We appreciate the referee's time and feedback, which have resulted in significant improvements to our manuscript, as detailed below:

Shusterman et al. present and analyze results from the low-cost, high-density CO_2 *monitoring network BEACO*$_2$*N to demonstrate that such a network allows investigating hyperlocal sources, e.g., highway traffic, and to track emission changes due to mitigation measures. Key findings, like an experimentally determined correlation length and the strong correlation of local* CO_2 *enhancements with traffic are important results for the urban GHG research community. Overall, the manuscript is very well written, nicely structured, and concise.*

1.) However, some further detail on the methodology would be instructive for other (and future) researchers attempting to use similar approaches, which would ensure that the paper has the best possible impact. The methods applied are properly referenced, but, e.g., the work of de Foy is very recent and some more information might be useful for the reader.

The referee's request for a greater level of detail in our methodological descriptions is a sentiment shared by the other referee as well. We have adjusted the text accordingly, as detailed in response to the specific comments below as well as our other referee response.

2.) Furthermore, the authors do not clearly define the terms used for spatial scales, e.g., "regionwide" (see specific comments). As different groups/communities use different definitions of "regional," it seems imperative that this is added to the manuscript to avoid confusion.

Please see our responses to the specific comments below.

3.) The authors refer to MRV and that this network would/could be useful. While the work described here echoes the concept of MRV, MRV itself, as introduced by the Bali action plan (UNFCCC), seems not to be the best goal. I would argue that providing atmospheric-based constraints on emissions would be very valuable by itself and can enormously help (local) stake-holders without the complications of being integrated into an MRV framework.

We acknowledge that some readers, the referee included, may adhere to a much stricter definition of MRV activities and have removed all references to MRV from the text in favor of language referring to atmospheric-based constraints on emissions more generally; see our responses to the specific comments below for details.

After addressing these comments, I would fully recommend this work for publication in ACP as it is an important advance in the field and will be of great interest to the community.

Specific comments:

P1 L9: Consider adding "at subnational scale" as national CO_2 *emissions are usually fairly easy to report based on consumption data compared to, e.g., CFCs,* N_2O*, or* CH_4*, and MRV frameworks exist under UNFCCC (e.g., https://unfccc.int/sites/default/files/non-annex_i_mrv_handbook.pdf). For cities MRV has also been developing (e.g., the GHG protocol), but the authors could highlight the added/complementary value of atmospheric information.*

We have updated the text to refer to subnational scales and have removed this and all subsequent references to MRV frameworks:

"Urban carbon dioxide comprises the largest fraction of anthropogenic greenhouse gas emissions and yet quantifying urban emissions at subnational scales is highly challenging, as numerous emission sources reside in close proximity within each topographically intricate urban dome."

"To support this effort, there is a clear need for monitoring strategies capable of describing emission changes and attributing those changes to the relevant policy measures (Pacala et al., 2010)."

"However, cities also present the largest atmospheric monitoring challenge in that many disparate emission sources combine with complex topography."

"A considerable amount of emission estimation work has been invested in the development of activity-based emission inventories for selected metropolitan areas […]"

P4 L19: Please give an estimate of what scale "regionwide" refers to.

While the specific sentence to which the referee refers no longer exists in its original form, we have updated the first reference to regional spatial scales in the text to clarify our intended meaning of the term:

"LCC, however, also exhibits relationships with more distant sites, indicating a sensitivity to more regional-scale (10–30 km) influences."

P4 L25: Why did you choose the 10th percentile to define "regional" and not, e.g., the 5th or 20th percentile?

As mentioned in the text, the 10th percentile is chosen "to account for measurement error […] as well as any nearfield draw down from the local biosphere." We note in Sect. 2.1 that Shusterman et al. (2016) found the 1-min mean measurements from the BEACO_2N CO_2 monitors to possess an uncertainty of less than ±4 ppm, which amounts to between 5% and 10% of the typical ambient CO_2 signals observed in our urban domain. We therefore adopt the conservative upper limit of 10% to allow for some influence from the biosphere, although a precise quantification of this component of the signal is beyond the scope of this study. Thus, assuming an overall 10% uncertainty in an arbitrarily chosen site's ability to characterize the regional signal, we define the bottom 10th percentile of the observations as our best estimate of this quantity.

The manuscript text has been updated to direct the reader to the reasoning behind this quantity:

“The bottom 10th percentile is chosen (rather than the absolute minimum) to account for measurement error (±4 ppm at 1-min resolution; see Shusterman et al., 2016) as well as any nearfield draw down from the local biosphere; negative values in the local signals are likely attributable to some combination of these effects.”

P4 L33: Please correct to “Figure 4.”

This typographical error has been corrected.

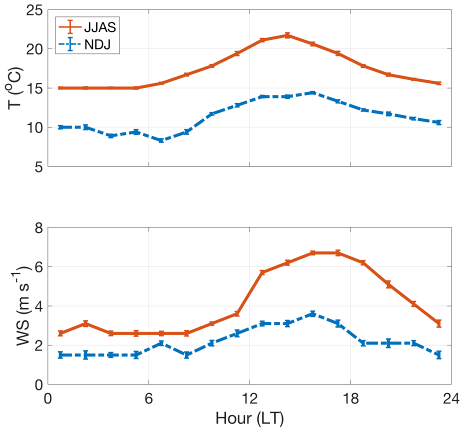
P5 L3: The daily cycle is mainly driven by boundary layer height dynamics–the local traffic flux is the superimposed fluctuation here, in my opinion. It surely causes a modification, e.g., by causing the morning and evening peaks to be more pronounced. However, different studies in rural regions have largely similar diel cycle shapes (e.g., Garcia et al., 2012 https://www.tandfonline.com/doi/abs/10.3155/1047-3289.58.7.940; Perez et al., 2012 https://www.sciencedirect.com/science/article/pii/S0048969712007498).

We agree with the referee that the traffic flux is the superimposed fluctuation here, and the text as written reflects this sentiment:

“This diurnal profile corresponds well with known patterns in traffic emissions–which are largely consistent across seasons–superimposed on diel fluctuations in boundary layer height and/or biosphere activity that vary in timing and magnitude according to the season.”

P5 L6: It seems counter-intuitive that the PBLH changes earlier in winter (also compared to other studies), as more energy is introduced into the system during summer months to break the NBL (as the solar insulation is stronger and the sun rises earlier). Please provide additional data, e.g., PBLH or other atmospheric proxy information in the appendix to support your interpretation.

Unfortunately, there exist no direct PBLH observations in the area with adequate temporal resolution to inform this analysis. Instead we show the median diel cycles in the summer vs. wintertime temperatures and wind speeds observed at the Port of Oakland International Airport’s NOAA Integrated Surface Database station (https://www.ncdc.noaa.gov/isd/) below:



We see that the increases in atmospheric proxies that might be associated with PBLH changes occur at almost identical times of day across seasons, even if the sun rises earlier and more energy is introduced into the system overall during the summer months, as the referee suggests. We do acknowledge, however, that the seasonal differences in PBLH changes are not the only possible explanation for the difference in the diel cycle in regional CO_2 concentrations, and have updated the text to reflect an additional possibility suggested by the other referee:

"Namely, these results might be interpreted to conclude the nighttime boundary layer in the BEACO_2N domain to be shallower during the winter months, producing a larger regional increase in response to rush hour traffic. The wintertime layer also appears to expand and re-contract earlier in the day than the summertime layer, resulting in both an earlier minimum and an earlier rise in afternoon–evening concentrations. The larger amplitude of the wintertime diurnal cycle may also reflect the greater influence of daytime photosynthesis and nighttime respiration during the San Francisco Bay Area's rainy winter season."

P5 L32: Why is the other methodology not shown in the supplement and why is this sentence in brackets? Seems to be an interesting finding/information.

We have removed the parentheses around this statement and have added a figure to the supplement that illustrates the results of this alternative methodology.

P5 L34: You could also refer to the large amount of traffic tunnel studies that have similar findings and are very straightforward (no other source besides traffic), e.g., references in https://www.atmos-chem-phys.net/14/12871/2014/acp-14-12871-2014.pdf.

We appreciate the referee's suggestions of additional related studies and believe that such tunnel-based measurement campaigns contribute very important information to mobile emission estimation efforts. However, in the interest of succinctness, we choose to forego a broader discussion of the many analyses that use CO_2 as a baseline against which the concentrations of co-emitted species are normalized and instead limit our discussion to studies that analyze the traffic dependence of CO_2 concentrations in their own right (i.e., Maness et al., 2015).

P5 L35: One question raised would be how long would you have to observe to confirm this 17% trend? Which is answered at P7 L4 for 11–30% emission changes. Consider removing the discussion of the 17% here.

As suggested, we have moved this discussion to occur later in the manuscript.

P6 L15: How exactly are the wind speed quartiles subdivided (and why)? See general comment 1.)

As noted in the text, the wind speed quartiles are subdivided to allow for a "nonlinear relationship" between CO_2 concentrations and this explanatory variable. In Gaussian dispersion modeling, for example, the downwind concentration of a given pollutant is inversely (rather than linearly) proportional to wind speed. Because our regression method is by definition linear, subdividing the wind speeds in this way allows us to decompose more complex mathematical relationships into linear components. We have updated the text to give more detail regarding the exact methodology of this approach:

"The nonlinear relationship between CO_2 concentrations and wind or boundary layer height is captured by dividing these meteorological datasets into quartiles and assigning each observation a value between 0 (at the maximum of the quartile) and 1 (at the minimum) using piecewise linear interpolation. The wind speed quartiles are further subdivided by wind direction and reassigned values of 0–1 accordingly before fitting a linear coefficient to each subset. The time of year is represented as a sum of sines and cosines with annual or semiannual periodicities whose values also vary between 0 and 1 and whose amplitudes are determined by the linear regression. Zeroes and ones are used to designate each hour of each type of day of the week as well. For example, timesteps corresponding to 0800 LT on a Monday may be assigned a 1 while all other timesteps are set to zero before the linear regression is performed. As a result, the MLR factors derived for each of the preceding explanatory variables can be interpreted in units of ppm CO_2. Meanwhile, the temperature and specific humidity variables are treated by calculating their difference from their mean values and dividing by their respective standard deviations before each is fit to CO_2 with a single linear coefficient, which will have units of ppm $^{o}C^{-1}$ and ppm $(kg_{water}\ kg_{air}^{-1})^{-1}$, respectively.

The independent variable leading to the greatest square of the Pearson correlation coefficient is then combined with each of the remaining variables and a second regression is performed. The two-input combination leading to the largest increase in the correlation coefficient is then combined with each of the remaining variables, and so on, until the addition of a new independent variable no longer increases the r^2 value by at least 0.005."

P6 L26: Why are Mondays and Saturdays not shown in the supplement?

A figure depicting MLR factors derived for Mondays, Fridays, and Saturdays has been added to the supplement.

P6 L29: Could you quantify to which degree the atmospheric dynamics have been controlled for? Claiming that it is "partially controlled for" does not automatically mean that the residual only/primarily reflects emission changes.

Without knowledge of the true emissions within a given site's footprint of sensitivity, we cannot quantify the degree to which atmospheric dynamics have been controlled for. The fact that the MLR factors remaining after "partially" controlling for dynamics may primarily reflect emission changes is a hypothesis rather than a premise of this study, a hypothesis that the discussion goes on to support with a first order, proof-of-concept analysis of the diel cycles in these factors. We have updated the text to clarify the speculative nature of this claim, and also to provide additional detail regarding the diel cycle analysis:

"The dependencies on time of day and day of week derived via this method are hypothesized to primarily reflect the changes in emissions, as the influence of the coincident changes in atmospheric dynamics has been at least partially controlled for. For reference, the corresponding Tuesday–Thursday and Sunday diel cycles in the total CO_2 observed at each site are shown in Fig. 9. Indeed, we do observe some of the same intuitive patterns in the linear regression coefficients, such as higher coefficients on weekday mornings corresponding to higher rush hour traffic emissions on those days, but with greater opportunity to differentiate between days of the week, especially around noon when raw concentrations are generally similar. As expected, the Tuesday–Thursday enhancement in the MLR factors is larger at the sites located close to a freeway (e.g., up to 520% of the corresponding Sunday MLR factor at FTK) but is less pronounced at LBL (70%), which is farther away from major mobile sources."

P7 L4: What is your confidence of the reported detection of such a trend within 2–3 years? 95%? How was this calculated?

The stated uncertainty in the regression slopes (11–30%) is the standard error, i.e., the 68% confidence interval. Assuming that the 35% reduction in CO_2 emissions per vehicle required by fuel efficiency regulation occurs evenly over ~10 years necessitates a 3.5% change in CO_2 emissions per vehicle per year. Thus, with a regression uncertainty of 11%, this 3.5% annual trend is detectable within just over 3 years using the observations from a single site. Even modest improvements in our ability to leverage information from $N > 1$ sites within the network would allow for trend detection with greater confidence and/or shorter timescales if, for example, different sites' observations are found to be sufficiently independent to scale down the uncertainty by $\sqrt{N}$. We have updated the manuscript text to clarify this point, include the re-located discussion of the LAN 17% slope uncertainty, and present the timescale of detection more precisely:

"When we examine the relationship between these multiple linear regression coefficients and morning traffic flow as we did at LAN (Fig. 10), we again find positive correlations. The standard error of the slope of the linear regression is calculated as the standard deviation of the model–data CO_2 residuals divided by the square root of the sum of the squared differences between each traffic flow increment and the mean traffic flow. The uncertainty in the slopes is thus found to be 11–30%, indicating that analysis of a single site could be used to detect as small as 11% changes in average emissions per vehicle, an improvement upon the 17% slope uncertainty calculated for the near-highway LAN site. For reference, under the Corporate Average Fuel Economy standards, the state of California aims to achieve a fleet-wide average fuel economy of 54.5 miles per gallon by the year 2025 (US EPA, 2012), corresponding to a 35% decrease in emissions relative to the 35.5 miles per gallon economy of 2012–2016 model year vehicles. Assuming a steady decrease in emissions of 3.5% per year, one BEACO_2N site is therefore sufficiently sensitive to detect such a trend with 68% confidence in as little as 3 years. By leveraging observations from multiple independent sites, even greater confidence and/or shorter timescales could be achieved."

P7 L17: The assumption that plumes can be detected within an urban area should be supported, e.g., by citations. At scales below 1 km^2 it seems that street canyon effects, building disturbances, etc. could play an important role and hinder the application of concepts such as "plumes," see, e.g., Lietzke and Vogt (2013; https://www.sciencedirect.com/science/article/pii/S1352231013002069) that also investigated traffic emissions at street scale.

Previous work using BEACO_2N measurements has provided preliminary evidence that plume-like events (if not "plumes" in the strictest sense) can be detected at this scale in urban areas (Kim et al., 2018). These plume-like events are not necessarily representative of a single vehicle's tailpipe, for example, but are nonetheless characterized by a sharp, distinct enhancement above background concentrations and have been shown to be correlated with average emission factors expected for a given vehicle fleet. We have updated the text to include a reference to this important proof-of-concept study:

"Prior studies have demonstrated a methodology for detecting plume-like events in the BEACO$_2$N NO_x and CO observations (Kim et al., 2018), and the ratio of these species to CO_2 provides a unique signature for each different CO_2 source (e.g., Ban-Weiss et al., 2008; Harley et al., 2005; Lopez et al., 2013; Nathan et al., 2018; Turnbull et al., 2015), allowing subsets of the data record to be directly attributed to specific (e.g., mobile) source types and allowing the relationship between these specific activities and CO_2 mixing ratios to be derived more precisely."

P8 L5: I would suggest reconsidering the wording here, especially as you refer to MRV earlier in the manuscript. This work strongly supports the conclusions of Turner et al. (2016), but it seems you have validated and not verified them.

We appreciate the referee's attention to detail in this case and have updated the text accordingly:

"This work thus provides an important data-based validation of the conclusions of Turner et al.'s theoretical analysis."

References:

Kim, J., Shusterman, A. A., Lieschke, K. J., Newman, C., and Cohen, R. C.: The BErkeley Atmospheric CO_2 Observation Network: field calibration and evaluation of low-cost air quality sensors, Atmos. Meas. Tech., 11, 1937–1946, doi:10.5194/amt-11-1937-2018, 2018.

Shusterman, A. A., Teige, V. E., Turner, A. J., Newman, C., Kim, J., and Cohen, R. C.: The BErkeley Atmospheric CO_2 Observation Network: initial evaluation, Atmos. Chem. Phys., 16, 13449–13463, doi:10.5194/acp-16-13449-2016, 2016.