Response to Reviews

Manuscript acp-2015-352

Black carbon, particle number concentration and nitrogen oxide emission factors of random in-use vehicles measured with the on-road chasing method

I. Ježek, T. Katrašnik, D. Westerdahl, and G. Močnik

We thank the referees for their comments. We have revised the manuscript to incorporate the feedback to all comments and advice. We have copied the remarks of each Referee in **black bold font** and our responses are given in regular black font. Manuscript text with revisions is given in regular blue font.

Section title "3 Emission factor measurement results" was missing in the AMTD publication. We have corrected this, changing the section numbering radically. We have also added the necessary additional references to the manuscript.

Anonymous Referee #1

Atmos. Chem. Phys. Discuss., 15, C5089–C5091, 2015

Received and published: 20 July 2015

There are a couple of parameters of the experiment for which I would like to know more details. What speed were these vehicles traveling at and what was their acceleration state. Engine RPM's would be great but I realize this is probably not obtainable. I think it is OK not to know that but it would still be useful to know the speed and whether these vehicles were traveling uphill, downhill or over relatively flat terrain.

Our measurements were performed mostly on highways and regional roads; because the sampled vehicles were picked randomly from the on-road fleet we could not measure their crankshaft rotational frequency, acceleration state or their exact traveling speed. We did, however, observe the speed of the mobile measurement platform while following a vehicle. The travelling speed was changing within each chasing episode, but for most trucks it was between 80 and 90 km/h, and for cars it was between 90 and 130 km/h. The terrain was never only uphill, downhill or flat but changing within a single measurement. The changes in the terrain were never large (the slopes were not steep). Relating vehicles emissions to their speed, acceleration state or either of the engine covariates is not as straightforward task to perform as it may seem. The exhaust gases needs some time to travel from the engine through the exhaust system, where an emission peak would already flatten as it would exit the tailpipe (Scharma et al., 2005), then it would be subjected to great dilution as it would be emitted in the outside air and only after that it would reach the mobile measurement platform. Due to the differences in the vehicle power, load and traffic situations the mobile platform cannot constantly be traveling at the same driving distance behind the chased vehicle, therefore the exact changes in speed cannot be recorded. We are however working on a publication addressing the issue of relating the on-road EF to the vehicle's operational state comparing chasing measurements to PEMS - this is an extended topic that needs a lot of consideration on the engine operating conditions, post processing and measurement of vehicular emissions.

The manuscript was amended in section 2 (P. 15363, L. 5) with the information on the vehicle speed:

The travelling speed was changing within each chasing episode, for most trucks between 80 and 90 km/h, and for cars between 90 and 130 km/h.

For Table 1 I would like to know what the size range of particles is that are measured for both the Aethalometer and the FMPS. Does this cover the range of particles we expect to see in engine exhaust?

The size range of the instrumentation does cover the range expected in fresh engine exhaust. We have added to the Table 1 the measurement range of the FMPS: from 6 to 560 nm. The size range of the exhaust particles was reported in Ježek et al. (2015), where the chasing method was tested in controlled conditions and compared to other methods. It is shown there that the particle number emission factor in the size range from 50-200 nm correlates well with the BC emission factor and exhibits the same temporal variation. This is expected, but it confirms that the losses in the sampling system (including the inlet) are negligible. Similar exhaust particle size values were reported in e.g. Kittelson et al. (2006) and Vogt et al. (2003).

The size distribution range of the FMPS was added as a comment in Table 1.

* Particle size range 5.6 – 560 nm

Finally I just wonder whether there was anything about the roadways sampled which would potentially skew any results. I mean for instance was a certain type of heavy duty truck more predominate on the roadways sampled than on the average roadway? Or a certain type of car? I don't have any reason to think this would be the case but I would be concerned about this if doing the measurements and I suspect the authors were, just want to be sure.

The roadways we were sampling on were highways and regional roads. The highways are part of the East-West and North-South trans-European corridors V and X. This is one of the major corridors for the transport of goods within this part of EU. The regional roads are predominantly the most common roadways for commuting from one city to another. We know that the composition of the traffic does differ in the cities and on highways (part of our so far unpublished traffic analysis results). To be sure that the measured fleet is representative and not subjected to bias, we performed the detailed fleet composition analysis as reported in section "3.1 Comparison of sampled vehicle fleet and Eurostat data" (note that the section numbering has changed radically due to an omission of the section title "3 Emission factor measurement results"). The analysis presented there shows that the measured fleet composition does indeed reflect the national and EU fleet compositions.

Technical Comments/Corrections

Recommend you reword pg 15359 lines 23-26 "They concluded that restricting the emissions of trucks only in Beijing is not sufficient to reduce traffic related air pollution in there due to the number of out of area trucks that operate in Beijing". There are a few typos recommend you reread carefully to catch nothing major but fleet versus feet bridge words that aren't needed things like that

We changed the sentence on page 15359, lines 23 – 26 to:
Because numerous trucks registered outside Beijing operate in the Beijing area, restricting Beijing-registered truck emissions is not sufficient to reduce traffic related pollution in the city.

We changed the word *feet* to fleet on page 15365, line 9. Other minor typos were fixed as well.

Anonymous Referee #2

Atmos. Chem. Phys. Discuss., 15, C5364–C5367, 2015.

Received and published: 29 July 2015

Few lines should be added to extend the added value of using the chasing method compared to other methods (remote sensing) and its potentials (in future).

We added to page 15359, line 1 the following text:
The chasing method allows us to capture a range of EF from a single vehicle and to measure the EF distribution rather than just a single value: depending on engine operation state, each vehicle produces a range of EFs with most values around a representative value (median) and a long super-emission tail. The comparison of the chasing method and the stationary method can be found in Ježek et al. (2015). With a single stationary measurement, we capture only a single value of the vehicle's EF distribution and several repetitions of the vehicle would be necessary to obtain that vehicle's EF distribution. We believe that using the EF distribution for a single vehicle, measured in real driving conditions, and using the collective distribution of the vehicle fleet to model traffic emissions could improve model predictions. Knowing the EF representative value and the super emission tail allows the quantification of the effect of potential abatement measures, e.g. how changing a driving regime would influence emissions at a certain section of the city.

The authors have spent a lot of efforts to demonstrate that the fleet composition they have investigated is representative at a national/EU level. I do not understand the purpose. As long as you investigate a sufficient number of vehicles for every category (and get corresponding EF), there is no need here to mimic the fleet composition. Hence this section (2.3) may bring to the idea that the results shown in the paper are representative for every category of the fleet. This may not be the case given the limited number of vehicles investigated in some subcategories

The main results of the paper are the BC, PN and NO_x EF distributions for diesel cars, gasoline cars and goods vehicles. According to Ban-Weis et al., 2009, about ≥ 30 trucks should be a large enough sample, presuming that the sampling was indeed random, for the sample mean to equally likely to fall below or above the sample mean. Our category samples were larger than the above threshold for diesel cars and goods vehicles, and very close to the threshold for gasoline cars. This makes us confident that the sample is large enough to be representative of the on-road fleet during the approximate period of the campaign on East-West and North-South trans-European corridors V and X (please see also the answer to the Referee #1's comment, inquiring about the representativeness). We made a step further by investigating the fleet composition in contexts relevant for modeling. Investigating a sufficient number of vehicles should reflect the fleet composition unless there would be some bias e.g. some vehicles would be more frequent on highways than regional roads or a city center (as referee #1 also pointed out). We investigated the fleet composition to be sure that the measured fleet was representative of the fleet we were sampling from on the national and the European level. We wanted to be sure that there was no potential bias from the roads we were sampling on, or from the ability of the mobile platform to chase cars with superior performance (and hence driving faster than we could chase on the highway). Our main focus was on diesel cars because there were no previously reported BC EF for them measured in real driving conditions. The comparison of the fleets serves as an independent test which we use to show the representativeness of the car fleet on a national and European level. The trucks are much more versatile when talking about their purpose and hence the mass they have to carry and power they have to produce. Unfortunately, we were unable to get the registration information for many of the sampled trucks; nonetheless we show the representativeness of Slovenian fleet for Europe and trust our sample is big enough to be representative. We did not mimic the fleet composition, because this was not necessary as we have shown that the sample was representative. This is important to enable the measured emission factors to be used for modeling emissions on a broader scale.

To further clarify the reasoning behind the analysis of the representativeness, we now start section "3.1 Comparison of sampled vehicle fleet and Eurostat data" (note the change of section numbering) with the following paragraph:
The fleet sample size determines the representativeness of the measured fleet. According to Ban-Weis et al. (2009), about ≥ 30 trucks should be a large enough sample (presuming that the sampling was indeed random) for the sample mean to equally likely to fall below or above the sample mean. Our category samples were larger than the above threshold for diesel cars and goods vehicles, and very close to the threshold for gasoline cars. This makes us confident that the sample is large enough to be representative of the on-road fleet during the approximate period of the campaign on East-West and North-South trans-European corridors V and X. In order to establish the relationship of our data as representative of the Slovenian and the

average European fleet, we used Eurostat data to compare the size and age composition of the three investigated vehicle fleets.

Additionally, we have added to the start of section “3.1.2 Goods vehicles” the following text to reflect the representativeness and variability of the sampled goods vehicle fleet:
The goods vehicles are much more versatile in their purpose and hence the mass they have to carry and power they have to produce. We were able to get the registration information for many of the sampled vehicles (28 out of 47) to identify the technical differences between the vehicles. Below, we show the representativeness of the Slovenian fleet for Europe. Our sample seems big enough to be representative, given the previously published criteria (Ban-Weis et al., 2009).

More efforts should be put in the paper regarding uncertainties associated with 1) measurements, 2) the impact of limited number of investigated vehicles (e.g. representativeness). We have no idea how “stable” are emissions during the chasing phase. Although this technique is described elsewhere, it would help the reader to know more on how stable (representative) are emissions taken from one vehicle during the measurement phase
The uncertainties of the measurement were covered in detail in Ježek et al., 2015, where the chasing measurement method was compared to the stationary plume measurements and the calculation algorithm was put to test. The uncertainty of the median value, which we here use as the representative EF value for a single (!) vehicle, was estimated to be -24/+26 %. The uncertainty of the single measurement depends on the measured CO_2 and its signal to noise ratio. The increase in CO_2 due to the exhaust plume depends on the dilution rate, which depends on vehicle speed and mostly on how much CO_2 is emitted (i.e. the engine operation). We constrain the calculation of the time evolving EF when CO_2 values are low by using a 10 s integrating time interval instead of shorter intervals. This smooths out the high emission peaks which are already smoothed out by travelling through the exhaust system and the atmosphere to the measurement instruments (Ajtay et al., 2005), and constrains the calculation error, yet keeping the calculated median value unchanged. The effect of the dilution was explained in the Supplement and referenced in section 2.1. The impact of limited number of vehicles was investigated by Ban-Weiss et al., 2009, where they show that sampling ≥ 30 trucks should be a large enough sample. We covered the fleet representativeness with the independent test of comparing the age, fuel type and size of our measured fleet to the national registry data in section 3.1.

We have added the following paragraph to the end of section “2.1 Emission factor calculation”:

The uncertainty of the median value, which we here use as the representative EF value for a single vehicle, was estimated to be -24/+26 % (Ježek et al., 2015). This uncertainty is reduced when calculating the fleet EF distribution. The uncertainty of the single measurement depends on the measured CO_2 and its signal to noise ratio (Ježek et al., 2015). We constrain the calculation of the time evolving EF when CO_2 values are low by using a 10 s integrating time interval instead of shorter intervals. This smooths out the high engine emission peaks, which are already smoothed out by travelling through the exhaust system and the atmosphere to the measurement instruments (Ajtay et al., 2005), and constrains the calculation error, yet keeping the calculated median value unchanged. The dilution does not affect the measurements of the single vehicle EF as long as the CO_2 increase is above the limit reported in Ježek et al. (2015). We show this in a comparison between a PEMS measurement and a chasing determination of EF (Figure S2, data from Ježek et al., 2015). The impact of limited number of vehicles was investigated by Ban-Weiss et al. (2009), where they show that sampling ≥ 30 trucks should be a large enough sample.

Ajtay D, Weilenmann M and Soltic P.: Towards accurate instantaneous emission models Atmos. Environ., 39 (13), 2443–2449, doi:10.1016/j.atmosenv.2004.03.080, 2005.

The supplement was extended with a section S2 including Fig. S2 and Table T1:

Supplementary material S3 – additional uncertainty analysis

In order to investigate the effect of exhaust dilution on the determination of the EF by chasing, and to further explain the results of the running integration calculation, we evaluated the chasing method using tailpipe measurements of CO_2 by PEMS. In this test we wanted to see how mobile measurements match the direct in-exhaust measurements of the chased vehicle. From these measurements we calculated the dilution rate (DR) as a ratio of the CO_2 measured by PEMS and by the chasing instrument (Chang et al., 2009), and compared it to the calculated BC EF.

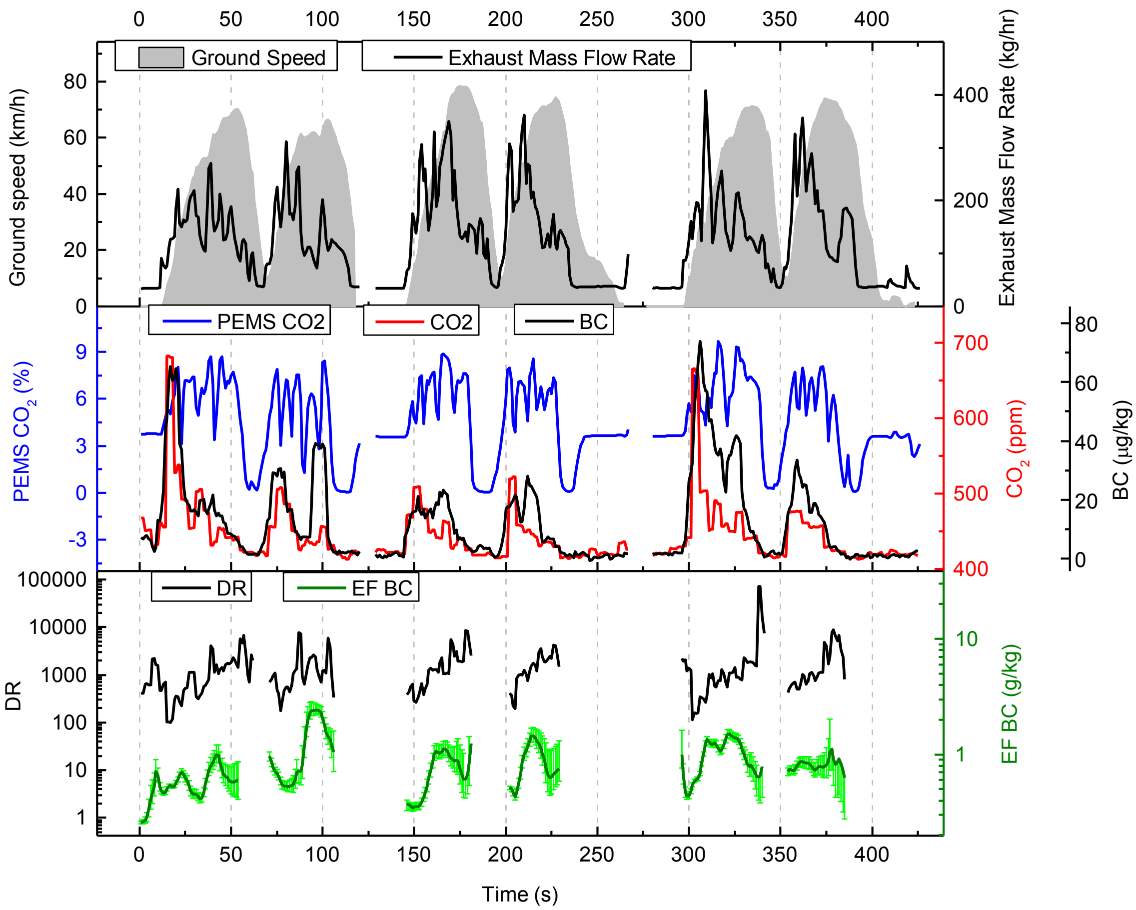

Figure S2. The tailpipe measurements performed with the portable emission measurement system (PEMS) are ground speed (shaded grey) and exhaust mass flow rate (black) – top; and CO_2 (blue) – middle. CO_2 and BC measured with the mobile station in red and black, respectively, also in the middle plot. The calculated dilution ratio (DR) in black and the BC EF in green – bottom. The BC EF does not show any significant dependence on the DR, and the uncertainty of EF (light green) increases when the CO_2 emissions are low. Note the log scales for DR and EF. Data from Ježek et al., 2015.

The results presented in Figure S2 first show how the exhaust mass flow rate changes with the vehicle speed for the analyzed turbocharged diesel engine. When the vehicle is accelerating, the power demand is high and so the exhaust flow rate increases and reaches the highest values at high engine speeds and loads. When the vehicle ceases to accelerate the flow rate drops; when the vehicle stops, and during certain braking sections, the engine idles and so the mass flow reach its minimum value. While driving, the concentration of CO_2 in the exhaust line varies from

roughly 4% to 9%, and drops to zero when the vehicle is braking. The jagged exhaust flow rate and CO_2 measured with PEMS reflect the gear changes as the mass flow is strongly dependent on the engine speed. The variability of the exhaust flow rate is often also reflected in the CO_2 measurements of the mobile platform, where we can observe similar drops in the CO_2 signal when a gear shift is made (e.g. after 25th to 30th and 160th to 170th seconds, and so on etc.).
The calculated DR values range from approximately 100, when we were in closer proximity to the chased vehicle and the speed of both vehicles was lower; to the maximum value of approximately 72000 when both the emitted CO_2 and the exhaust mass flow rate dropped. This occurred at the end of the track where we had to slow down to make a sharp U-turn. Notwithstanding this period, the maximum DR value was 8943 and the median 1077. This is similar to the measurements of Vogt et al. (2003), where they report dilution factors measured at approximately constant distances of 14, 50 and 100 m distance from a diesel car travelling 50-100 km/h to range from 926 to 9300.
The dilution does not affect the calculated BC EF. As we can see from Figure , the BC EF is at its highest just before the highest cruising speed is reached; and the dilution ratio is highest when the exhaust mass flow rate drops. This is consistent with the findings of Chang et al. (2009), who report that the dilution ratio depends not only on speed but also on the exhaust flow rate and other parameters, which are more important in the near wake region. The omitted parts, when the CO_2 drops below the background, overlap with the parts where there is little to no CO_2 emitted from the exhaust pipe, and so the CO_2 concentrations measured with the mobile station do not exceed the background level. However, the dilution rate does influence the uncertainty of the EF calculation. We can see that both the positive and the negative errors increase at the end of each run when the exhaust mass flow rate drops. We can also see that, at around the 170th second and after the 370th second, there is no positive error. This is because we do not calculate the EF when concentrations drop below the set baseline. If we had high background noise and low CO_2 emissions coming out of the vehicle, the error produced would have been large. We have, in part, limited calculating with low CO_2 by calculating the running integration EF using the 10 s time integrals instead of shorter intervals.
We will describe the EF variation measured with its range and selected percentile values. The range describes the spread of the sample data. The percentiles divide the sample so that for the *p*th percentile of a sample (p being a number between 0 and 100), as nearly as possible *p*% of the sample values are less than the p^{th} percentile and (100 − *p*)% are greater (Navidi, 2001). For each EF time series determined using different background levels, we calculated the distribution range, and the 25th, 50th (median), 75th and 90th percentile values. In
Table we can see that the negative relative error is smaller than the positive for values that are the median or higher. We can also see that the maximum value is calculated with the highest

uncertainty, but that the 90th percentile uncertainty already resembles the uncertainty of the 75th percentile. This means only a maximum of 10% of the values have an uncertainty that is higher than 25%. We can see that the error that arises from background determination is larger than that arising from instrument imprecision and the omission of CO and HC measurements.
In order to better resolve the EF variability, we have calculated the EF using a shorter integration time of 2 s. In order to calculate the 2 s integration interval we eliminated all values that were lower than the background plus two standard deviations of its variability, thereby excluding low CO_2 values from the calculation, which are the source of the highest EF calculation uncertainty. We can see in
Table, that an integration using a shorter time interval of 2 s yields in similar EF distribution values, only the maximum emissions are substantially higher. As Ajtay et al. (2005) reported for their laboratory experiments, the emission peaks flatten on their way from the engine through the exhaust system and the sampling lines of the measuring instruments. During our measurements there is a rapid, intense dilution of exhaust emissions in the atmosphere before they reach the mobile measurement platform. Even by integrating with shorter time interval we can only capture only a smoothened version of the emission peak. Since the uncertainty of such a calculation is rather high, we use the 10 s integration, which thus reflects an even more smoothed version of the super emission peaks produced by the engine.

Table T1. The emission factor (EF) calculated using different background levels shows that regardless of the set background, the EF distributions yield similar percentile values. The + and – signs denote the EF calculated using the background with subtracted 2 standard deviations of its variability (EF BC-), and from the background with added 2 standard deviations of its variability (EF BC+). Their positive and negative relative errors (rel. err.) are also reported. In the last column is the EF is calculated with a 2 s integral instead of a 10 s integral.

	EF BC- (g/kg)	**EF BC (g/kg)**	EF BC+ (g/kg)	-Rel err. (%)	+Rel err. (%)	**EF BC 2 s (g/kg)**
Minimum	0.24	0.23	0.23	-0.04	0.00	0.14
25th percentile	0.50	0.55	0.59	0.09	0.07	0.49
Median	**0.63**	**0.73**	**0.88**	**0.14**	**0.21**	**0.73**
75th percentile	0.85	1.01	1.25	0.16	0.24	1.15
90th percentile	1.17	1.35	1.69	0.13	0.25	1.65
Maximum	1.99	2.44	3.54	0.18	0.45	5.14

Navidi W.: Statistics for engeneers and scientists, vol 40, eds. D B Hash and L Neyens (New York: McGraw-Hill), 2001.

Aerosol size distribution may change from one type of vehicle to another (i.e. gasoline to diesel). This point was not discussed here but should be mentioned as it has some implication (for health for instance). Also can the author provide a reference to state that the mass absorption efficiency used here (i.e. the same for every type of vehicle) is not dependent of the aerosol size distribution (e.g. may not be biased by the type of vehicle investigated gasoline vs diesel).

The aerosol size distribution does indeed change, not only from one vehicle to another but within singe vehicle's operation. This was investigated in Ježek et al. 2015 (also Sharma et al., 2005), where it was shown that the exhaust constituents change with driving regime. It was also shown that some cars tend to emit more particles in the nucleation mode while others emit more soot particles which would then adsorb the precursors or coagulate with the smaller particles thus reducing them in number but not in mass. It was shown that, while the number size distribution changes with the engine load and is different for different times when measuring ambient air, the BC mass distribution depends on the load of the engine, but not on the time of day, when performing ambient measurements (Ning et al., 2013).

Ježek et al. (2015) have shown that BC and particle number concentration of particles between 50 nm and 200 nm correlate well during a chase measurement. This is the size range in which BC particles are expected to lie. The Rayleigh–Debye–Gans theory (Sorensen, 2001) treats particles as fractal-like aggregates of small primary individual spheres (monomers) and calculates the absorption of the aggregate as the absorption of a single monomer particle multiplied with the number of such particles in the aggregate (an example of such calculation can be found in Kim et al., 2015). Since the absorption and mass of the aggregate increase linearly with the number of monomers, the mass absorption cross-section is independent of the aggregate dimension. This is consistent with the near-road observations by Ning et al. (2013).

We added a short description on this to the last paragraph in section 2 before section 2.1, so that the paragraph reads:

The mobile measurement platform used for the on-road chasing measurements is described in detail in Ježek et al., 2015. We used instruments with high time resolution (1 to 10 s) the

Carbocap GMP343 (Vaisala) to measure CO_2, the Aethalometer AE33 prototype version β (Aerosol d.o.o.), the Fast Mobility Particle sizer (TSI), for the on-road campaign we added also a Nitric Oxide Monitor and an NO_2 converter (models 410 and 401, 2B Technologies). For the Nitric Oxide Monitor the sampling line was a Teflon tube, while for the rest we used static-dissipative tubing. The instrumental details and measurement uncertainties are summarized in Table 1. The Aethalometer data was compensated for the loading effect using the Drinovec et al. (2015) compensation algorithm. While the size distribution of the exhaust particles change with the engine operation (Ježek et al. 2015; Sharma et al., 2005), a fact that might have implication in the context of the health effects of exhaust particles, Rayleigh–Debye–Gans theory (Sorensen, 2001; an example of such calculation can be found in Kim et al., 2015) predicts the mass absorption cross-section independent of the size distribution of the fractal aggregates. This is consistent with the near-road observations by Ning et al. (2013).

Kim, J., Bauer, H., Dobovičnik, T., Hitzenberger, R., Lottin D., Ferry, D., Petzold, A.: Assessing Optical Properties and Refractive Index of Combustion Aerosol Particles Through Combined Experimental and Modeling Studies, Aerosol Sci. Technol., 49 (5), 340-350, doi:10.1080/02786826.2015.1020996, 2015.

Ning Z., Chan, K. L., Wong K. C., Westerdahl, D., Močnik, G., Zhou, J. H., Cheung, C. S.: Black carbon mass size distributions of diesel exhaust and urban aerosols measured using differential mobility analyzer in tandem with Aethalometer, 80, 31-40, doi:10.1016/j.atmosenv.2013.07.037, 2013.

Sorensen, C. M.: Light Scattering by Fractal Aggregates: A Review, Aerosol Sci. Technol., 35, 648-687, 2001.

EF differences from one category to another are more in the range of tens of percent than orders of magnitude. The figures displaying EF should be consistent with this and should rather be in linear scale (and not log). Although I agree that box plots are relevant, the figures currently prevent from comparing easily medians (as discussed in the text). The same stands for the dispersion of EF values (Fig. 2); it is difficult here to evaluate here how far the super emitters from the median are. Last but not least, there are no 'statistical' discussions comparing average/median, dispersion (standard deviation), etc.

We chose the log scale because we didn't find the figures with linear scales as clear. In linear form, the super emitter tails dominate the whole scale so that the distribution and the smallest values are not as visible. Especially the lowest EFs are thus completely obscured – this should not be the case, because these are the values that indicate improvement in vehicular emissions, and the differences are often observed on the scale of an order of magnitude. This illustrates that the emission abatement strategies are indeed successful. We have added the linear scale figures in the supplemental material and refer to both figures at the same time in the text. In order for the distribution to be visible, we broke each of the y axes. The super emitter tails are also the reason the median and the average value do not overlap. We added a sentence on the "statistics" just before the end of section "3.5 Contribution of high emitters to the measured fleet":

…

The results in Figure 5 show that 25% of highest emitting vehicles in each vehicle category produce 50 to 65% of BC emissions, 47 to 55% of NOx emissions and 61 – 87% of PN emissions. The high contributions of super emitters are the statistical cause of the non-symmetrical distributions and are responsible for the mismatch between the median and the average EF values. Excluding high emitting vehicles or improving their emission rates by retrofitting them with additional after treatment devices, such as was the case in Port of Oakland, US, (Dallmann et al., 2011) can decrease traffic emissions.

We have added the plot in the linear scale to supplemental materials as Figures S3 and S4.

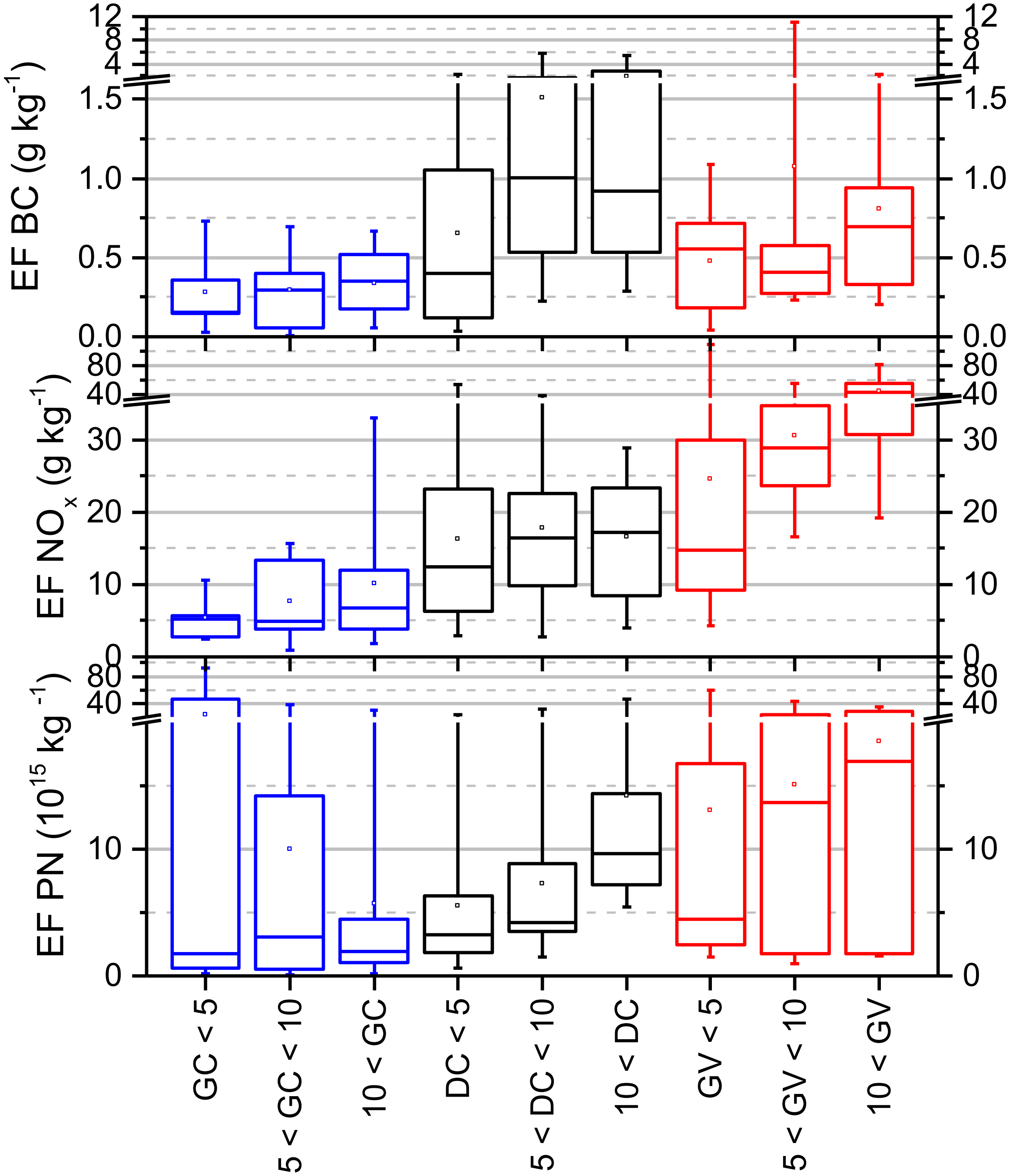

Figure S3. BC and NOx EF according to different vehicle categories and age group subcategories: gasoline passenger cars (GC, blue), diesel passenger cars (DC, black), and goods vehicles (GV, red). Note the EF linear scale; same figure in logarithmic scale can be found as Figure 3.

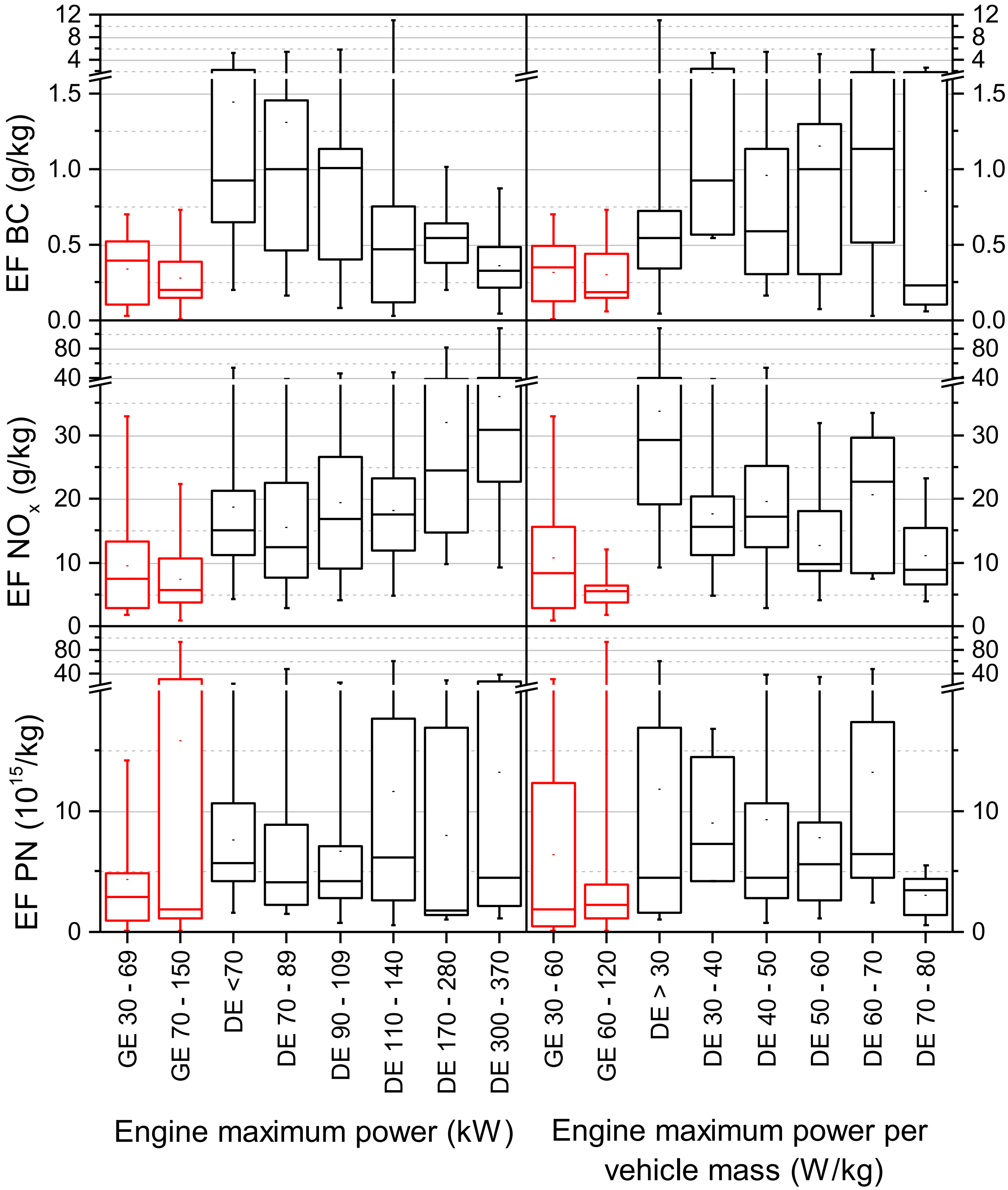

Figure S4. BC and NOx EFs according engine power (left) and size (right); red boxes present gasoline engines (GE) and black boxes present all diesel engines (DE) regardless of their vehicle category. Note the EFs are on logarithmic scale; same figure in logarithmic scale can be found as Figure 4.

The way the fleet is categorized here is not conventional. Although it is fully justified, correspondence with EURO is coming late in the paper. It should be earlier. The same stands for the reason why the paper is investigating the ratio of maximum engine power to vehicle curb weight. Explanations are provided but almost in the end of the paper.
There are several different vehicle categorizations in use – e.g. traffic counters, toll stations, emission standards, etc. The classification we used is described in section 2.2 and is based according to the Directive 2001/116/EC (2002). Data gathered by agencies that control the vehicle registry use this classification and report the data to Eurostat, which in a way also uses the same classification. The age delimiters were used according to the time the EURO 3 and EURO 4 standards came in to force.

The delimiters for engine maximum power and maximum power per vehicle mass were chosen, so that each group would have sufficient number of vehicles inside a group for a statistical analysis. The purpose of this exercise was to avoid all the thus far used classifications, and see how mechanical and physical features of the vehicles affect the emissions. Vehicles are versatile but some vehicles have different chassis but the same engine, or the same chassis but different weight, a car (like a Reanult Kangoo) or a van can be registered as a personal vehicle or a light goods vehicle, even though it is the same vehicle, and so on.

We reworded and added to the second paragraph in section 3 (page 15367, line 2):
We compared our measured fleet composition on the vehicles' age and size with the information on the Slovenian and European vehicle fleet statistics (section 3.1). We present our results as BC, PN and NOx EF distributions for the vehicle categories and compare them to results of other similar studies in section 3.2. We further demonstrate how the EFs of each group depend on their age, by grouping them according to years when EURO3 and EURO 4 standards became effective. Even though the purpose of use is indeed important when classifying vehicles; but with such categorization the mechanical features may be overlooked. A single car (for example Renault Kangoo or similar) can be classified as a personal vehicle or a light goods vehicle. To see how mechanical and physical features of the vehicles affect the emissions, we disregarded the purpose based categorizations and observed the effect of engine maximum net power, and the ratio between engine maximum net power and vehicle mass in section 3.4. In section 3.5 we present the contribution of high emitters to the sampled fleet cumulative emissions.

It would have been interesting to compare the EF results with those available for instance in COPERT (V4) which is often used to feed models dedicated to traffic emissions.

We mention EMEP/EEA NOx EF in the text (P 15372, L 3). We did not include them in the Table 7 because we did not want to include other study types such as tunnel measurements, chassis dynamometer tests or measurements with portable emission measurement systems, as they have already been discussed in other studies (Shorter et al., 2005; Wang et al., 2012). The EFs used in COPERT are from EMEP/EEA emission inventory guide book. Depending on the available data for traffic emission modeling the guide book offers different EF calculation approaches that are based on vehicle speed. There are no reported BC or PN EF in the EMEP/EEA emission inventory guide book 2013, however they do report NO_x EF in g/km in Tier I. Using the fuel consumption values that are also report in the aforementioned guidebook, we get the following NOx EF values: gasoline cars 8.73 g/kg (4.48–29.89 g/kg), diesel cars 12.96 g/kg (11.2–13.88 g/kg), LGV 14.91 g/kg (13.36–18.43 g/kg), and HGV 33.37 g/kg (28.34–28.29 g/kg). We can see that the values are comparable to the results of our study and the study of Carslaw and Rhys-Tyler, 2013. Although there are some variations e.g. our diesel cars' EFs are a bit higher (median 15.43 g/kg) and goods vehicles NO_x EFs reported in our manuscript are slightly lower (median value 27.71 g/kg) than those reported in the EMEP/EEA guide book. We are preparing a separate manuscript where we will compare the results of a traffic emission model using the EMEP/EEA EFs and those from our on-road measurements, and where we will discuss the comparison of these two EF datasets in more detail. Given the interest, we have included the COPERT NOx EFs in Table 7.

Specific comments and technical corrections :

- Is there another word to describe the fleet of "goods" vehicle? This term is not easy to understand (especially in the abstract).

The term "goods vehicles" was used because we merged the N2 "light goods" and N3 "heavy goods vehicles" in to one group – using the definitions and terminology from Directive 2001/116/EC (2002). Using the word "truck" instead of "goods vehicles" was in our opinion not appropriate because this could imply we only measured lorries and that we excluded road tractors or trailer trucks, which was not the case. The terminology depends a lot on which data-set one is using. To avoid the additional confusion of different categorizations and naming of vehicles we decided to follow the terminology in Directive 2001/116/EC (2002) as close as possible.

- Be consistent through the manuscript when using hyphen (e.g. "real world" and "realworld").

We changed all “real world” and to hyphened “real-world”. Changes are on page: 15357, line 17, 18 and 26; 15361, line 18; 15375 line 15, and page 15379 line 17.

- Several terms may not be adequate. For instance, “resemble” which is used 5 times in pages 15371-1732. “In this segment : : :” (page 15373), “: : : arranging : : :” (line 17, page 15377)

The following changes were made in the manuscript:

P 15371, L 5: The BC EF median of goods vehicles we measured (0.47 gkg^{-1}) is similar to…

P 15371, L 15: The median value of the NO_x EF distribution (27.7 gkg^{-1}) observed for goods vehicles lies closer to…

P 15372, L 1: The NO*x* EF values of the gasoline and diesel cars in this campaign (6.3 and 15.4 gkg^{-1} respectively) coincide with…

P 15372, L 7: The goods vehicles NO*x* EFs (27.7 gkg^{-1}) from this campaign agree with…

P 15372, L 27: HGV PN EF from Ban-Weiss et al. (2009) (4.7x 10^{15} kg^{-1}) and from the study of Hudda et al. (2013) (4.2 and 5.2 x 10^{15} kg^{-1}) coincide with…

P15373, L 8: “In this section…”

P 15377, L 17: “The cumulative emission distribution of our vehicle fleet were calculated for vehicles from highest to lowest emitters…”

- Page 15362, line 17: “: : : the Slovenian vehicle fleets stills from the measurements : : :”. I do not understand this sentence.

The following change was made in the manuscript: “stills changed to “photographs”

- Page 15362, line 19. Probably a dot is missing after “region”

Indeed. A dot was added to page 15362, lin2 19 so it reads: … region. Slovenian …

- Page 15363 line 11; “sizer” instead of “seizer”

The following change was made in the manuscript page 15363, line 11: “seizer” was changed to “sizer”

- Page 15366, line 20. No end in the sentence “Emission factor measurement results”

“Emission factor measurement results” is the title to section “3. Emission factor measurement results”. This section title is missing in the ACPD manuscript and all sections are mislabeled. *All the following section numbers were changed accordingly*!

- Page 15367 line 25. ": : : good agreement : : :". It is more related to the representativeness than agreement.

The following change was made in the manuscript P 15367, L 25: The gasoline and diesel car engine displacement segregation of the campaign fleet is representative of the European and Slovenian fleets, where…

- Page 15374, line 29. I believe that the values taken here (0.7 and 43.95) were from Fig. 3. It should be better stated.

The following change was made in the manuscript page 15374, line 29: The 10 year or older goods vehicles (BC and NOx EF respectively 0.7, 43.95 gkg^{-1}, please see Figure 3 and S3)…