We would like to thank the reviewers for their comments. We apologize for the extended time between reviews and revision but this was due to incorporating the important extra work requested in the revised manuscript. We have made the changes suggested and responded to comments in a point by point format below. We believe this is a stronger paper as a result and thank the reviewers for their work.

**Reviewer #1**

1.) The main scientific work and ideas that were put into the paper are the growing of the forests and their validation including variable reduction and creation of the confusion matrix. Accordingly, four out of five figures are about these topics. And the results are interesting and offer new ways of looking at this kind of datasets. In contrast the abstract, mainly the introduction and the conclusion strongly focus on the prediction of aerosol classes.

In response to this reviewer comment we have added text in the three stated areas (abstract, intro and conclusions) to the emphasize the use of the general technique as opposed to the specific classification of aerosol spectra. For example, we now lead the introduction with the generalized technique and then move to the specific use here. We believe it is also critical to maintain the classification objective as this was our motivation for the work and the data we present.

> 2: 8-12 in the abstract now reads:
> *"Our primary focus surrounds the growing of random forests using feature selection to reduce dimensionality, and the evaluation of trained models with confusion matrices. In addition to classifying 20 unique but chemically-similar aerosol types, models were also created to differentiate aerosol within four broader categories: fertile soils, mineral/metallic particles, biological, and all other aerosols."*
>
> 3: 1-15 in the introduction:
> *"Following the introduction of random forests in the 1990s, recent developments in deep learning and neural networks have triggered a renewed interest in machine learning. This has led to the development of numerous easy-to-use, freely-available, open-source packages in popular programming languages like Python, and these tools are becoming increasing used in academia and industry. While random forests have been used for complex classification and regression analysis in various fields, studies that employ random forests in aerosol mass spectrometry remain sparse. Utilizing these tools, the primary purpose of our study is to introduce a framework for growing random forests, reducing dimensionality, ranking chemical features, and evaluating performance using confusion matrices. Such properties are desirable for SPMS studies, where input variables can become redundant and interpretability is more limited methods with methods such as cluster analysis and neural networks. Powerful analysis techniques such as those falling out of recent artificial intelligence research can prove useful for*

*helping to tease out the subtle yet significant impact that aerosol chemistry has on the climate system.”*

19: 16-21 in the conclusion:

“This study lays out a framework for training and implementing random forests on SPMS data, with a focus on dimensionality reduction and the evaluation of model performance with confusion matrices. A key benefit to the proposed method is chemical feature selection, which allows researchers to identify potentially important chemical markers between arbitrary groups of aerosols or identify sources of contamination. Additionally, the approach allows for differentiation of aerosols within a SPMS dataset, augmenting existing tools and reducing the need for a qualitative comparison between mass spectra.”

2.) The paper is in its present form hard to follow. Often the nomenclature is not consistent throughout the paper or doesn’t fit to the cited literature. For example, they do not use the proper term "random forest" but call it machine learning classifier, predictive model, classification model, rule-based probabilistic classification of a decision tree ensemble or supervised classification and even more important.

We thank the reviewer for the suggestion and agree more specific language should be applied to describe the random forest algorithm, which is the primary focus of the paper. The terminology through the paper has been modified to be consistent and precise per the comments, and we have removed sections that went into unnecessary detail.
While do note the terms mentioned are appropriate for introducing and motivating the random forest approach, they describe broader categories of machine learning models. The distinction between each was not necessary for this paper, beyond the mention of how supervised methods (i.e. classification via random forests) differ from unsupervised methods such as clustering. We have now tried to make this distinction and believe it is now more clear.

3.) While the basic algorithm to grow a random forest is presented. The underlying concepts (randomness, law of big numbers, assumptions, input parameter) and details of the validation process remain unclear…

As suggested, we have modified the text in several ways to address this comment, which has been broken into specific points and addressed:

Upon further inspection, the authors agree the concepts you mention such as randomness and out-of-bag sampling are key details surrounding random forests approach and needed further detail. Out-of-bag samples refers to spectra that are held out during the validation process to prevent training and testing on the same data, and the section has been revised to more concisely explain the procedure.

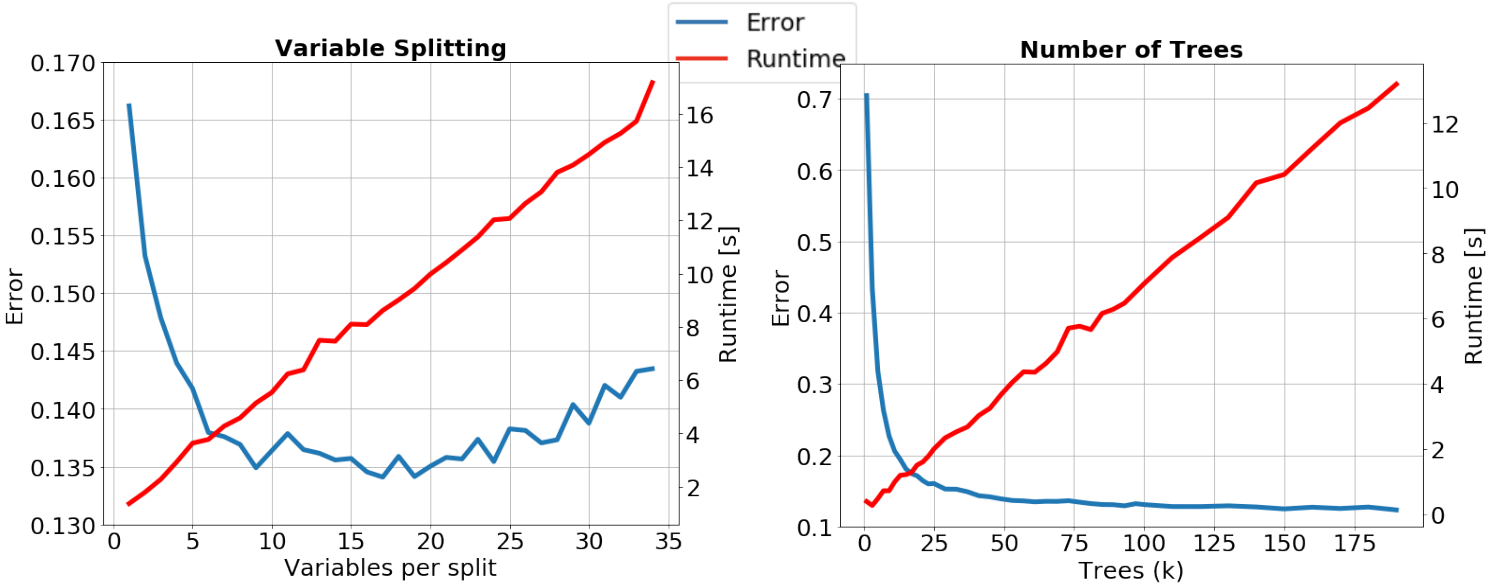
12: 2-10 now reads:

> *“Overall, the generalizability and robust performance of random forests is owed significantly to the series of random statistical procedures used to construct such models. An ensemble classifier reduces variability by averaging predictions over a series of independently trained models, and bagging introduces additional randomness by producing “perturbed” versions of the original data via random sampling of input data. The randomness used in constructing forests, both in bagging the training set and choosing variable splits, work to decorrelate the output of each tree even as the inputs become correlated [Breiman, 2001]. As the number of trees increases, the law of large numbers guarantees a convergence of the out-of-bag error to the generalization error.”*

…For example, k=1000 trees have been used for each forest but no further explanation is given why exactly this number of trees is the right one. Or a plot of the test set error against number of trees presented which would make this decision obvious. The number of random variables used to select the best split from is not specified nor its implications discussed….

- We have included various pieces of the information you requested in the paper, and agree the details surrounding runtime, memory requirements, and model selection are important. Additionally, supplementary plots that characterize how model performance and runtime scale are provided below. The plot also shows how runtime scales with number of trees and number of split variables.

> 11:12 - 20 now includes the requested information:
>
> *“The number of variables per split is chosen to be 11 and the number of trees is 110. The optimal model was determined by enumerating combinations of these parameters on a coarse grid and selecting the values that produce the lowest test error. Model behavior is primarily sensitive to the number of variables per split, and shows weak dependence on the number of trees and number of input variables beyond small values. As the number of variable splits increases, error decreases exponentially to a local minimum before again rising due to over fitting. Alternatively, as the number of trees is increased the error asymptotes to some nonzero value, a known characteristic of random forests where test error converges to the generalization error.*



…The treatment of the "out-of-bag" observations, which is the central means of validation is not comprehensible….

More detail has been added.

10:15 - 10:22 now reads:

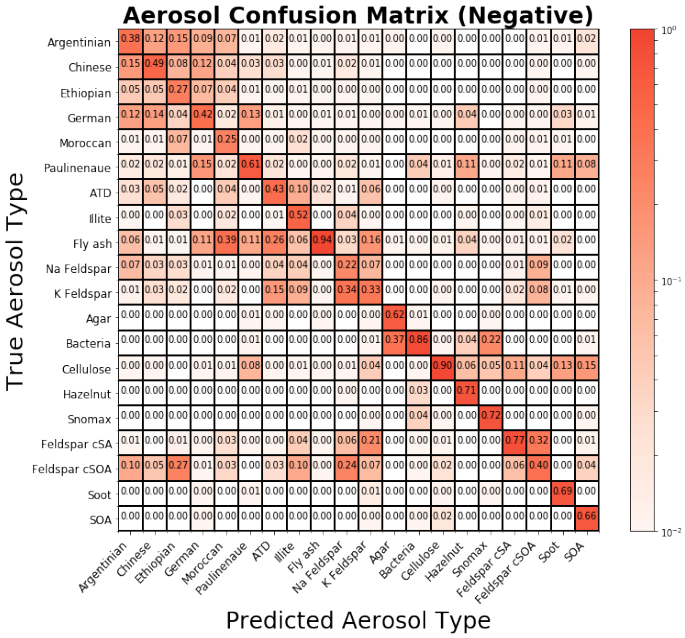
> *“On average, each tree is built with ~63% of the original data, leaving a portion of the training set unsampled. The unsampled data for each tree, known as ‘out-of-bag’ observations, are recorded and later provide a means to assess classification error for the forest. To determine model error, predictions are made on each point in the dataset using only the subset of trees that did not use the point for training. Each training point is left out at least once. This is analogous to making predictions with a separately trained forest that did not observe the point and prevents testing with the same data used for training.”*

…The resultant classification accuracies are not put into perspective; thus the reader can’t judge if the algorithm performs is a major improvement to other methods, of which the simplest would be to just use mean values of each aerosol class and use the most similar one as a prediction…

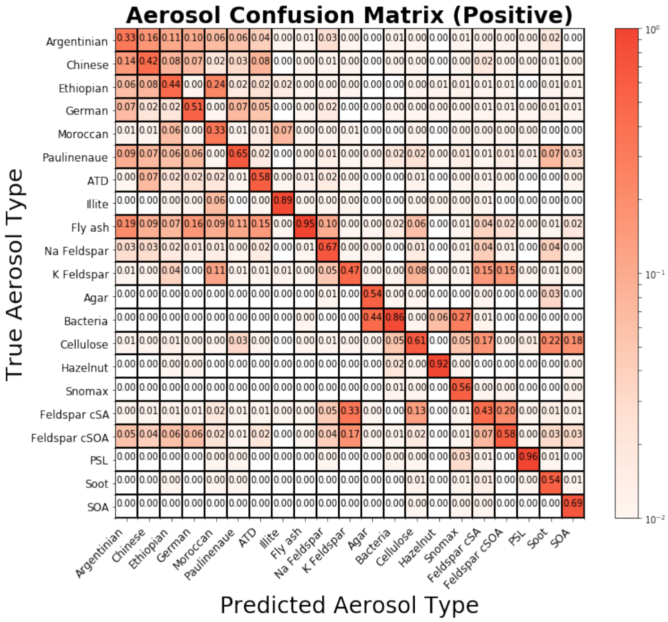
We have compared the technique to a simple classifier that uses the euclidean distance to assigned an unknown aerosol to the closest “mean” class vector. Confusion matrices for the broad categories have been included in the paper for comparison as part of figure 4, and matrices for all labels have been included below as a supplementary figure.

13:10 - 20 now reads:

*“To access relative model performance, we contrast the results with a simple classifier that compares unseen aerosols to a set of class mean vectors. Using the Euclidean distance metric, the unknown aerosol is assigned to the nearest class. This simple baseline classifier helps put results in the context of machine learning techniques that rely on distance-based metrics such as k-means and hierarchical clustering. K-means clustering attempts to divide the data points into k distinct clusters, representing spectra as vectors. Using Euclidean distance, the standard algorithm assigns points to centroids, or clusters, which are essentially mean vectors representing the average of all points in the cluster. Assuming perfect convergence of k-means clustering, where k is the number of aerosol classes, each cluster represents the mean of aerosol in that class. The random forest results below demonstrate many areas of improvement over the simple classifier.”*


Aerosol Confusion Matrix (Negative)

| | Argentinian | Chinese | Ethiopian | German | Moroccan | Paulinenaue | ATD | Illite | Fly ash | Na Feldspar | K Feldspar | Agar | Bacteria | Cellulose | Hazelnut | Snomax | Feldspar cSA | Feldspar cSOA | Soot | SOA |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Argentinian | 0.38 | 0.12 | 0.15 | 0.09 | 0.07 | 0.01 | 0.02 | 0.01 | 0.00 | 0.01 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 | 0.02 |
| Chinese | 0.15 | 0.49 | 0.08 | 0.12 | 0.04 | 0.03 | 0.03 | 0.00 | 0.01 | 0.02 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Ethiopian | 0.05 | 0.05 | 0.27 | 0.07 | 0.04 | 0.01 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| German | 0.12 | 0.14 | 0.04 | 0.42 | 0.02 | 0.13 | 0.01 | 0.00 | 0.01 | 0.01 | 0.00 | 0.00 | 0.01 | 0.00 | 0.04 | 0.00 | 0.00 | 0.00 | 0.03 | 0.01 |
| Moroccan | 0.01 | 0.01 | 0.07 | 0.01 | 0.25 | 0.00 | 0.00 | 0.02 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 | 0.00 |
| Paulinenaue | 0.02 | 0.02 | 0.01 | 0.15 | 0.02 | 0.61 | 0.02 | 0.00 | 0.00 | 0.02 | 0.01 | 0.00 | 0.04 | 0.01 | 0.11 | 0.00 | 0.02 | 0.01 | 0.11 | 0.08 |
| ATD | 0.03 | 0.05 | 0.02 | 0.00 | 0.04 | 0.00 | 0.43 | 0.10 | 0.02 | 0.01 | 0.06 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 |
| Illite | 0.00 | 0.00 | 0.03 | 0.00 | 0.02 | 0.00 | 0.01 | 0.52 | 0.00 | 0.04 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 |
| Fly ash | 0.06 | 0.01 | 0.01 | 0.11 | 0.39 | 0.11 | 0.26 | 0.06 | 0.94 | 0.03 | 0.16 | 0.01 | 0.00 | 0.01 | 0.04 | 0.00 | 0.01 | 0.01 | 0.02 | 0.00 |
| Na Feldspar | 0.07 | 0.03 | 0.03 | 0.01 | 0.01 | 0.00 | 0.04 | 0.04 | 0.00 | 0.22 | 0.07 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.09 | 0.00 | 0.00 |
| K Feldspar | 0.01 | 0.03 | 0.02 | 0.00 | 0.02 | 0.00 | 0.15 | 0.09 | 0.00 | 0.34 | 0.33 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.08 | 0.01 | 0.00 |
| Agar | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.62 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Bacteria | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.37 | 0.86 | 0.00 | 0.04 | 0.22 | 0.00 | 0.00 | 0.00 | 0.01 |
| Cellulose | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 | 0.08 | 0.00 | 0.01 | 0.00 | 0.01 | 0.04 | 0.00 | 0.00 | 0.90 | 0.06 | 0.05 | 0.11 | 0.04 | 0.13 | 0.15 |
| Hazelnut | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.03 | 0.00 | 0.71 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Snomax | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.04 | 0.00 | 0.00 | 0.72 | 0.00 | 0.00 | 0.00 | 0.00 |
| Feldspar cSA | 0.01 | 0.00 | 0.01 | 0.00 | 0.03 | 0.00 | 0.00 | 0.04 | 0.00 | 0.06 | 0.21 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.77 | 0.32 | 0.00 | 0.01 |
| Feldspar cSOA | 0.10 | 0.05 | 0.27 | 0.01 | 0.03 | 0.00 | 0.03 | 0.10 | 0.00 | 0.24 | 0.07 | 0.00 | 0.00 | 0.02 | 0.00 | 0.00 | 0.06 | 0.40 | 0.00 | 0.04 |
| Soot | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.69 | 0.00 |
| SOA | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.66 |


Aerosol Confusion Matrix (Positive)

| | Argentinian | Chinese | Ethiopian | German | Moroccan | Paulinenaue | ATD | Illite | Fly ash | Na Feldspar | K Feldspar | Agar | Bacteria | Cellulose | Hazelnut | Snomax | Feldspar cSA | Feldspar cSOA | PSL | Soot | SOA |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Argentinian | 0.33 | 0.16 | 0.11 | 0.10 | 0.06 | 0.06 | 0.04 | 0.00 | 0.01 | 0.03 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.02 | 0.00 |
| Chinese | 0.14 | 0.42 | 0.08 | 0.07 | 0.02 | 0.03 | 0.08 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.00 | 0.00 | 0.01 | 0.00 |
| Ethiopian | 0.06 | 0.08 | 0.44 | 0.00 | 0.24 | 0.02 | 0.02 | 0.02 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 | 0.01 | 0.01 | 0.00 | 0.01 | 0.01 |
| German | 0.07 | 0.02 | 0.02 | 0.51 | 0.00 | 0.07 | 0.05 | 0.00 | 0.00 | 0.02 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 | 0.00 | 0.01 | 0.01 |
| Moroccan | 0.01 | 0.01 | 0.06 | 0.00 | 0.33 | 0.01 | 0.01 | 0.07 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Paulinenaue | 0.09 | 0.07 | 0.06 | 0.06 | 0.00 | 0.65 | 0.02 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.02 | 0.02 | 0.00 | 0.01 | 0.01 | 0.01 | 0.01 | 0.07 | 0.03 |
| ATD | 0.00 | 0.07 | 0.02 | 0.02 | 0.02 | 0.01 | 0.58 | 0.00 | 0.01 | 0.02 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 |
| Illite | 0.00 | 0.00 | 0.00 | 0.00 | 0.06 | 0.00 | 0.00 | 0.89 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 |
| Fly ash | 0.19 | 0.09 | 0.07 | 0.16 | 0.09 | 0.11 | 0.15 | 0.00 | 0.95 | 0.10 | 0.00 | 0.00 | 0.02 | 0.06 | 0.00 | 0.01 | 0.04 | 0.02 | 0.00 | 0.01 | 0.02 |
| Na Feldspar | 0.03 | 0.03 | 0.02 | 0.01 | 0.01 | 0.00 | 0.02 | 0.00 | 0.01 | 0.67 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.01 | 0.04 | 0.01 | 0.00 | 0.04 | 0.00 |
| K Feldspar | 0.01 | 0.00 | 0.04 | 0.00 | 0.11 | 0.01 | 0.01 | 0.01 | 0.00 | 0.05 | 0.47 | 0.00 | 0.00 | 0.08 | 0.00 | 0.01 | 0.15 | 0.15 | 0.00 | 0.01 | 0.00 |
| Agar | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.54 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.03 | 0.00 |
| Bacteria | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.44 | 0.86 | 0.00 | 0.06 | 0.27 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 |
| Cellulose | 0.01 | 0.00 | 0.01 | 0.00 | 0.00 | 0.03 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.05 | 0.61 | 0.00 | 0.05 | 0.17 | 0.00 | 0.01 | 0.22 | 0.18 |
| Hazelnut | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.00 | 0.92 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Snomax | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.56 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Feldspar cSA | 0.00 | 0.01 | 0.01 | 0.01 | 0.02 | 0.01 | 0.01 | 0.00 | 0.00 | 0.05 | 0.33 | 0.00 | 0.00 | 0.13 | 0.00 | 0.01 | 0.43 | 0.20 | 0.00 | 0.01 | 0.01 |
| Feldspar cSOA | 0.05 | 0.04 | 0.06 | 0.06 | 0.02 | 0.01 | 0.02 | 0.00 | 0.00 | 0.04 | 0.17 | 0.00 | 0.01 | 0.02 | 0.00 | 0.01 | 0.07 | 0.58 | 0.00 | 0.03 | 0.03 |
| PSL | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.03 | 0.01 | 0.00 | 0.96 | 0.01 | 0.00 |
| Soot | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.01 | 0.01 | 0.00 | 0.00 | 0.54 | 0.01 |
| SOA | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.69 |

…It is not given which implementation of the algorithm is used. Nor how long a typical random forest generation lasts and how this runtime scales with respect to number of particles, number of trees, number of split variables, etc. . Along with the memory requirements which are missing too, these are basic and easy to provide information that help to compare this method to other methods….

*11:20 – 12:1*

*“The models were trained with the Python 2.7 Scikit-learn module on a MacBook Pro with 16 GB 1600 MHz DDR3 memory and a 2.5 GHz Intel Core i7 processor. A typical*

*random forest model took about 5-10 seconds to train, and we found a linear relationship between runtime and both the number of trees and variables per split."*

4.) The random forests have been grown on chemical information and the size of individual aerosol particles, but some of the aerosol classes are not chemically defined. (e.g. multiple fertile soil classes, ATD) This basic contradiction is not clearly addressed.

We agree with the reviewer and have expanded the paper to define this more clearly. This is in the form of the new paragraph in the introduction on page 5 which also includes detail that some of this issue stems from the complex nature of atmospheric aerosols that are often defined by source as opposed to type. Furthermore, we have color coded Table 1 to make the distinction more clear.

4:16 – 5:9

"Chemical composition of an individual atmospheric aerosol particle is a complex interplay between its primary composition at the source (i.e. dust, biogenic organic, anthropogenic organic, soot, etc.) and its atmospheric processing up to the time of detection. Atmospheric processing can include any combination of coating with secondary material, coagulation and cloud processing. Even distinct primary aerosol types can have similar mass spectral markers. For example, fly ash, mineral dust and bioaerosol can all contain strong phosphate signal [Zawadowicz et al., 2017]. Secondary material is often difficult to differentiate from primary material, but even minor compositional changes can be atmospherically important. As one example, mineral dusts are known to be effective at nucleating ice clouds [Cziczo et al., 2013]; however, despite minor addition of mass, atmospherically processed mineral dust is less suitable for ice formation [Cziczo et al., 2013]. As a second example, ice nucleation in mixed-phase clouds has been suggested to be predominantly influenced by feldspar, a single component among the diverse mineralogy of atmospheric dust [Atkinson et al., 2013]. Using current SPMS data analysis approaches, it can be difficult to detect these minor yet important compositional differences and new robust and generalizable analysis techniques are critical."

5.) To me the section dealing with the blind test data does not fit to the abstract and introduction which present the random forest as a tool specifically suited for this use-case. After showing 80+

There are two primary factors that help explain differences between the test set and blind set, which are both due to the way the experiment and sampling were conducted : a) transmission efficiency b) coagulation. During the course of the experiment, we expect the mineral dust and SOA to coagulate. Since aerosol types were reported by AIDA before particles enter the chamber, it is not possible to quantify exactly what fraction of the particles picked up an SOA coating. Moreover, there would have been a time dependence to the coagulation process.

Additionally, through coagulation, there is the possibility of effectively producing a particle type not in the training set, depending on the exact mineral component of the mineral dust used by AIDA. While it is known a mineral dust was included in the chamber, the exact composition of the dust was not known. While our training set contains K-Feldspar coated with SOA, a different type of SOA-coated mineral dust will appear unique to the model. Because the generalization performance of supervised classifiers is ill-defined for particles not included in the training set, this could lead to performance that is not captured by the confusion matrices. Given the experimental uncertainties from transmission efficiency and coagulation, as well as the model uncertainties highlighted in the confusion matrices, we believe the results reveal skill in using random forests to pick out distinct aerosols. In future studies, uncertainties can be reduced by adding additional particle labels or accounting for transmission efficiency, but coagulation will likely remain an inherent uncertainty. The limitations of transmission efficiency and coagulation are also noted at the end of the results section.

The caption in Figure 5 has been updated to state these factors more clearly

> 35: 6-10 now reads
>
> "Notes (1) the soot in the blind mixture was known to be below the instrument detection limit and therefore is not expected to be found in the data, (2) coagulation of SOA and mineral dust, which occurred after aerosol input to the chamber, appears as the "other" category, (3) the aerosols types reported by AIDA do not account for PALMS transmission efficiency."

So my suggestion would be to split the paper and resubmit both parts in a thoroughly revised version one part with a clear focus on the algorithm and its general applicability to SPMS data including a real comparison to methods currently used (fuzzy-cmeans, manual decision tree, k-means). And the other with a thorough analysis of the blind data set explaining in a comprehensible way the measured spectra based on all available information, statistics and assumptions. If it is not possible to explain the measured spectra in a controlled laboratory experiment like the one described, the use of the instrument to characterize atmospheric aerosol populations would be quite limited.

The co-authors had a discussion regarding this suggestion but have decided that keeping the paper in something similar to the original format was the best course of action. Our reason is that the current format allows us to present a new technique for aerosol mass spectra classification and then use it on a single well constrained data set. We expect to expand its usage on future (e.g. field) data sets.

We acknowledge the reviewer's comments have focused on the general method and they raise an interesting point about using multiple method on a data set and then doing a cross comparison. We need to therefore comment that, in keeping this a new method to use with a demonstration on this data set, a multi-method comparison is well beyond the scope of our goal. The reviewer notes several different methods that have been used for aerosol mass spectral analysis. However, in practice, each instrumental group has focused on one or two techniques.

Setting up a means to apply several techniques to a single data set is therefore non-trivial and would occupy a significant amount of time. Again, we agree the cross comparison is interesting but would certainly represent months of research time beyond the scope of what we are attempting to do here.