We gratefully thank all reviewers for the careful reading and valuable comments. Below we provide our point-by-point responses to the reviewers' comments. In the following context, raised comments/suggestions are marked in **black**, responses are presented in **red**, and changes to the manuscript/supplement information are indicated in **blue**.

Reply to Anonymous referee #4

The paper presents an interesting study where several methods for calculating the dilution ratio and the emission factor from a chased vehicle are tested and compared. It is an important contribution to the remote emission sensing research field, namely to the emission factor determination using the plume chasing technique.

We thank the referee for the positive comments towards our manuscript.

134 Could you estimate how good the assumption that the vehicle distance was constant between the two vehicles is since you had GPS on both the vehicle and the measurement station?

Good question. We tested this during the data preprocessing phase. We found that the GPS devices we used for measuring location were too inaccurate for measuring the distance between two vehicles.

As an example, we had a stop at the same location (+-1 meter from the certain sign). Before the stop, the distance between vehicles was recorded to be between 1 and 10 meters (5 and 95% quantile) and during the stop, the distance was between 3 and 12 meters (5 and 95% quantile). There was no clear difference in distance between vehicles. Especially during the stop, the distance was standard, and that could be checked from the video material. Hence the 10-meter difference between the rounds is at least mostly caused by the measurement accuracy of GPS devices.

For that reason, we assumed that the distance was constant during measurement period. Probably, the distance between the chased car and mobile laboratory in chase measurements should be recorded better in the future.

We added a sentence to the text to clarify this ("Unfortunately, the GPS data from the chased vehicle and ATMo-Lab was not accurate enough, so that the changes in the chase distance could have been estimated from the GPS data.").

In line 220 you state that Nraw uses CO_2 calculated from OBD data while in Table 2 it is in a row stating it is not using OBD data. Can you make this consistent?

N_{raw} method is using OBD data as the information of the exhaust flow rate is measured by OBD device. As the Table was placed so that it continued the next page, it was not clear to which category N_{raw} belongs. To make it clearer, we moved the whole Table to the next page, and inserted one empty line below N_{raw}.

Table 2: Division of the methods for calculating EF of a vehicle. OBD data means the data collected from the chased vehicle (see also Figure 2) and learning data means the data collected from other drives of the same vehicle and from other vehicles (including data from ATMo-Lab and, also from OBD if its data is used). Methods are introduced in more detail in subsections 2.3.1-2.3.7.

		Uses learning data: yes	Uses learning data: no
Uses OBD data	yes	MARS-OBD, NWD	N/CO_2 integral, N/CO_2 linear, N/CO_2 RRPA, N_{raw}
	no	MARS-chase	N/CO_2 Traficom

Is Audi type-approved as Euro 5? Euro 5 type approval started in September 2009, while Audi's registration year was 2008. Where is the information on the studied vehicles from?

The information of the studied vehicles is from the Traficom database. The information from the Traficom database states that Audi fulfills Euro 5 requirements.

186 – 193 In the N/CO2 Traficom description you say you used the Traficom data, can you also state how the values you got there compared to what you measured and what the producer of the vehicle states the values are?

Traficom provided consumption values for each vehicle. Traficom consumption values are based on the values provided by the manufacturer of the vehicles.

The consumption value used here is so called “combined consumption” which combines the urban and road driving consumption, the two other values Traficom provides. This was selected because it was the only value that was provided for all vehicles. In addition, it is most likely the most optimal out of those three for our driving route.

Table Y below shows the values from Traficom database, compared to measured (OBD) consumptions of the vehicles. We can see that usually the measured values were on average higher than the values provided by Traficom. Of course, for certain sections such as the downhill section, the fuel consumption can be much lower than the median value. Cold outside temperatures are at least one reason for larger median consumption but also the temperature of the engine might have an effect, especially at the beginning of the drive.

Table Y. Measured fuel consumptions and reported combined fuel consumptions from Traficom database.

Vehicle	Median measured fuel consumption (l/km)	Reported combined fuel consumption (l/km)
Audi	9,9	7,1
Seat	6,7	5,7
VW	8,4	7,6
Ford	7,1	4,6
Skoda1	6,8	5,4
Skoda2	7,4	7,5

We added the following sentences to subsection 2.3.2 to clarify the origin of Traficom consumption values and the nature of those values, as an average fuel consumption.

"Traficom consumption values are based on the values provided by the manufacturer of the vehicles. The values in the database are the average consumptions (in unit of $l/100\ km$), and hence the actual consumption at certain time might be over or under the consumption value in the database."

195-213 One-minute constant EF seems long. Looking at Figure 1, could you use a shorter time maybe 30s? How long did it take you to make a single round of 13.8 km? Was there much variability between runs? Would this affect the result? How are the zero values included in the final EF distributions since these are on log plots?

The drive time was between 19 min 41 sec and 22 min and 42 sec, including two 30 second stops and two 1 minute idles before and after the drive. However, there is variation between vehicles and within drives of one vehicle. See boxplots for ranges of drive times for each car in the figure 2 a) below.

EFs were calculated for the shorter time periods and then pooled to get an EF for the whole round. We didn't find any clear dependence between drive time and EF. See also Figure 2 b) below for median EFs from N to CO2 ratio linear method calculated as in Figure 5 for all the vehicles.

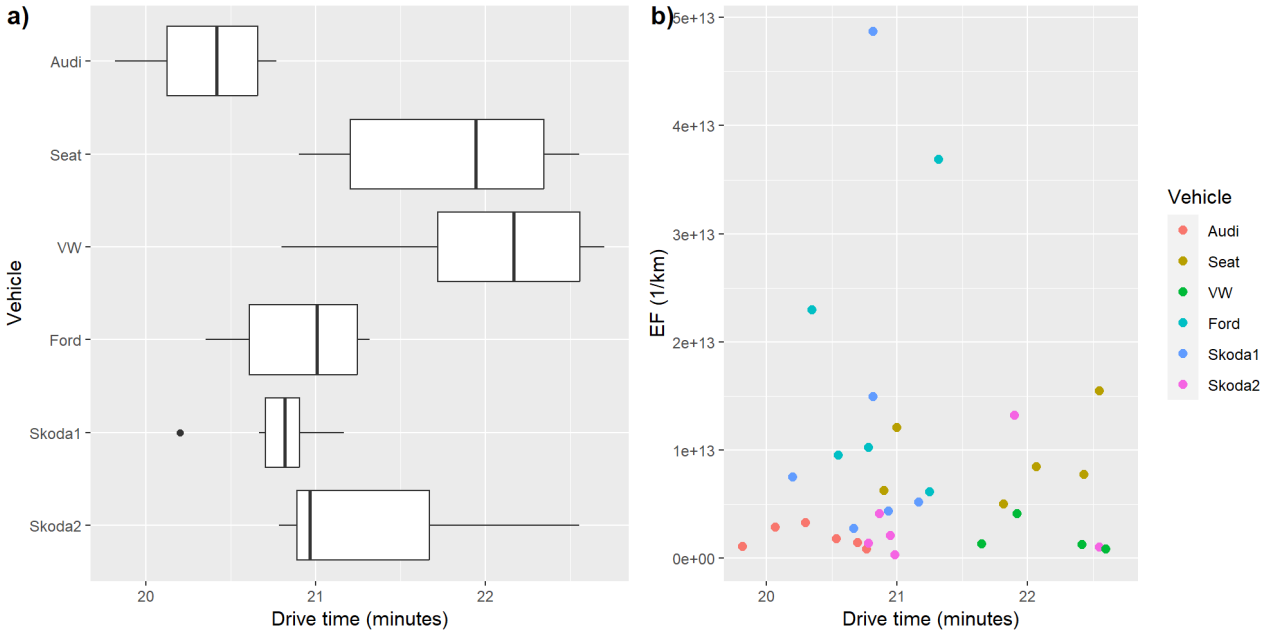


Figure 2. a) Boxplot of drive times for each vehicle. b) Scatter plot of EF from N to CO2 linear method (y-axis) as a function of drive time (x-axis).

Zero values were excluded from the EF plot in Figure 7. This has been mentioned in the figure caption. In Figures 5 and 6 all EFs were positive.

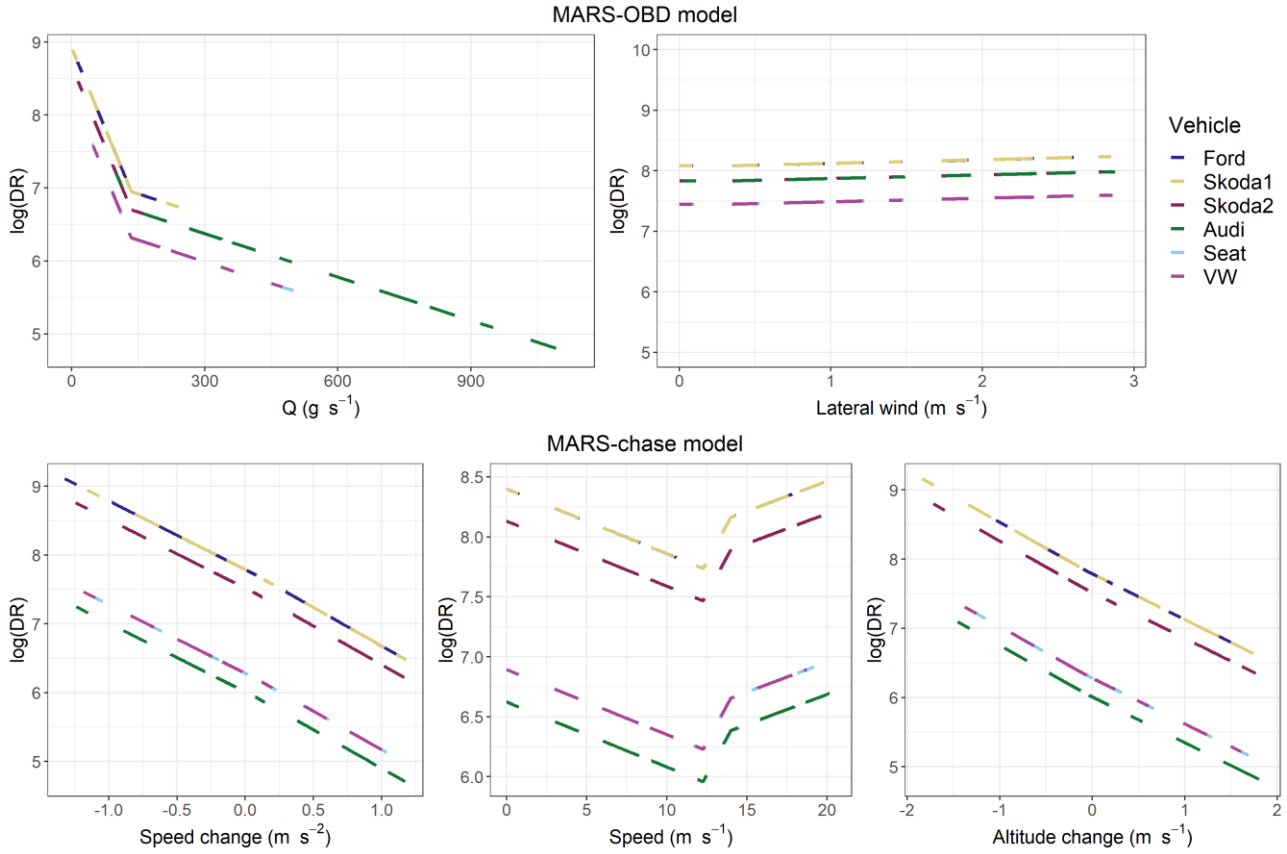
What values do whiskers on box plots represent (Figures 5 and 6)?

EFs were calculated to 100 bootstrap samples. For each sample we got an EF value, and those were represented as a boxplot.

We added a sentence "Whiskers are representing the distribution of EFs in different bootstrap samples." to both figure captions to clarify this.

Figure 4 is not clear – which splines are overlapping? Are all vehicles on all plots? It is difficult to see which vehicle is which with this color code.

We tried to make those splines more visible by removing the points from the figure and by changing the line type. See the new figure below this text. The overlapping splines are mentioned in the figure caption of Figure 4 but are hopefully now also easier to recognize also from the figure.



Maybe add a time resolution column to Table 3. Would different time resolutions influence your results?

We added the clarification to Table 3, for N/CO2 linear and RRPA methods using one-minute time intervals for linear models. Otherwise, all methods are using data with 5 second time resolution.

When using 5 second time resolution, we felt that 30 second intervals might be too short for linear fitting, as maximum of six points are measured in each 30 second period. For time resolution differences, we tested different time resolutions for linear and RRPA methods (30 sec, 1 minute). The results were slightly different when using different time resolutions, but the conclusions were not changed.

Minor issues

Use subscripts consistently throughout the text for "2" in CO_2 (there are two cases in Table 3 and in l. 175). Also degrees for degrees C in the caption of Figure 5.

We thank the reviewer for noticing these. These have now been corrected.

204 "without no need to" do you mean double negative or "with no need to" or "without a need to"

Changed to 'without a need to'.