxal + ammonium sulfate brown carbon aerosol), (8) F2 vs WT (white t-shirt particles), and (14) B3 vs WT. Looking first at experiment (7), F2 particles show A-type fluorescence characteristics and are dominated by a mode between 1.5 and 4 µm. BC3 particles are primarily non-fluorescent <1.5 µm, but are primarily A-type between 1.5 and 3 µm, suggesting similar size and fluorescence properties. The white t-shirt particles separated poorly (~41% misclassification) from both the fungal spore and bacterial particles. All three particle types (WT, F2, and B3) exhibit medium fluorescent intensity in the FL1 channel. The poor ability to separate WT from both F2 and B3 was surprising, however, given that WT exhibited significantly higher mean fluorescence in each of the FL2 and FL3 channels. As first mentioned by Savage et al. (2017), great care should be taken when interpreting fluorescent particle results from indoor environments where increased concentrations of bleached fibers from clothing, bedding, paper, and cleaning products may be present.

While the results show that the spores and bacterial particles investigated could generally be well separated from most potentially interfering non-biological species, the results were much less successful for differentiation from pollen. P9 pollen particles separated poorly in all experiments (versus D12, H2, or P5), with rate of misclassification ranging from 22 to 47%. It is important to keep in mind, however, that the WIBS was operated using a standard gain setting that limits analysis of particle size to below approximately 20 µm. As a result, the WIBS is insensitive to whole pollen grains and so most of the particles observed during pollen experiments are small pollen fragments. Any intact pollen grains that navigate the flow system to be detected are likely to be binned together in the channel representing the largest particles. Clustering results including pollen should be interpreted accordingly. Pollen grains can fragment in ambient air as function of increased relative humidity (Miguel et al., 2006; Suphioglu et al., 1992; Taylor et al., 2004), but the relative ratio of whole/fragmented particles is hard to predict under ambient conditions. Smaller fragments can also exhibit different fluorescent properties than whole grains (Pöhlker et al., 2013). O'Connor et al. (2014) operated a WIBS-4 (Univ.

Hertfordshire) at lower gain in order to improve pollen detection efficiency, but these results are not explored directly here.

The WIBS instrument is frequently used to differentiate between airborne biological particles and material of non-biological origin. A secondary goal of differentiating more finely between types of biological aerosols is also frequently pursued. To investigate this goal, six additional experiments were conducted by pairing two different types of non-biological particles (Table 2b). In contrast to the results shown in Table 2a, the clustering algorithm showed generally poor ability to separate between two biological particle types. Only one of the six experiments resulted in error <15% (F2 vs B3, 10.3% error), whereas error for the other five experiments ranged from 18% to 65%. The worst accuracy was demonstrated by experiments (22) B1 vs B3 and experiment (23) P5 vs P9. Both of these experiments attempted to separate between different species of a single particle type (i.e. between two bacteria or two pollen, respectively). Overall, these results suggest that the clustering strategy may be quite useful at aiding the differentiation of biological material from non-biological material, but that separating more finely to quantify differences between types of individual biological particles is ~~likely to be~~ significantly more challenging and not likely to be possible in most situations.

4.3 Investigating impact of fluorescence thresholding strategy on cluster quality

In previously published studies, removing particles from clustering analysis that exhibited particle fluorescence intensity below the threshold (i.e. non-fluorescent) or at the saturating point improved the efficiency of clustering (Crawford et al., 2015; Ruske et al., 2017). In Sections 4.1-4.2, particles with either of these characteristics were left in the analysis to prevent the underestimation of particles clustered. In this section, however, we investigated whether removing non-fluorescent particles could improve cluster accuracy for the experiments that performed poorly in Section 4.2. Of the 23 trials represented in Table 2, 10 experiments exhibited 15% or greater misclassification and were subjected to further analysis in order to investigate whether using a more discriminating fluorescence thresholding strategy could improve cluster results. In all 10 cases fluorescence saturating particles were retained, and three separate thresholding conditions were compared by: (I) keeping all non-fluorescent and saturating particles, (II) removing non-fluorescent particles by applying a fluorescence threshold of FT baseline + 3σ, and (III) and removing non-fluorescent particles by applying a fluorescence threshold of FT baseline + 9σ. Savage et al. (2017) showed evidence that applying a FT + 9σ improved WIBS results by removing a higher fraction of non-biological material from analysis than by applying the more commonly used FT + 3σ, without negatively impacting observations of biological particles. Table 3 shows the percentage of particles misclassified in each of three scenarios investigated here (Table 3a) as well as the number of particles subjected to the clustering algorithm (Table 3b).

Each scenario, with exception of the B3 vs B9 experiment (21), shows a decrease in particle misclassification from scenario I (no fluorescence threshold applied) to scenario II (FT + 3σ). In contrast, eight of the ten scenarios *increase* in particle misclassification when raising the fluorescence threshold from 3σ (II) to 9σ (III). The exceptions to this trend are experiments (8) F2 vs WT and (19) F2 vs P9, which show nominal improvement in error (2-4% reduction) with increased threshold. We hypothesize that the 9σ results degrade, in most cases, because the threshold becomes high enough that most weakly fluorescing particles have been removed from analysis. This reduces the ability of the cluster to group into low and high fluorescence categories, and so remaining particles are separated less efficiently. Secondly, removing particles

at higher fluorescence thresholds leads to increasingly poor counting statistics, as represented in Table 3b by the number of particles included in each experiment. Overall, these results suggest that inputting particles into the clustering analysis with at least a nominal fluorescence threshold (i.e. FT + 3σ) can improve the clustering results in many cases, however, increasing the threshold further may decrease cluster quality.

4.4 Investigating the capability ~~cluster ability~~ to separate particles in simulations of complex ~~synthetic~~ mixtures

To this point, our investigation has focused on a variety of individual match-ups between two distinct particle types. To better simulate real-world scenarios, we computationally~~analytically~~ simulated~~synthesized~~ six mixtures of particles by pooling existing WIBS data from selected particle types in prescribed ratios. Each simulated mixture was assembled~~synthesized~~ 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. Table 4 provides an overview of the percentage of each particle type included as well as the total number of particles in the mixture. Mixtures 1 and 2 were simulated~~synthesized~~ arbitrarily to test if a minority (25%) of one type of fungal spores (F2) could be separated from a majority (75%) of a mixture of three different non-biological materials. Mixtures 3 and 4 synthesized arbitrary mixtures of two types of bioaerosol (F2 and B3) with three or five types of non-biological particles, respectively. Mixture 5 was simulated~~synthesized~~ to examine the separation of pollen (P9) from a set of five non-biological particles. Mixture 6 was simulated~~synthesized~~ to ~~simulate~~ be similar to an indoor environment that might have a mixture of biological particles (F2 and B3) with non-biological materials, including bleached fibers (WT). These mixtures are not intended to closely mimic any set of individual ambient conditions, but are rather used as very rough ~~synthetic scenarios~~simulations used for discussion and to prompt discussion related to future experiments within the community. In a real-world sampling environment one would also expect a high concentration of non-fluorescent particles as well (e.g. most organic aerosols, sea salt, dusts), but these were generally~~largely~~ not sampled as a part of the Savage et al. (2017) study, which focused on fluorescent particles. As a result, relatively non-fluorescent particles like D12 and H2 were included here as “fillers” in most mixtures as surrogates for other types of non-fluorescent particles. Clustering analysis was performed using the ratios listed in Table 4, the B scenario of pre-normalization conditions, and filtering non-fluorescent particles below the FT + 3σ threshold. In all cases, the number of clusters retrieved after HAC was pre-defined to be the same as the number of particle types input.

Cluster results from all six mixtures are summarized in Figure 5. Figure 5 (Part A) shows the number of particles from each type assigned to each cluster, and Parts B and C show results grouped by general particle classification (brown for non-biological and dark green for biological). Overall, the ability of the HAC analysis to separate the biological particles from the non-biological particles was high. In some cases, the quality of separation of one or two biological species from a mixture of non-biological materials was even higher than the 2-material match-ups shown in Sections 4.1-4.3. The two 4-component mixtures showed 22.4% and 14.8% misclassification of fungal spores. In both cases, a small fraction of each of the non-biological materials were mixed into the spore cluster, whereas almost none (1.5% and 0.6%) of the spores were incorrectly mixed into the sum of the non-biological clusters.

Mixtures 3 and 4 showed similar misclassification for fungal spores (11.9% and 13.8%, respectively), whereas the bacterial particles clustered with amazing quality. For Mixture 3, no

particles other than bacterial particles were grouped into Cluster 1, and only 16 of 213 bacterial particles were assigned to other clusters. For Mixture 4, 135 of 137 particles in Cluster 6 were bacterial in origin and 135 of 142 bacterial particles were assigned to the cluster. The combination of fungal and bacterial particles in Mixtures 3 and 4 resulted in a total of 5.0% and 5.3% misclassification of all biological particles.

In contrast to the poor separation of pollen from other particle types discussed in Section 4.2, Mixture 5 showed a higher quality of separation between pollen (9.4% misclassified) and the sum of five other non-biological particle types. Lastly, the mixture designed to roughly mimic an indoor environment including white t-shirt particles. In this mixture the WT particles confounded the spore separation, but the bacterial separation was nearly flawless.

Another surprising observation from the analysis of these simulated~~synthetic~~ mixtures was that the diesel soot particles (Mixtures 1, 2, 4, and 5) separated into their own cluster in almost all cases with very high quality (1.8%, 2.9%, 0.6%, and 9.4%, respectively, of diesel soot particles misclassified into a different cluster). The quality of separation of bacterial particles and diesel soot (Mixture 4) was especially amazing, given the qualitative similarity of the two particle populations. For example, size-distributions of each particle type show primarily A-type particles with similar mean fluorescent intensity values in FL1, FL2, and FL3 (Savage et al., 2017).

5. Conclusions

Application of results from a recent set of systematic laboratory experiments (Savage et al., 2017) by the commonly used hierarchical agglomerative clustering analysis helps to reveal areas where the tool can be used well and other areas where it struggles. First (Section 4.1) it was observed that differing ratios of particle input into the clustering algorithm can produce dramatically different results. It will be important for anyone applying HAC to ambient particle sets where particle ratios are not independently verified to interpret results somewhat loosely. In Section 4.1 the clustering quality of scenario B, where fluorescence intensity was not normalized to particle size and where all input variables were binned into log space, was determined to consistently demonstrate the highest quality results. Further, the ability to the HAC analysis to separate between two groups of individual particle types using no fluorescence threshold (Section 4.2) and comparing three separate threshold strategies (Section 4.3) was shown to be relatively high in many cases, but confounded in others. Lastly, Section 4.4 explored the ability of HAC analysis to separate biological components from more complex mixtures of four to seven types of input particles.

A standard fluorescence threshold of FT + 3σ has been commonly applied during WIBS analysis to separate between fluorescent and non-fluorescent particles. Savage et al. (2017) concluded that application of a more aggressive threshold strategy (FT + 9σ) could help discriminate between biological and non-biological particles more successfully in many circumstances, however certain types of interfering, non-biological particle species can still confound WIBS analysis irrespective of the threshold. Here we have investigated an orthogonal strategy to separate particle types by subjecting particles to HAC computer analysis. By comparing the results of the HAC analysis with raw separation based on fluorescence thresholding alone, the HAC analysis can clearly increase quality of differentiation. Interestingly, while Savage et al. (2017) reported that the FT + 9σ strategy helped improved differentiation, using the same threshold in conjunction with HAC analysis actually degraded results. We therefore conclude that if HAC analysis is to be performed, the standard FT + 3σ threshold is

likely to produce the highest quality results, however if HAC is not to be applied that the FT + 9σ threshold is probably a better choicethe most likely to enable investigation of biological particles while computationally filteringreduce a large fraction of non-biological particles.

The overall message here is that HAC can be applied successfully to differentiate particle types sampled by WIBS instruments and that it is most successful at separating biological species (i.e. fungal spores and bacteria) from non-biological particles. In all cases the HAC method allows separation of particles at least at the order-of-magnitude level, and often with misclassification of <5%. As mentioned by Savage et al. (2017), however, it should always been kept in mind that different instruments may produce slightly different signals due to physical differences between instruments (i.e. fluorescence calibration, tuning, and detector gain sensitivity) and between calibration strategies (Könemann et al., 2018; Robinson et al., 2017). and that rResults here are only also generally extendable to other UV-LIF instruments, whether they offer single or many channels of emission spectral resolution, in that the methods of particle pre-preparation and the impact of particle number ratio are likely to relay similar effects on clustering strategy. -Subtle differences in particles observed in a real-world environment may also complicate HAC analysis or the extension of results presented here. The UV-LIF community is encouraged to continue laboratory investigations, including detailed interrogation of clustering analytical techniques, to further understand limitations to better differentiating between particles.

6. Acknowledgments

The authors acknowledge the University of Denver for financial support from the faculty start-up fund. Nicole Savage acknowledges financial support from the Phillipson Graduate Fellowship at the University of Denver. Martin Gallagher, and David Topping, and Simon Ruske in the School of Earth and Environmental Sciences at the University of Manchester are acknowledged for initial discussion regarding clustering strategy. Cathy Durso at the University of Denver Center for Statistics and Visualization is acknowledged for help running clustering algorithms. All contributors to the Savage et al. (2017) paper, in which all experimental data discussed here were originally presented, are acknowledged for their contributions.

7. References

Bond, T. C., Doherty, S. J., Fahey, D. W., Forster, P. M., Berntsen, T., DeAngelo, B. J., Flanner, M. G., Ghan, S., Karcher, B., Koch, D., Kinne, S., Kondo, Y., Quinn, P. K., Sarofim, M. C., Schultz, M. G., Schulz, M., Venkataraman, C., Zhang, H., Zhang, S., Bellouin, N., Guttikunda, S. K., Hopke, P. K., Jacobson, M. Z., Kaiser, J. W., Klimont, Z., Lohmann, U., Schwarz, J. P., Shindell, D., Storelvmo, T., Warren, S. G., and Zender, C. S.: Bounding the role of black carbon in the climate system: A scientific assessment, J. Geophys. Res.-Atmos., 118, 5380-5552, 2013.

Crawford, I., Lloyd, G., Herrmann, E., Hoyle, C. R., Bower, K. N., Connolly, P. J., Flynn, M. J., Kaye, P. H., Choularton, T. W., and Gallagher, M. W.: Observations of fluorescent aerosol-cloud interactions in the free troposphere at the High-Altitude Research Station Jungfraujoch, Atmospheric Chemistry and Physics, 16, 2273-2284, 2016.

Crawford, I., Robinson, N. H., Flynn, M. J., Foot, V. E., Gallagher, M. W., Huffman, J. A., Stanley, W. R., and Kaye, P. H.: Characterisation of bioaerosol emissions from a Colorado pine forest: results from the BEACHON-RoMBAS experiment, Atmos. Chem. Phys., 14, 8559-8578, 2014.

Crawford, I., Ruske, S., Topping, D. O., and Gallagher, M. W.: Evaluation of hierarchical agglomerative cluster analysis methods for discrimination of primary biological aerosol, Atmos. Meas. Tech., 8, 4979-4991, 2015.

Després, V. R., Huffman, J. A., Burrows, S. M., Hoose, C., Safatov, A. S., Buryak, G. A., Fröhlich-Nowoisky, J., Elbert, W., Andreae, M. O., Pöschl, U., and Jaenicke, R.: Primary Biological Aerosol Particles in the Atmosphere: A Review, Tellus Series B-Chemical and Physical Meteorology, 64, 15598, 2012.

Douwes, J., Thorne, P., Pearce, N., and Heederik, D.: Bioaerosol health effects and exposure assessment: Progress and prospects, Annals of Occupational Hygiene, 47, 187-200, 2003.

Eick, C. F., Zeidat, N., and Zhao, Z.: Supervised clustering-algorithms and benefits, 2004, 774-776.

Fennelly, M. J., Sewell, G., Prentice, M. B., O'Connor, D. J., and Sodeau, J. R.: The Use of Real-Time Fluorescence Instrumentation to Monitor Ambient Primary Biological Aerosol Particles (PBAP), Atmosphere, 9, 1, 2017.

Foot, V. E., Kaye, P. H., Stanley, W. R., Barrington, S. J., Gallagher, M., and Gabey, A.: Low-cost real-time multi-parameter bio-aerosol sensors, Proceedings of the SPIE - The International Society for Optical Engineering, 7116, 711601, 2008.

Fröhlich-Nowoisky, J., Kampf, C. J., Weber, B., Huffman, J. A., Pöhlker, C., Andreae, M. O., Lang-Yona, N., Burrows, S. M., Gunthe, S. S., Elbert, W., Su, H., Hoor, P., Thines, E., Hoffmann, T., Després, V. R., and Pöschl, U.: Bioaerosols in the Earth system: Climate, health, and ecosystem interactions, Atmospheric Research, 182, 346-376, 2016.

611 Gabey, A. M., Gallagher, M. W., Whitehead, J., Dorsey, J. R., Kaye, P. H., and Stanley, W. R.:
612 Measurements and comparison of primary biological aerosol above and below a tropical forest
613 canopy using a dual channel fluorescence spectrometer, Atmospheric Chemistry and Physics, 10,
614 4453-4466, 2010.

615 Gosselin, M. I., Rathnayake, C. M., Crawford, I., Pohlker, C., Frohlich-Nowoisky, J., Schmer,
616 B., Despres, V. R., Engling, G., Gallagher, M., Stone, E., Poschl, U., and Huffman, J. A.:
617 Fluorescent bioaerosol particle, molecular tracer, and fungal spore concentrations during dry and
618 rainy periods in a semi-arid forest, Atmospheric Chemistry and Physics, 16, 15165-15184, 2016.

619 Healy, D. A., O'Connor, D. J., Burke, A. M., and Sodeau, J. R.: A laboratory assessment of the
620 Waveband Integrated Bioaerosol Sensor (WIBS-4) using individual samples of pollen and fungal
621 spore material, Atmospheric Environment, 60, 534-543, 2012a.

622 Healy, D. A., O'Connor, D. J., and Sodeau, J. R.: Measurement of the particle counting
623 efficiency of the "Waveband Integrated Bioaerosol Sensor" model number 4 (WIBS-4), Journal
624 of Aerosol Science, 47, 94-99, 2012b.

625 Hernandez, M., Perring, A. E., McCabe, K., Kok, G., Granger, G., and Baumgardner, D.:
626 Chamber catalogues of optical and fluorescent signatures distinguish bioaerosol classes,
627 Atmospheric Measurement Techniques, 9, 3283-3292, 2016.

628 Hill, S. C., Pinnick, R. G., Niles, S., Fell, N. F., Pan, Y. L., Bottiger, J., Bronk, B. V., Holler, S.,
629 and Chang, R. K.: Fluorescence from airborne microparticles: dependence on size, concentration
630 of fluorophores, and illumination intensity, Applied Optics, 40, 3005-3013, 2001.

631 Huffman, D. R., Swanson, B. E., and Huffman, J. A.: A wavelength-dispersive instrument for
632 characterizing fluorescence and scattering spectra of individual aerosol particles on a substrate,
633 Atmos. Meas. Tech., 9, 3987-3998, 2016.

634 Huffman, J. A. and Santarpia, J.: Online Techniques for Quantification and Characterization of
635 Biological Aerosols. In: Microbiology of Aerosols, John Wiley & Sons, Inc., 2017.

636 Huffman, J. A., Sinha, B., Garland, R. M., Snee-Pollmann, A., Gunthe, S. S., Artaxo, P., Martin,
637 S. T., Andreae, M. O., and Poeschl, U.: Size distributions and temporal variations of biological
638 aerosol particles in the Amazon rainforest characterized by microscopy and real-time UV-APS
639 fluorescence techniques during AMAZE-08, Atmospheric Chemistry and Physics, 12, 11997-
640 12019, 2012.

641 Huffman, J. A., Treutlein, B., and Pöschl, U.: Fluorescent biological aerosol particle
642 concentrations and size distributions measured with an Ultraviolet Aerodynamic Particle Sizer
643 (UV-APS) in Central Europe, Atmospheric Chemistry and Physics, 10, 3215-3233, 2010.

644 Kaye, P. H., Stanley, W. R., Hirst, E., Foot, E. V., Baxter, K. L., and Barrington, S. J.: Single
645 particle multichannel bio-aerosol fluorescence sensor, Optics Express, 13, 3583-3593, 2005.

646 Kiselev, D., Bonacina, L., and Wolf, J.-P.: A flash-lamp based device for fluorescence detection
647 and identification of individual pollen grains, Review of Scientific Instruments, 84, 2013.

Könemann, T., Savage, N. J., Huffman, J. A., and Pöhlker, C.: Characterization of steady-state fluorescence properties of polystyrene latex spheres using off- and online spectroscopic methods, Atmospheric Measurement Techniques, 11, 3987-4003, 2018.

Liu, Y., Li, Z., Xiong, H., Gao, X., and Wu, J.: Understanding of internal clustering validation measures, 2010, 911-916.

Miguel, A. G., Taylor, P. E., House, J., Glovsky, M. M., and Flagan, R. C.: Meteorological influences on respirable fragment release from Chinese elm pollen, Aerosol Sci. Technol., 40, 690-696, 2006.

Müllner, D.: fastcluster: Fast hierarchical, agglomerative clustering routines for R and Python, Journal of Statistical Software, 53, 1-18, 2013.

Norusis, M.: Cluster Analysis. In: IBM SPSS Statistics 19 Guide to Data Analysis, Norusis & SPSS Inc., 2011.

O'Connor, D. J., Healy, D. A., Hellebust, S., Buters, J. T. M., and Sodeau, J. R.: Using the WIBS-4 (Waveband Integrated Bioaerosol Sensor) Technique for the On-Line Detection of Pollen Grains, Aerosol Sci. Technol., 48, 341-349, 2014.

Pan, Y.-L., Pinnick, R. G., Hill, S. C., and Chang, R. K.: Particle-Fluorescence Spectrometer for Real-Time Single-Particle Measurements of Atmospheric Organic Carbon and Biological Aerosol, Environ. Sci. Technol., 43, 429-434, 2009a.

Pan, Y. L., Huang, H., and Chang, R. K.: Clustered and integrated fluorescence spectra from single atmospheric aerosol particles excited by a 263-and 351-nm laser at New Haven, CT, and Adelphi, MD, Journal of Quantitative Spectroscopy & Radiative Transfer, 113, 2213-2221, 2012.

Pan, Y. L., Pinnick, R. G., Hill, S. C., and Chang, R. K.: Particle-Fluorescence Spectrometer for Real-Time Single-Particle Measurements of Atmospheric Organic Carbon and Biological Aerosol, Environ. Sci. Technol., 43, 429-434, 2009b.

Pan, Y. L., Pinnick, R. G., Hill, S. C., Rosen, J. M., and Chang, R. K.: Single-particle laser-induced-fluorescence spectra of biological and other organic-carbon aerosols in the atmosphere: Measurements at New Haven, Connecticut, and Las Cruces, New Mexico, J. Geophys. Res.-Atmos., 112, D24S19, 2007.

Perring, A. E., Schwarz, J. P., Baumgardner, D., Hernandez, M. T., Spracklen, D. V., Heald, C. L., Gao, R. S., Kok, G., McMeeking, G. R., McQuaid, J. B., and Fahey, D. W.: Airborne observations of regional variation in fluorescent aerosol across the United States, J. Geophys. Res.-Atmos., 120, 1153-1170, 2015.

Pinnick, R. G., Fernandez, E., Rosen, J. M., Hill, S. C., Wang, Y., and Pan, Y. L.: Fluorescence spectra and elastic scattering characteristics of atmospheric aerosol in Las Cruces, New Mexico, USA: Variability of concentrations and possible constituents and sources of particles in various spectral clusters, Atmospheric Environment, 65, 195-204, 2013.

685 Pinnick, R. G., Hill, S. C., Pan, Y. L., and Chang, R. K.: Fluorescence spectra of atmospheric
686 aerosol at Adelphi, Maryland, USA: measurement and classification of single particles
687 containing organic carbon, Atmospheric Environment, 38, 1657-1672, 2004.

688 Pöhlker, C., Huffman, J. A., Förster, J.-D., and Pöschl, U.: Autofluorescence of atmospheric
689 bioaerosols: spectral fingerprints and taxonomic trends of pollen, Atmospheric Measurement
690 Techniques, 13, 3369-3392, 2013.

691 Pöhlker, C., Huffman, J. A., and Pöschl, U.: Autofluorescence of atmospheric bioaerosols -
692 fluorescent biomolecules and potential interferences, Atmospheric Measurement Techniques, 5,
693 37-71, 2012.

694 Robinson, E. S., Gao, R.-S., Schwarz, J. P., Fahey, D. W., and Perring, A. E.: Fluorescence
695 calibration method for single-particle aerosol fluorescence instruments, Atmospheric
696 Measurement Techniques, 10, 1755, 2017.

697 Robinson, N. H., Allan, J. D., Huffman, J. A., Kaye, P. H., Foot, V. E., and Gallagher, M.:
698 Cluster analysis of WIBS single-particle bioaerosol data, Atmospheric Measurement Techniques,
699 6, 337-347, 2013.

700 Ruske, S., Topping, D. O., Foot, V. E., Kaye, P. H., Stanley, W. R., Crawford, I., Morse, A. P.,
701 and Gallagher, M. W.: Evaluation of machine learning algorithms for classification of primary
702 biological aerosol using a new UV-LIF spectrometer, Atmospheric Measurement Techniques,
703 10, 695, 2017.

704 Ruske, S., Topping, D. O., Foot, V. E., Morse, A. P., and Gallagher, M. W.: Machine learning
705 for improved data analysis of biological aerosol using the WIBS, Atmospheric Measurement
706 Techniques Discussions, In Review, 2018.

707 Savage, N. J., Krentz, C. E., Könemann, T., Han, T. T., Mainelis, G., Pöhlker, C., and Huffman,
708 J. A.: Systematic characterization and fluorescence threshold strategies for the wideband
709 integrated bioaerosol sensor (WIBS) using size-resolved biological and interfering particles,
710 Atmos. Meas. Tech., 10, 4279-4302, 2017.

711 Shiraiwa, M., Ueda, K., Pozzer, A., Lammel, G., Kampf, C. J., Fushimi, A., Enami, S., Arangio,
712 A. M., Frohlich-Nowoisky, J., Fujitani, Y., Furuyama, A., Lakey, P. S. J., Lelieveld, J., Lucas,
713 K., Morino, Y., Poschl, U., Takaharna, S., Takami, A., Tong, H. J., Weber, B., Yoshino, A., and
714 Sato, K.: Aerosol Health Effects from Molecular to Global Scales, Environ. Sci. Technol., 51,
715 13545-13567, 2017.

716 Sivaprakasam, V., Lin, H.-B., Huston, A. L., and Eversole, J. D.: Spectral characterization of
717 biological aerosol particles using two-wavelength excited laser-induced fluorescence and elastic
718 scattering measurements, Optics Express, 19, 6191-6208, 2011.

719 Sodeau, J. R. and O'Connor, D. J.: Chapter 16 - Bioaerosol Monitoring of the Atmosphere for
720 Occupational and Environmental Purposes. In: Comprehensive Analytical Chemistry, de la
721 Guardia, M. and Armenta, S. (Eds.), Elsevier, 2016.

722 Stanley, W. R., Kaye, P. H., Foot, V. E., Barrington, S. J., Gallagher, M., and Gabey, A.:
723 Continuous bioaerosol monitoring in a tropical environment using a UV fluorescence particle
724 spectrometer, Atmospheric Science Letters, 12, 195-199, 2011.

725 Suphioglu, C., Singh, M. B., Taylor, P., Knox, R. B., Bellomo, R., Holmes, P., and Puy, R.:
726 Mechanism of grass-pollen-induced asthma, The Lancet, 339, 569-572, 1992.

727 Swanson, B. E. and Huffman, J. A.: Development and characterization of an inexpensive single-
728 particle fluorescence spectrometer for bioaerosol monitoring, Optics Express, 26, 3646-3660,
729 2018.

730 Taylor, P. E., Flagan, R. C., Miguel, A. G., Valenta, R., and Glovsky, M. M.: Birch pollen
731 rupture and the release of aerosols of respirable allergens, Clin. Exp. Allergy, 34, 1591-1596,
732 2004.

733 Toprak, E. and Schnaiter, M.: Fluorescent biological aerosol particles measured with the
734 Waveband Integrated Bioaerosol Sensor WIBS-4: laboratory tests combined with a one year
735 field study, Atmospheric Chemistry and Physics, 13, 225-243, 2013.

736 Wright, T. P., Hader, J. D., McMeeking, G. R., and Petters, M. D.: High Relative Humidity as a
737 Trigger for Widespread Release of Ice Nuclei, Aerosol Sci. Technol., 48, i-v, 2014.

738 Yu, X. W., Wang, Z. B., Zhang, M. H., Kuhn, U., Xie, Z. Q., Cheng, Y. F., Poschl, U., and Su,
739 H.: Ambient measurement of fluorescent aerosol particles with a WIBS in the Yangtze River
740 Delta of China: potential impacts of combustion-related aerosol particles, Atmospheric
741 Chemistry and Physics, 16, 11337-11348, 2016.
742

Tables

Table 1. Six scenarios explored, with varying combinations of pre-analysis treatment. (1) Fluorescence normalization refers to whether fluorescence intensity was normalized to particle size. (2) Variables logged refers to whether data was manipulated to produce a normal distribution.

Parameters	A	B	C	D	E	F
1. Fluorescence Normalization	1. No	1. No	1. Yes	1. Yes	1. Yes	1. No
2. Variables Logged	2. No	2. Yes	2. No	2. Yes	2. Yes, only AF/Size variables	2. Yes, only AF/Size variables

Table 2. Misclassification of 2-cluster solutions for 23 match-ups of two individual particle types (equal ratio of particle number, B-scenario) computationally combined before clustering analysis. Misclassification calculated as the sum percentage of particles misclassified in each cluster divided by the total number of particles. Three biological particle types (F2, B3, P9) compared separately to (a) non-biological particle materials and (b) biological particle materials. Particle number input was a subset of total population of particles experimentally analyzed.

(a)

	Non-biological particle materials							
	Diesel soot (Soot 4) S4	California sand (Dust 2) D2	Arizona Test Dust (Dust 12) D12	Suwannee River Humic Acid (HULIS 2) H2	Methyl-glyoxal + glycine aerosol (Brown carbon 1) BC1	Glyoxal + amm. sulfate aerosol (Brown carbon 3) BC3	White t-shirt (Misc. 2) WT	Wood smoke (Soot 6) WS
Aspergillus niger (Fungi 2)	*(1)* *0.1%*	*(3)* *2.6%*	(4) 6.1%	(5) 4.8%	(6) 2.5%	(7) **23.0%**	(8) **40.5%**	(9) 7.2%
P. stutzeri (Bacteria 3)	*(2)* *1.2%*		(10) 1.9%	(11) 1.2%	(12) 1.3%	(13) 6.1%	(14) **41.7%**	(15) 4.7%
Phelum pretense (Pollen 9)			(16) **22.7%**	(17) **23.2%**				

(b)

	Biological particle materials				
	S. cerevisiae (Fungi 4) F4	*Phelum pretense* (Pollen 9) P9	*P. stutzeri* (Bacteria 3) B3	*Taxus baccata* (Pollen 5) P5	*B. atrophaeus* (Bacteria 1) B1
Aspergillus niger (Fungi 2)	(18) **27.9%**	(19) **36.4%**	(20) 10.3%		
P. stutzeri (Bacteria 3)		(21) **18.3%**			(22) **65.4%**
Phelum pratense (Pollen 9)				(23) **46.8%**	

Table 3. Further exploration of 2-cluster solutions for the 10 match-ups of two individual particle types shown in Table 2 with misclassification >15%. Each match-up shown using three separate fluorescence threshold strategies in advance of particle input into cluster algorithm: (I) all particles included (no fluorescence threshold), (II) particles with fluorescence intensity < FT + 3σ removed, and (III) particles with fluorescence intensity < FT + 9σ removed. (a) Particle misclassification. (b) Total particle number used for clustering experiment.

(a) Percent misclassified

	Input	(7) F2 + BC3	(8) F2 + WT	(14) B3 + WT	(16) P9 + D12	(17) P9 + H2
Bio + Non-bio	(I) All particles	23.0%	40.5%	41.7%	22.7%	23.2%
	(II) Fluor. > FT + 3σ	10.3%	36.2%	24.3%	19.3%	3.4%
	(III) Fluor. > FT + 9σ	41.4%	32.6%	31.8%	45.3%	14.0%
	Input	(18) F2 + F4	(19) F2 + P9	(21) B3 + P9	(22) B1 + B3	(23) P9 + P5
Bio + Bio	(I) All particles	27.9%	36.4%	18.8%	65.4%	46.8%
	(II) Fluor. > FT + 3σ	13.3%	31.0%	20.0%	77.5%	24.9%
	(III) Fluor. > FT + 9σ	29.0%	28.6%	29.0%	66.7%	33.9%

(b) Number of particles

	Input	(7) F2 + BC3	(8) F2 + WT	(14) B3 + WT	(16) P9 + D12	(17) P9 + H2
Bio + Non-bio	(I) All particles	1,959	565	565	10,359	8,902
	(II) Fluor. > FT + 3σ	1,000	393	393	171	207
	(III) Fluor. > FT + 9σ	471	319	319	38	37
	Input	(18) F2 + F4	(19) F2 + P9	(21) B3 + P9	(22) B1 + B3	(23) P9 + P5
Bio + Bio	(I) All particles	10,000	8,900	10,000	10,000	10,000
	(II) Fluor. > FT + 3σ	9,600	8,500	9,800	10,000	10,000
	(III) Fluor. > FT + 9σ	9,200	8,100	9,700	10,000	7,895

Table 4. Particle fraction for each type and total particle number used as inputs for simulated~~synthetic~~ mixtures.

		F2	B3	P9	S4	D12	H2	BC1	WS	WT	
Mixture Number	Mixture Name	*Asp. niger* (Fungi)	*P. stutzeri* (Bacteria)	*Phelum pretense* (Pollen)	Diesel soot	AZ Test Dust	Suwannee River Humic Acid	Brown Carbon 1	Wood smoke	White t-shirt	Total Particle Number
1	4-Comp. A	25%			25%	25%	25%				680
2	4-Comp. B	25%			25%	25%			25%		680
3	High PBAP	25%	25%			20%	20%	10%			850
4	Low PBAP	12.5%	12.5%		15%	15%	15%	15%	15%		1134
5	Pollen			30%	10%	20%	20%	10%	10%		850
6	Indoor Air	20%	20%			20%	20%			20%	850

Figures

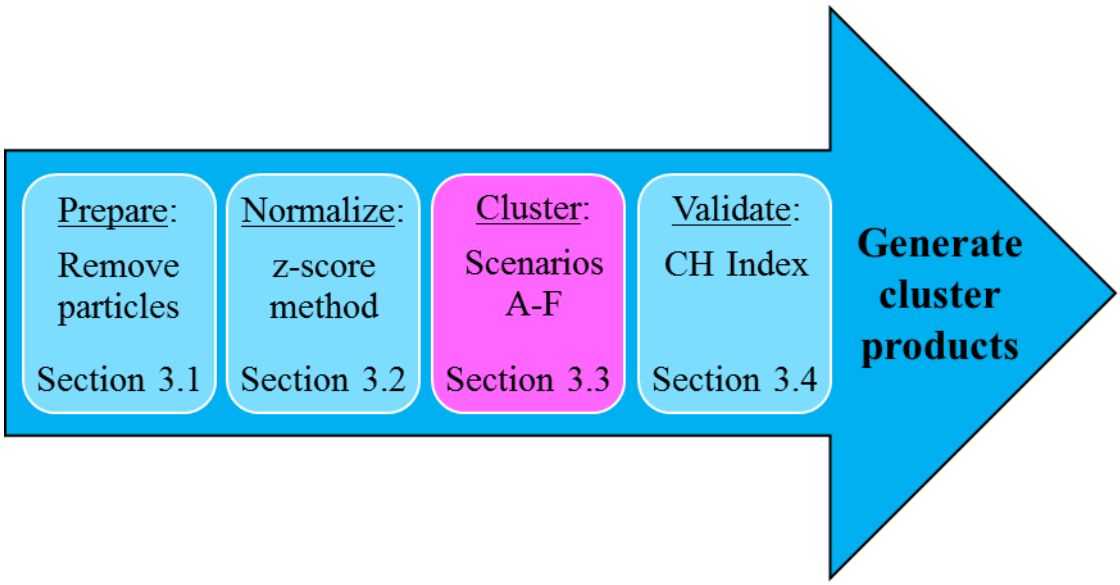


Figure 1. Schematic diagram showing the data preparation process resulting in the generated clustering products. Parameters within the pink box are the focus of this manuscript.

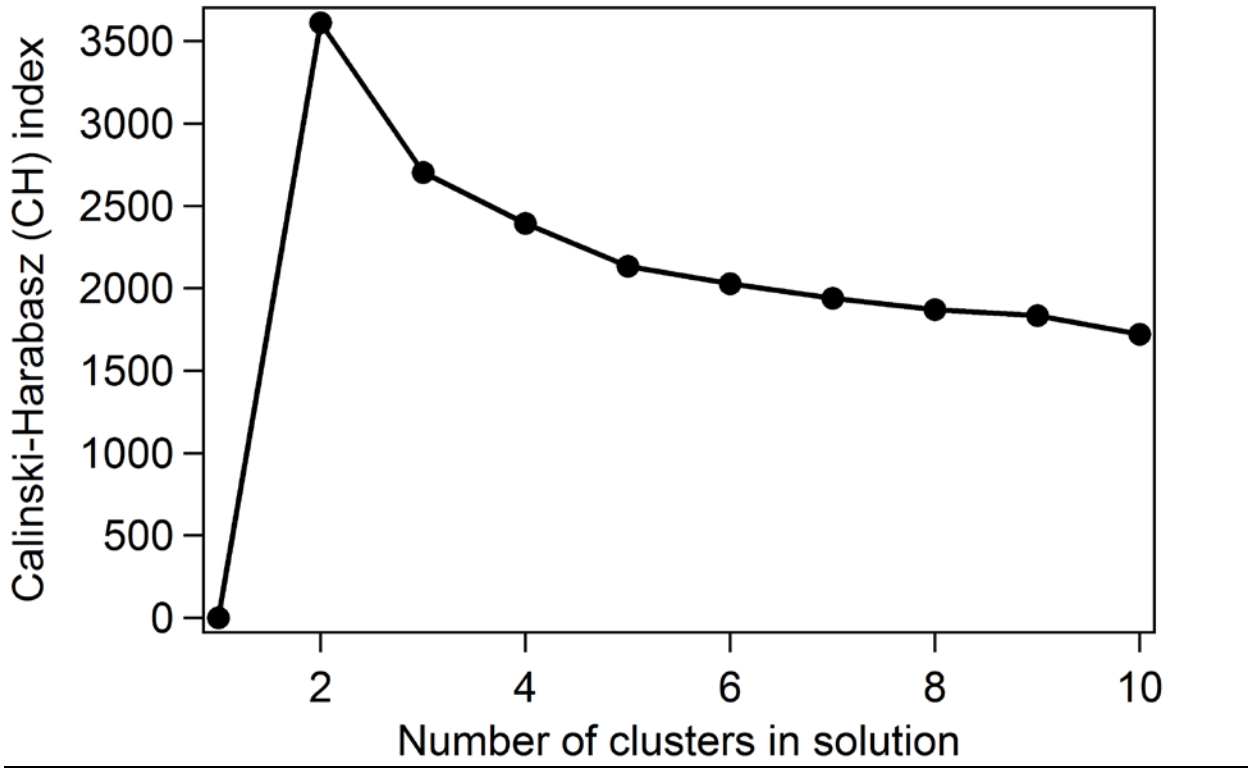


Figure 2. Example of Calinski-Harabasz Index plot for cluster experiment with input of *Aspergillus niger* and diesel soot (50:50 ratio). Optimal number of clusters is determined by the highest CH value.

	A	B	C	D	E	F
Fungi : Diesel						
50:50 Ratio	1.1	0.9	7.2	4.5	3.6	0.8
80:20 Ratio	64.8	4.1	4.5	2.9	3.8	76.5
20:80 Ratio	2.1	3.8	68.5	6.0	19.5	2.1
Bacteria : Diesel						
50:50 Ratio	50.0	1.2	6.8	4.5	31.6	50.0
80:20 Ratio	0.2	0.2	0.7	1.0	0.9	0.2
20:80 Ratio	80.0	0.3	68.2	0.3	43.7	80.0
Fungi : Dust						
50:50 Ratio	12.7	2.6	24.3	23.5	18.4	30.6
80:20 Ratio	76.6	9.0	20.0	25.4	25.4	29.3
20:80 Ratio	35.9	1.5	55.7	23.4	44.6	58.6

Figure 3. Cluster misclassification shown for three computational combinations of fungal spores (F2), bacteria (B3), ~~and~~ diesel soot (S4), and mineral dust (D12). Each combination explored with respect to ratio of input particle number using the scenario B and a 2-cluster solution for each experiment. Scenario letter A-F refers to scenarios summarized in Table 1. Red shaded region (and values) indicates the percent of particles misclassified. Blue shaded region represents the percentage of particles correctly classified.

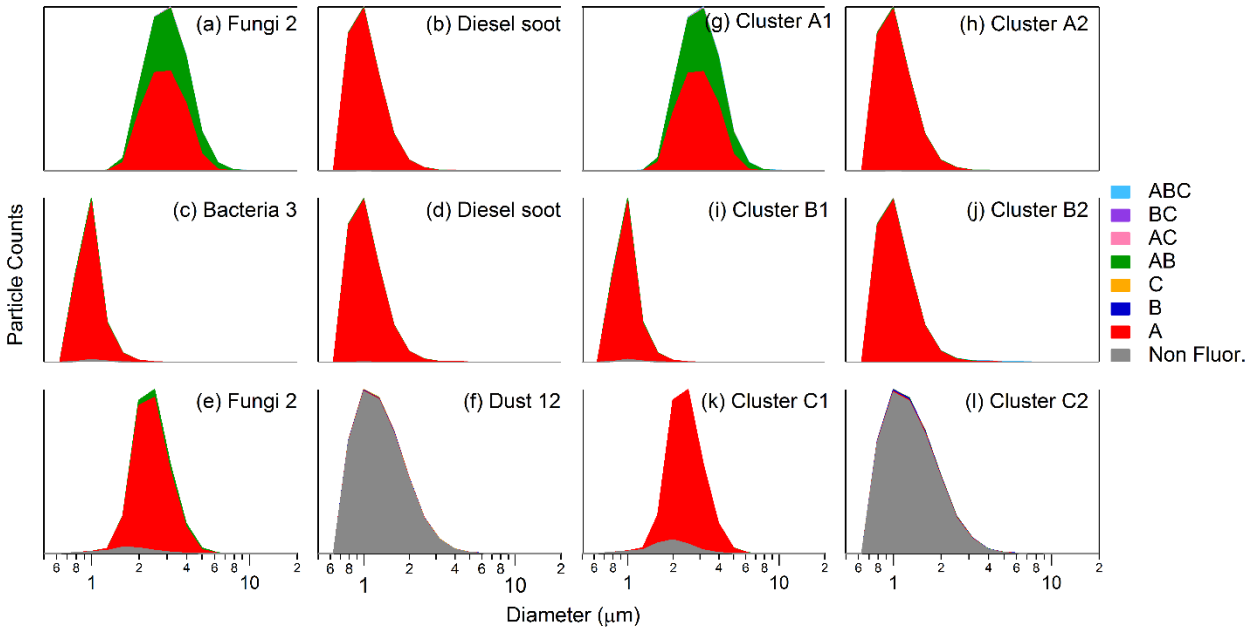


Figure 4. Particle type stacked category size distributions for input and output clustering results, using FT + 3σ threshold definition. Each experiment (row) shows match-ups of two particle types computationally mixed using 50:50 ratios, scenario B, and 2 cluster solutions. Left two columns show properties of input particles, right two columns show properties of cluster outputs.

Part A: Individual Clusters (Particle Number)

Part B: Grouped Clusters (Particle Number)

Part C: Summary (Cluster Quality)

Mixture #1: 4 Comp. - A

Cluster	F2	S4	D12	H2
1	163	2	22	23
2	7	1	123	67
3	0	0	21	80
4	0	167	4	0

Cluster	Fungi				Non-bio
1	163				47
2-4	7				463

Mixture #1

Total P.	Miscl.	Cat.
210	22.4%	Fungi
470	1.5%	Non-bio

Mixture #2: 4 Comp. - B

Cluster	F2	S4	D12	WS
1	167	2	23	4
2	2	3	88	10
3	1	0	55	156
4	0	165	4	0

Cluster	Fungi				Non-bio
1	167				29
2-4	3				481

Mixture #2

Total P.	Miscl.	Cat.
196	14.8%	Fungi
484	0.6%	Non-bio

Mixture #3: High PBAP

Cluster	F2	B3	D12	H2	BC1
1	0	197	0	0	0
3	200	6	13	2	6
2	9	10	133	79	6
4	4	0	21	88	25
5	0	0	3	1	47

Cluster	Fungi	Bacteria		Bio	Non-bio
1	0	197			0
3	200	6			21
2,4,5	13	10			403
1,3				403	21

Mixture #3

Total P.	Miscl.	Cat.
227	11.9%	Fungi
197	0.0%	Bacteria
424	5.0%	Bio
426	5.4%	Non-bio

Mixture #4: Low PBAP

Cluster	F2	B3	S4	D12	H2	BC1	WS
1	0	0	0	10	15	20	0
2	23	2	0	125	77	6	165
3	0	0	0	3	1	128	1
4	4	0	0	18	68	11	2
5	3	0	169	8	9	0	0
6	0	135	1	0	0	0	1
7	112	5	0	6	0	6	1

Cluster	Fungi	Bacteria		Bio	Non-bio
7	112	5			13
6	0	135			1
1-5	30	2			836
6,7				252	14

Mixture #4

Total P.	Miscl.	Cat.
130	13.8%	Fungi
136	0.7%	Bacteria
266	5.3%	Bio
868	3.7%	Non-bio

Mixture #5: Pollen

Cluster	P9	S4	D12	H2	BC1	WS
1	0	0	13	16	13	0
2	2	0	28	83	15	1
3	0	0	4	1	51	1
4	6	2	113	70	0	79
6	5	77	3	0	0	0
5	242	6	9	0	6	4

Cluster			Pollen		Non-bio
5			242		25
1-4,6			13		570

Mixture #5

Total P.	Miscl.	Cat.
267	9.4%	Pollen
583	2.2%	Non-bio

Mixture #6: Indoor Air

Cluster	F2	B3	D12	H2	WT
1	160	7	13	0	31
4	0	154	0	0	0
2	4	0	32	95	35
3	6	9	125	75	62
5	0	0	0	0	42

Cluster	Fungi	Bacteria		Bio	Non-bio
1	160	7			44
4	0	154			0
2,3,5	10	9			466
1,4				321	44

Mixture #6

Total P.	Miscl.	Cat.
211	24.2%	Fungi
154	0.0%	Bacteria
365	12.1%	Bio
485	3.9%	Non-bio

Figure 5. Overview of computationally simulated~~synthetic~~ mixtures. Six mixtures shown as groups of rows, with input particle fractions defined in Table 4. Part A (left columns) show particle number retrieved by each individual cluster and categorized by each input particle type. Part B (middle columns) show particle number categorized and grouped by particle classes (i.e. non-biological and biological). Part C (right columns) show misclassification of groups of particles. Colors: light green (fungal spores), blue (bacteria), pink (pollen), dark green (grouped biological), brown (all non-biological).