Anonymous Referee #1

General comments

Description: This discussion paper describes emission factors of diesel-powered trucks and buses in Mexico City measured using both the Aerodyne mobile laboratory and on-road remote sensing. The targeted compounds include CO, NOx, SO2, selected VOCs, PM, black carbon (BC), and particulate organic carbon (OC). The two methods produced similar results. BC emission factors were consistent with those measured in other studies, while the OC/BC ratio was larger than found in California. Emission factors generally agreed with those used in the EPA MOVES-2014b model for NOx and BC and were higher for CO, OC, and selected VOCs.

Relevance:

Heavy duty diesel-powered vehicles are responsible for substantial amounts of BC and NOx emissions, yet there are limited on-road measurements of emissions from these vehicles. This work adds to the database of such measurements and shows that the chasing method with a mobile lab and the on-road remote sensing method produce comparable results, so it is fair to synthesize results across these different types of studies.

Assessment: The work contributes useful information about emissions from diesel engines. The writing and figures are very clear and informative. The paper illustrates the strengths and weaknesses of each of the two methods for measuring emission factors. The paper could be strengthened through some reorganization of the Results and Discussion and addition of statistical tests.

We thank the reviewer for the constructive comments on the paper.

Specific comments

1. p. 4, line 10: A little more information about the prescribed driving routes and operation of the vehicles would be useful. What was the range of speeds? Were the engines always warmed up beforehand?

We thank the reviewer for this suggestion as this will allow the results to be more adequately compared to future studies of emission characteristics of diesel vehicles in Mexico. We have now included a more detailed description of the driving conditions (the range of speeds and accelerations of the vehicle sampled) during the tests in the supplemental material document.

2. p. 7, lines 24-26: "Since the measurements were obtained in similar prescribed driving routes, the results show a wide range of average emission factors associated with each vehicle engine and emission control characteristics." The wording and logic are not quite right here. I think the authors intend to emphasize that driving conditions were very similar for all vehicles, so differences must reflect variability

between engines and control systems. But later, they assert that there is large variability even for the same vehicle.

We thank the reviewer for calling our attention to this ill-constructed phrase. As pointed out, we want to emphasize that the driving conditions were very similar for all vehicles. We have modified the paragraph accordingly:

"Since the measurements were obtained under similar prescribed driving routes, differences in results mainly reflect variability among vehicle engines and emission control characteristics."

As stated, the results indicate that even after controlling for driving routes the observed variability of emission factors still can be large. This is in agreement with current understanding of real-world emissions as compared to laboratory-based studies and the growing acknowledgment that engine performance can produce large variability under real-world operation conditions.

3. p. 8, line 7: The comparison of emission factors among different vehicle types begs for statistical tests of differences. This is true for the presentation of differences by control technology, too.

We thank the reviewer for this valuable suggestion. We have now performed statistical testing for the significance of the results between vehicle types, by control technology, and between measurement techniques.

We have analyzed the statistical significance between control technologies for $PM_{2.5}$ EF using non-paired non-parametric Wilcoxon Rank tests and found that with a 95% confidence level the results for the EPA98 and EPA04 are significantly different. Similarly, differences between EURO3, EURO4, and EURO5 EF were found to be significantly different with a 95% confidence level. However, the rest of the tests indicated that the results for the EPA98 and the EURO3 (the older technologies sampled) were not significantly different with a 95% confidence level. Therefore, the following paragraph has been included in the text:

"Non-paired Wilcoxon Rank tests indicate that there is statistically significant difference (at the 0.05 significance level) between the $PM_{2.5}$ emission factors obtained for the EPA98 and EPA04 control technologies as well as among the EURO3, EURO4, and EURO5 technologies. However, the results for the EPA98 and the EURO3 technologies were not significantly different."

We similarly performed non-paired Wilcoxon Rank tests for comparing the emission factors by vehicle type for each pollutant. We found that CO, NO_x, and SO_2 from service trucks, urban buses, and Metrobuses were significantly different among them, whereas their VOCs measured (C_2H_2, acetaldehyde, benzene, toluene, C2-benzenes), and PM components (BC, OC, and inorganics) were not statistically significantly different. On the contrary, VOCs and PM-components emission factors obtained from Turibuses were statistically different from the corresponding emission factors from service trucks, urban buses, and Metrobuses. Thus we have included the following paragraph:

“Non-paired Wilcoxon Rank test indicate that there is statistically significant difference (at the 0.05 significance level) between emission factors from service trucks, urban buses, and Metrobuses for the CO, NO_x, and SO_2 pollutants, whereas their corresponding VOCs, BC, OC, and PM-inorganic emission factors were not significantly different. VOCs, BC, and PM-inorganic emission factors from biodiesel-fueled Turibuses were significantly different from the corresponding emission factors from service trucks, urban buses, and Metrobuses.”

In addition to the analysis suggested by the reviewer we also evaluated the statistical significance of the results for CO and NO emission factors that were obtained with both the chasing and the remote sensing techniques. Since these represent co-sampled data we used paired t-test with a 0.05 significance level. The results indicate that in both cases of CO and NO co-samplings there is no significant difference between the results obtained by the two measurement techniques, with a confidence level of 95%. Therefore, we have now added the following paragraphs to the results:

“Paired t-tests indicate that there is no statistical significant difference (at the 0.05 significance level) between the two measurement techniques for both cases of CO and NO emission factors.”

4. p. 8, line 26: The paragraph about the limitations of the sample size should be moved to the Discussion section.

As suggested by the reviewer, we have moved the discussion on the limitations of the sample size to the Discussion section.

5. p. 9, line 5: The comparison between the two methods in Section 4.1 seems it belongs more in the Results section than in the Discussion section because it is a straightforward presentation of results that address one of the objectives of the study.

We have moved the comparison of the two methods to the Results section.

6. p. 9, line 11: For comparison of the two methods, the authors chose to use the 10 seconds of AML data leading up the instant of remote sensing, which lasted 1 second. Why not isolate the 1-2 seconds of AML data that best correspond to when the remote sensing measurement was captured?

For the estimation of the emission factors using the chasing technique, the second-by-second measurements are integrated over a time period to account for the dispersion of the emission plume. Thus, if too few data points are included in the integration the resulting emission factor may not properly reflect the plume development and unnecessary uncertainty is introduced in the analysis. Based on our past experience with data analysis of this technique, we consider a good conservative number of data points for plume development is about 10 seconds as the basis for the choice of integration time. Thus, we have now complemented the following sentence:

"Since the remote sensing technique measures the emission factor of the sampled vehicle only while it passes through the detectors, only the emission factors obtained with the mobile laboratory ~10 seconds before and up to the corresponding actual moment of co-sampling with the remote sensing detector were considered for the comparison between the two techniques. Thus, we assume that a time period of 10 seconds is sufficient to capture a large portion of the emission plume sampled by the mobile laboratory."

7. p. 11, line 15: I assume that all the vehicles tested in this study were running on petroleum diesel, so results for B10 and B20 biodiesel are irrelevant to the present study and do not merit mention here, or they require greater justification for inclusion in the comparison.

In the discussion section of the paper we compare our results on the emission factors from biodiesel vehicles to the only other available literature study of similar measurements in Mexico that used B10 and B20 blends. We believe that, given the very limited information currently available, the comparison information and discussion presented is a valuable addition and thus we have decided to keep it in the manuscript.

8. p. 12, line 10: Can the authors comment on why there are differences in the OC/BC ratio compared to that found in other studies? Might altitude explain some of it or dilution? The mobile lab and remote sensing detect fresher, less diluted plumes compared to tunnel studies.

There are several possible reasons why the results show higher OC/BC ratios in comparison to other studies. These include differences in conditions derived from the environment (e.g., altitude, temperature), technical sampling methods (capturing fresh versus more diluted emissions), and diesel fuel composition. Unfortunately it is not possible from our results to quantitatively assess the contributions from these factors as it is beyond the scope of this study. Dedicated experiments controlling for these factors as well as vehicle technology and driving conditions could help to quantify the impacts of these factors. Nevertheless, it is possible to argue that the higher OC/BC content in the Mexican results obtained with the mobile laboratory are not due to differences in the sampling technique used in tunnel studies. As the reviewer pointed out, the former capture more fresh emissions than tunnel studies and, therefore, secondary formation of organic aerosols in the air masses would only increase the OC/BC during in the tunnel study sampling (Massoli et al., 2012), which is in the opposite direction needed to explain the observed differences.

No samples were obtained in our study of the diesel fuel employed, and thus it is not possible to know its exact organic chemical composition and its effects on emissions. Although a detailed chemical composition of diesel fuel by PEMEX (Mexican National Oil Company) is not publicly available, an official report indicates a predominant fraction of paraffin compounds of linear molecular chains with 11 to 12 carbons and a maximum 30% (in volume) of aromatics (IMP, 2014). In principle, a dedicated experiment could be set up to investigate the effects of OC formation due to differences in PEMEX's diesel fuel composition, but this is beyond the scope of this study. We have therefore included the following paragraph:

"Several factors including driving conditions, vehicle technology, and diesel fuel composition can contribute to the observed differences, but the quantification of these contributions is beyond the scope of this study. Nevertheless, the higher organic content of the emissions in the sampled Mexican vehicles with respect to those measured in California by Dallman et al., (2014) illustrate the large emission differences in PM composition that can be found in diesel fleets around the world, thus further indicating the need for locally adjusting the emission factors databases in mobile emission models."

References:

IMP, Instituto Mexicano del Petroleo: Factores de Emision para los diferentes tipos de combustibles fosiles que se consumen en Mexico. Informe Tecnico F.61157.02.005. Dirección de Servicios de Ingeniería Gerencia de Servicios en Ingeniería Región Centro-Norte. 2014. Available: http://www.inecc.gob.mx/descargas/cclimatico/2014_inf_parc_tipos_comb_fosiles.pdf

Paola Massoli , Edward C. Fortner , Manjula R. Canagaratna , Leah R. Williams , Qi Zhang , Yele Sun , James J. Schwab , Achim Trimborn , Timothy B. Onasch , Kenneth L. Demerjian , Charles E. Kolb , Douglas R. Worsnop & John T. Jayne (2012) Pollution Gradients and Chemical Characterization of Particulate Matter from Vehicular Traffic near Major Roadways: Results from the 2009 Queens College Air Quality Study in NYC, Aerosol Science and Technology, 46:11, 1201-1218, DOI: 10.1080/02786826.2012.701784.

9. Table 2: This could be moved to the supplemental information, as a more digestible summary of the results appears in the figures. Footnote 4 mentions "hundredths" of Metrobuses. Should this be 101 Metrobuses, the number sampled?

We agree that the information shown in Table 2 is somewhat dense. However, we believe it is important to present directly in the main manuscript a summary table of the average results obtained with both techniques. Therefore, we have decided to maintain Table 2 in the main text of the manuscript.

We thank the reviewer for the suggested edit on the footnote 4, the change has been made.

10. Figure 4: The NO figure shows small variability in mobile lab measurements and much larger variability in remote sensing measurements (large spread in the y-axis direction). This does not comport

with Fig. 1, which shows similar variability in the NO emission factors measured by both the mobile lab and remote sensing. Is it because these data points are limited to a much shorter period?

As pointed out by the reviewer, the differences in variabilities shown in Figures 1 and 4 are the result of the chosen periods for the comparisons but also on the averaging of results. Figure 1 shows "smaller" variability than Figure 4 because, as described in the manuscript, the figure is based on the averages of emission factors obtained from each vehicle, whereas Figure 4 shows the comparison of individual emission factors whenever they were co-sampled by the two techniques for the same vehicle.

Technical corrections

11. p. 4, line 10: The wording "the AML was positioned behind target diesel vehicles" makes it sound like the AML was stationary or attached to the vehicles. I suggest something like, "the AML followed behind target diesel vehicles," instead.

We thank the reviewer for the suggestion, the change has been made.

12. p. 5, lines 1-2: Rewrite "we have referred rBC to BC in this manuscript

The change has been made from "…, we have referred rBC to BC in this manuscript" to: "…, we refer to rBC as BC in this manuscript".

13. p. 5, line 2: "detection limit" should be "detection limits".

The change has been made.

14. p. 7, line 26: Change "observed" to "reported".

The word has been changed.