This study developed and demonstrated a point sampling method to automatically measure emissions from a large-scale of individual vehicles. In this works, the authors present their system that can be used for particulate matter (PM) and gas emissions measurements, which is notably independent of vehicle type. They find that when using their peak detection algorithm (TUG-PDA), they can separate vehicle-specific emissions down to a spacing of just a few seconds between vehicles. In this study, they present initial findings from the use of this method that collected ~100,000 vehicle records from several measurement locations, mainly in urban areas. When compared to equivalent remote sensing measurements, the authors found good agreement even with the newest standards which are harder to capture due to their lower emissions and the current remote sensing abilities. This paper is well written and organized.

This manuscript presents novel work on the development of a plume detection system. The authors have done a lot of work to respond to and update their work based on the last round of revisions. With that being said, if the authors are able to update their work with the minor revisions listed in this report, this manuscript should be accepted for publication.

Line 19:
Define NOx here instead of in line 21

Line 47-48: "Other PM metrics such as PN or BC cannot be accurately determined using these systems…"
This point needs to be further clarified for the reader.

Table 1 /Table 2: Because Table 2 is not referenced in the text and is just hanging in the section that is placed in, it's recommended to add to Table 1 or putting it in the Appendix and referring to it in the main text.

2.2.1 Pre-processing:
This section has a lot of technical information that does not fit well into the bulk of the manuscript. It's recommended to simplify to the following:
1. Raw data from the instruments are time aligned
2. The CO2 is the default time resolution (keep statement about if instruments have large response time differences)
3. Outliers are filtered (state metrics for this) and measurements are smoothed

Lines 194-201:
This section leading up to describing the algorithm do not add value to the methods section and ends up being more distracting. The audience should understand the concept of PS by this part of the paper and therefore, there is no need to provide these detailed ideas. It's recommended to cut or slim to one introduction statement on why you developed the algorithm.

Line 211: "There must be either at least two gradients"

Which two gradients are being referred to? As in two pollutants need to be rising and having a gradient? Please clarify in the text.

Figure 3:
This figure can go into the appendix. It is too much to take in and understand in the main text. Also, it's recommended to add step numbers to make it easier to follow the diagram.

Figure 4:
It would be helpful to have the axes go lower than background so the reader can see how the background level is determined and potentially used as a stop (while recognizing that this specific example is not due to background but because of a passing vehicle)

Line 269:
Another recent paper on plume detection was published and should be referenced here. https://www.sciencedirect.com/science/article/pii/S1364815222003000

Section 2.2.3:
This section can be revised to be more direct about exactly what this method does. Also, generally, the methods section still seems quite lengthy. It is recommended to highlight the specific and unique points that applies to this PDA system.

Lines 306-311:
These sentences make the TUG-PDA sounds like it hypothetically can get down to a detection of 3 s but it is not able to in many cases, which contrasts a lot of the high level take aways from this paper. It is suggested to reword in the more active and present tone to express exactly what the system can do and what factors are able to be adjusted for varying situations.

Figure 5a:
For V4, the BC integrated areas seems to be cut off early. Though it may be following the rules, the BC plume both starts later and ends later than the corresponding CO2 plume and therefore, should be integrated to basically be a time shifted version of the CO2 plume instead of being cut off which would lead to an underestimation of the EF (this is discussed later on when the author states "The median EFs were 19% lower than in cases without interference"

Because of this, I believe it is very important for the authors to directly respond to the previous comment from Referee 2: Have you characterized how different fractions of peaks captured and assumed baseline concentrations impact resulting emission factors, using the subset of peak events that were 100% isolated with all pollutants starting and ending at the background condition? Authors: We thank the reviewer for this suggestion. We have not evaluated that in detail, but that is a very interesting suggestion for future investigations.

The authors have done some interesting analysis already that answers some of the referee's questions but they should go a step further and address how this ultimately would effect a TUG-PDA users emissions output.

Figure 6:
Define what the whiskers (Confidence intervals? What percentage?) are in this box and whisker plot

Line 366:
These laboratory measurements will need to be further defined especially with respect to determining limits of detection for BC as this is something that is not typically done.

Line 378:
This sentence and the paragraph need to be clearer about what differences between the two instruments and their performance are. This sentence I believe is applied to only BC emissions from the AE33 vs from the BCT. The BCT measurements can be separated for the two passing vehicles while the. Measures from the AE33 are not able to be seperated. Is that correct?

Section 3.2.1.
The focus of this section is specifically to compare two BC instruments while the bigger scope of this work is to developing the PDA. I think that this section should be remove or added to the appendix or framed as a case study to emphasize the importance of understand the instruments used with the PDA.

Line 389:
Please edit to be clearer. What is VSP?

Figure 8.
What is roadside?
Also, the fits do not have any statistical information and therefore do not add any meaning takeaways. Edit figure to clearly show the message the authors wish to convey.
Is there a strong statement that supports the placement to be in the middle in order to captures the higher levels of CO2?

Table 3:
Needs to be moved up in the text to where is it referred to. Do not leave it dangling at the end.