Reviewer #1:

Major concern is that this method would be difficult to be applied for detection of metallic components in atmospheric aerosols having low concentrations. Although their new machine leaning technique and development of low cost delay generator can be so promising, the feasibility of this technique to detect and/or to quantify atmospheric metallic components in PM should be still low, based on their current dataset (solution-based sample results using highly concentrated ones with uncertain amount of deposited or ablated mass).

Their deposition mass calculation is still unclear to readers.

1. Can authors be sure if their calculated mass is the same as the deposited mass on the 1 mm diameter tungsten ground electrode? The detailed experimental procedure with schematic should be required.

Authors' Response: The authors have used a 5µL pipette in order to carefully control deposition of sample on the electrode. Using a larger diameter pipette would have caused some the deposited mass stick to the Tungsten electrode circumference, reducing the mass on the electrode surface. Therefore, using this small pipette and slowly depositing mass on the electrode, assures that the calculated mass is similar to the deposited mass.

2. Their deposit samples on the 1 mm diameter tungsten ground electrode were ablated completely by the spark plasma (i.e., is it the same as the ablated mass)?

Authors' Response: What we present is a calibration-based technique in that we develop a calibration between the known mass deposited and the spectra obtained. There is no feasible way to measure the ablated mass, but this is accounted for in the calibration. We agree with the reviewer that the ablation process might change with the total deposited mass. However, there is substantial literature showing that such matrix effects should not be significant for the low amounts of mass that we are depositing and that will be deposited in the final instrument design.

3. I wonder if their calculated masses are relevant to the concentration range of metallic components in PM2.5 or PM10 in atmosphere.

Authors' Response: The concentration of the ambient toxic metals depends on where the instrument is tested. For industrial application, the concentration of metal PMs can reach to few µg/m^3. Much of the potential application of this instrument is in communities in proximity to sources that are potentially emitting a substantial amount of toxic metals. For the proposed instrument, the detection limit and sampling time are coupled together. Increasing the sampling time will increase the total collected mass on the ground electrode and hence improve the detection process.

Reviewer #2:

The authors have to some degree addressed my comments and the manuscript has improved. However, I still find it very hard to read, mostly because it has no clear structure and logical flow to it. Sections that describe developments, experiments, results and analysis are ultra short and leave out much detail. Much is left to the reader's imagination and interpretation, which is not a good scientific reporting style. The objectives of the work are still not entirely clear to me. Throughout the manuscript, the same phrases are repeated, but details on how these objectives are achieved is missing in the right places. Also, in my report on the original manuscript I have pointed out several of these shortcomings. The reply of the authors was to repeat the same phrases, which is not very impressive.

In conclusion, I am disappointed by the manuscript and by the authors' efforts to improve it. As somebody not immediately working in the field of atmospheric science, I accepted the review of the manuscript, because the title and the premise sounded interesting. I believe the manuscript is publishable, but it takes more than the current revisions to make the manuscript readable and understandable. More details are given below:

Authors' Response: The author would like to thank the reviewer for his/her valuable comments and we apologize if we did not address the prior comments sufficiently well. As mentioned before both in the manuscript and rebuttal, the objective of the current study is to develop a cost-effective instrument to detect toxic metallic elements in air. There are other techniques such as XRF and ICP-MS for such a task, however they both are expensive and difficult to handle in field measurements. The authors attempted to develop an instrument that addresses these shortcomings using low-cost components and improve the performance of the new instrument by employing advanced machine learning techniques.

- The introduction has now improved and it is easier to read. However, the section on the spark-induced breakdown spectroscopy (SIBS) and laser-induced breakdown spectroscopy (LIBS) is still very difficult to read. This is in part, because of the citation style. The citations are not separated from the actual text in any way and sometimes it is impossible to tell where a citation ends and the text starts again.

Authors' Response: We apologize and have changed the citation style to match that of the journal. We hope during the final production these citation style become more readable for readers.

- I had recommended to add more context and background to the introduction. The authors talk at length about different detection methods, but do not say much about machine learning in the introduction, although half of the manuscript is about machine learning. I had recommended to cite and then discuss several recent machine learning approaches for spectra and spectroscopy. Since then there have been even more in the literature. This request has been blatantly ignored by the authors. Makes me wonder why we reviewers write referee reports in the first place.

Authors' Response: The authors greatly appreciate the reviewer time and effort for providing useful inputs. In order to address the reviewer concern, the following has been added to the introduction starting at Pg. 3, Line 59:

"Recently…"

- The objectives are now stated more clearly: "We employed two complementary approaches: (1) decreasing the cost of the electronics associated with SIBS and (2) incorporating advanced data analysis techniques to improve quantification and limit of detection." However, what is still not clear to me is how the cost is decreased (see below). An example of frequent circular arguments in this manuscript is a sentence in the same paragraph in the introduction: "The expensive components such as spark generation and delay generator have been developed to reduce the overall cost." To me this sentence sounds like `we developed expensive components to reduce the cost', which would, of course, be a contradiction. I suppose the authors meant to say that they improved the expensive components so that they become cheaper. Maybe the sentence could also be improved.

Authors' Response: The following has been revised to address the reviewer comment at, Pg. 3, Line 75:

"To reduce the overall cost, inexpensive replacements for necessary components, such as the spark generator and delay generator have been developed in the lab"

- Both for the spark generation system and the delay generator, the manuscript states that these are expensive components. But it does not say why. What is expensive about them? I am not an expert on these instruments and most likely most readers will not be either.

Authors' Response: The cost of the spark generation system according the quote we received was ~$12K and delay generator ~$5K. At the beginning of this study, the authors asked the very same question and in fact that was the main motivation to develop these components in the lab. The cost of our spark system and delay generator are ~$600 and ~$50 respectively. The quoted spark system ($12K) can generate high-energy spark with high repetition rate (1Hz). However, considering low concentration of PM and the required time to sample enough materials makes this repetition rate completely unnecessary. For the commercial delay generators, they can generate delays from ps to ms range with high accuracy (fs). However, our measurements only require a few µs delay with accuracy of less than ~0.2µs. Therefore, purchasing a commercial delay generator also is unnecessary. Therefore, the authors decided to develop those components in the lab to reduce the overall cost. To address the reviewer comment, the following has been added at Pg. 3, Line 83:

"Setting up a spark emission spectroscopy system requires expensive components. However, depending on the application some of these components can be replaced. Components such as spark generator and delay generator can cost up to $10K and $5K respectively. Here according to our application and needs we developed these components for less than $600 and $50 respectively."

- How is the spark generator improved? The manuscript devotes literally only 1 sentence to this “In our setup, a 10Ω resistor maximizing power dissipation in the spark gap, while minimizing oscillations.” And this sentence is not even grammatically correct. It has no verb. Such a short description is simply not good enough and not scientific at all.

Authors’ Response: Having 0Ω resistor will cause oscillation in current and voltage in the spark gap. Therefore, it is necessary to add a resistance to damp these oscillations. However, adding too much resistance will increase the energy dissipated in the resistor and hence reduce the overall energy dissipated in the spark gap. The following has been revised at Pg. 4, Line 98:

“We found that 10 Ω resistor maximizes the power dissipation in the spark gap, while minimizing oscillations.”

- Spectra collection: this section is also very short. How many spectra were collected in the end and at what settings? What is the spectral resolution? How are the spectra discretised, etc.? In other words, what is the data set?

Authors’ Response: For each concentration more than 10 spectra have been collected and used for model development. Considering 4 sets of concentrations for each element and 4 elements in total, we collected 160 spectra for the entire study for each element. In order to address the reviewer question, the following has been added to the manuscript at Pg. 5, Line 123:

“For each concentration more than 10 spectra have been collected and used for model development.”

- What I am missing throughout is a section and a schematic figure that illustrates the whole process. What does the instrument look like and how do the different components that the authors improve fit in? How do we go from spark generation to a spectrum? How is the spectrum further processed with machine learning? This schematic would ideally also illustrate the overall objectives of the research.

Authors’ Response: In order to address the reviewer question, the asked schematic has been added to the revised manuscript at Pg. 3, Line 84.

- The results section has more figures than it has text. Very hard to understand anything.

Authors’ Response: We have added some text in response to other reviewer comments. The figures are necessary to show what we found. We hope that the result is more acceptable to the reviewers.

- In the clustering section on page 6, I could not follow anything anymore after “For each element, 0.1, 1, 10 and 100 ng…” What does “element” refer to here? I think this part might now be describing data collection? Maybe? And therefore answer one of my earlier questions? But then seamlessly without paragraph break the paragraph continues with “Feature scaling is a standard preprocessing step…” How do we get from clustering to spectra collection to feature scaling? I cannot follow the logic anymore.

Authors' Response: The objective of this study was to detect and quantify toxic metallic PM in air. These PMs are comprised of different metallic elements as listed in the Table 1 and 2. "Element" is referring to the metallic elements such as Cu, Ni, etc that we used in this study to test our spark emission spectroscopy setup. After ablating the deposited mass and recording the spectrum, feature scaling has been used as a preprocessing step to improve the optimization process for out machine learning model. To address the reviewer concern, we have added the following at Pg. 6, Line 141:

"After ablating the deposited mass and recording the spectrum, feature scaling has been used as a preprocessing step to improve the optimization process for out machine learning model."

- "Figure 7 illustrates the Boltzmann plot…" What is a Boltzmann plot?

Authors' Response: Boltzmann plot is a well-known approach in plasma spectroscopy for obtaining plasma temperature. The authors explained how it is constructed in the previous rebuttal. Basically, the slope of the fit to the Boltzmann plot provides the plasma temperature. To address the reviewer comment the following has been added at Pg. 6, Line 146:

"Plasma temperature can be obtained as: …"

- Throughout the whole results section, the authors speak of features and feature selection. A "feature" in machine learning is an abstract concept that refers too patterns in the data input. It is ok to use the word features, but then it has to be clear what the data is, what the data input is to the machine learning and what the features are. I have figured out now, that the input into the machine learning is x-y-z data in the form of discretised spectra. x are the frequencies and y the intensities of the spectra and z are the detected masses of metals. A feature is then simply one point in the spectrum. This information may seem trivial to the authors, but without it, the manuscript is much harder to read and understand.

Authors' Response: The authors are glad that the manuscript can explain the model now. This information has been provided in Pg. 7, Line 155:

"where $x(i) \in R^{2048}$ and $h_\theta(x(i))$ represent the normalized spectrum and the LASSO concentration prediction based on spectrum (i) (x(i)), respectively, and where y(i) is the known concentration corresponding to spectrum (i)."

- in equation 1 "m" appears to be the number of spectra, i.e. the data set size. Would be good to say that in the text.

Authors' Response: The following has been added at Pg. 7, Line 156:

"*m* refers to total number of spectra and …"

- Figure 8 is used to justify the choice of the regularisation constant c. I find this procedure also hard to follow. Instead of plotting the training and test loss as function of the number of features (which is related to c) it would be more illustrative to plot it as a function of c. Or have an alternative x-axis at the top of the graph that shows the c values.

Authors' Response: The c constant does not provide much physical information while the number of feature provides the total number of spectrum features (lines) that has been used in the model. Here, feature refers to spectral lines in a spectrum. Hence, the total No. of feature indicates the total No. of spectral lines used to build the models. In order to clarify this, the following has been added at Pg. 7, Line 164:

"k indicates the total number of features (spectral lines) used to build the model."

- The whole discussion on the univariate "technique" is very hard to follow. What was actually done? What is plotted in Figures 12 and 13?

Authors' Response: Considering the vast interest toward machine learning techniques these days, the authors wanted to point out that machine learning is not the solution to everything and there is no guarantee that multivariate model always perform better than univariate model specially if the multivariate model has not trained efficiently. In order to address the reviewer comment, the following has been added at Pg. 9, Line 188:

"To evaluate LASSO performance, we compared LASSO with …"