Response to comments from anonymous referee #3

The manuscript is well written, with a good methodological approach and of interest for its immediate application for air quality management. Results on real-world emission factors and rates from separated vehicle categories are of primary interest for the air quality community. This manuscript has the additional value of evaluating the impact of environmental policy for HDV fleet. This kind of studies are urgently needed in Europe, where beside the drawback of CRT systems for NO_2, the EURO standards did not reduce real-world primary NO_2 emission per single light-vehicle.

I would like authors to address the following points:

1. *The increase of NO2/NOx ratio is likely due to the CRT retrofit, but the strategy of CARB was also that of banning old vehicles. Do authors have sufficient data to distinguish between these two phenomena when evaluating their impact on single pollutants?*

   The CARB ban on old vehicles was focused on the adjacent Ports of Los Angeles and Long Beach. We would expect that the ban on older vehicles would most strongly reduce the high end of the emission factors, especially NO and BC, while CRT retrofits would tend to shift the entire distribution of EFs downward, except for NO_2. We appear to see both of these changes on I-710, the primary truck route, but separately quantifying the extent of each of these contributions was not possible with our methodology.

2. *Table S2 should also include precipitation data and some estimate of traffic congestion found during each route*.

   Rain is infrequent in Southern California and rain did not occur during our sampling campaign. A note to that effect has been added to Table S.2: "Sampling days, hours and meteorological conditions."

   Speed is the best measure of congestion in Los Angeles. The speed plot for measurements on freeways other than 110N has been moved from the Supplement to the main article. This plot shows our speed distribution was reasonably representative of the speed distribution on LA freeways and that both free-flowing and congestion conditions were measured.

3. *Section 3.1. The first sentence does not seem true for CO and PB-PAH or at least the statement cannot be observed in the plots. The same for line 20 of the same page*

The first sentence referenced has been modified to read:

> "LDV EFs on SR-110 showed ~~more than~~ about an order of magnitude range, as shown in Figure 1."

Please note that specific units (g/mile) were mentioned in Line 20 for the comparison made. However, it has been edited to be consistent with g/kg units. The edited version reads as:

> "HDV emission factors (per kg fuel) exceeded LDV EFs by at least an order of magnitude for particulate pollutants and about four fold for NO and NOx. On average the HDD fleet on freeways other than I-710 emitted 4 times higher NOx, 10 times higher PN, and 6 times higher BC per quantity of fuel burned than LDGs." In addition, CO emission factors have been removed from the results.

4. The results presented seem to suggest that NO_2 can no be longer used as reliable indicator of traffic PM emissions, given the impact of retrofit systems and (in Europe) the change of NO_2/NO_x primary emissions. Can authors discuss if, basing on their study, BC is a more reliable air quality metric?

Yes, this is one possible implication of the results. For HDDs, a large portion of the PM emissions are BC, and decreases in BC will reflect reductions in PM, while changes in NO2 will not do so if CRT retrofits are both increasing NO2 and decreasing PM. For gasoline-powered vehicles, the situation is more complex. We have added the following text to the manuscript:

> "This finding may indicate that NO_2 trends may no longer be a valid way to track reductions in diesel PM emissions. For example, studies like Burnekeef et al., (2009) have used NO_2 long-term trends as indicators of traffic PM emissions."

5. *Fig S4 should be updated also with LDV.*
Fig. 5. Does 110 label refer to both stretches or only the southern? If so, please add the northern stretch as well.

Figure S4 has been edited as per suggestion.

Regarding Figure 5, 110 only refers to southern segment (110S). The figure caption has been edited to clarify this. The northern segment, SR-110 has no trucks and the motive behind this figure was to illustrate that despite substantial differences in truck fractions, total emissions are comparable. We don't believe that addition of SR-110, the LDG-only traffic freeway segment, is necessary.

6. I generally agree with the comments of other Referees, mostly concerning:
The evaluation of this technology for the European case, where LDV are more separated between diesel and gasoline engines, with respect to US.

On the issue of applicability in Europe**,** as long as reliable statistics on the fuel type division are available the same methodology of partitioning emissions could be extended to three groups (HDV – diesel, LDV – diesel and LDV – gasoline). Though mathematically feasible, we think it will be challenging to obtain the fuel type statistics required. The technique requires at least two of the three $\Delta P/\Delta CO_2$ ratios to be known for direct application. These could be obtained from roadway measurements in areas where only vehicles with certain fuel type are allowed. (If they exist in Europe, though authors are not as familiar with Europe). It is unlikely that there are places where only gasoline or diesel LDV are allowed, ruling out the possibility for two known $\Delta P/\Delta CO_2$ ratios. But it is probably possible to find locations near ports/railway yards that are heavily dominated by heavy duty diesel truck traffic. Once at least one, i.e., $\Delta P/\Delta CO_2$ ratios for diesel HDV are known, our suggestion is to first apportion diesel HDV emissions out from mixed fuel traffic measurements as opposed to our approach where we had access to only gasoline LDV traffic emission and we apportioned those out using the measured ratios. The remaining emission will then have to be apportioned out in diesel LDV and gasoline LDV fractions, but since both $\Delta P/\Delta CO_2$ ratios are unknowns, they will have to be estimated. Each observation post HDD apportionment would give a set of linear equations (of the type below for each observation) which could be solved to get an approximate solution through linear least squares for an over-determined linear system.

Equation for each observation, where 'o' is number of observations:

$$(\Delta P/\Delta CO2)\_LDG * \text{fraction of } LDG_o + (\Delta P/\Delta CO2)\_LDD * \text{fraction of } LDD_o = (\Delta P/\Delta CO2)\_\text{apportioned}_{LD}$$

However, the accuracy of regressed LDG and LDD ΔP/ΔCO2 ratios would depend on the accuracy of fraction of LDG and LDD input, which is likely to be less accurately known than the division between LD and HD vehicles. If fraction of LDG and LDD estimates is very uncertain, it might be a better approach to chase individual vehicles and calculate individual EFs.

7. More details are needed about the background concentrations subtracted. Authors mentioned they used the first percentile, but it is not clear on what time series. Were data from the same freeway link assembled together or divided per time of the day? Background concentrations likely vary from one hour to another. Representativeness of the data used must be discussed.

Roadway concentrations are elevated compared to off-road or ambient concentrations, and these concentrations act as a kind of baseline concentration to which emissions from current vehicles are superimposed. Our goal was to measure the increment in pollutant concentration over and above this baseline. Background concentration was calculated as the first percentile of the concentrations on each freeway segment for each run. This percentile approach also gave us a baseline measure that was the spatial and temporal equivalent of elevated roadway concentration measurements, and directly comparable to the traffic emissions we were measuring since they were made with the same instruments. The percentile approach was briefly mentioned in the manuscript at Lines 11-13/Pg. 18722, and we have provided more details in the revised version. Also, for clarity "background" has been replaced by "baseline." The relevant section now reads as:

> "Roadway baseline values in Equation 1 were estimated as the first percentile of pollutant concentrations observed on each freeway link. Since we were attempting to measure concentration increases from current traffic during a short time interval—over and above the elevated concentrations already present on the freeways—a percentile concentration value was the most appropriate indicator of baseline roadway concentrations. The lowest few concentration percentiles were relatively insensitive to superimposed traffic (CO_2 percentile profiles for a subset of runs have been plotted in Figure S3 in SI). Concentrations at a distant location away from the freeway were lower and would have provided an artificially low baseline estimate and upwardly biased the EFs."

Also, the following text has been added to Methods section that clarifies which time series were used.

> "Instruments logged data at different intervals (1-10 seconds) and all data were averaged over 10 seconds. Freeway segments were demarcated in the 10 second data time series based on location information collected using the GPS."
>
> "All statistics (median or first percentile concentration) required to calculate EF using Equation 1 were determined from the time series for each freeway segment, typically tens of miles long. If multiple runs were conducted on a freeway within a day, the time series for each run was analyzed separately."